

Europe: so many languages, so many cultures

HEIN STEINHAUER

ABSTRACT

The number of different languages in Europe by far exceeds the number of countries. All European countries have national languages, and in nearly all of them there are minority languages as well, whereas all major languages have dialects. National borders rarely coincide with linguistic borders, but the latter (including dialect borders) mark by their nature also more or less distinct cultural areas. This paper presents a survey of the different language families represented in Europe: Indo-European, Uralic, Altaic, and the four Caucasian language families, each with their sub-branches and individual languages. Some information is given on characteristic structural phenomena and on the status and history of these languages or language families and on some of their extinct predecessors. The paper ends with a short discussion on the language policy and practices of the institutions of the European Union. Europe lacks a language with the status and power comparable to Indonesian in Indonesia. The policy is therefore based on equal status of all national languages and on respect for all languages, including national minority ones. The practice, however, is unavoidably practical: “the more languages, the more English”.

KEYWORDS

Language families in Europe: Uralic, Caucasian, Altaic, Indo-European; language spread through history; structural variety of European languages; language policy in Europe and European countries.

HEIN STEINHAUER (1943) studied Slavic and Baltic languages at the Universities of Amsterdam and Zagreb. Since his PhD dissertation on Croatian dialects (University of Amsterdam 1973), he has been working on Austronesian and Non-Austronesian languages of Indonesia. From 1988 to 1994 he was stationed at Pusat Bahasa, Jakarta, for the Indonesian Linguistics Development Project (ILDEP). Although retired he is still affiliated to Leiden University, and to Radboud University Nijmegen. He published extensively on Indonesian, on varieties of Malay, and on various other regional languages of Indonesia. Hein Steinhauer may be contacted at h.steinhauer@hum.leidenuniv.nl.

1 INTRODUCTION: ONE STATE ONE LANGUAGE?¹

With the former communist countries of Central Europe having joined the European Union, the number of its member states is approaching the number of provinces in Indonesia. Different from Indonesia, however, there is no *bahasa Eropa* which can function as:

lambang kebulatan semangat kebangsaan [Eropa] ..., alat penyatuan berbagai-bagai masyarakat yang berbeda latar belakang kebahasaan, kebudayaan, dan kesukuannya ke dalam satu masyarakat nasional [Eropa] ..., dan alat perhubungan antarsuku, antardaerah serta budaya (see Amran Halim 1976: 16).

On the level of the European nation states, the linguistic ecology today is a bit more transparent than in most Indonesian provinces. Each state has at least one undisputed national language, practically all with a documented history of several centuries or more. With regard to the European minority languages the situation is less stable, however. Not all of them have an acknowledged status. Language policy with respect to local vernaculars and languages of more recent immigrant communities differs from state to state and from government to government. In exceptional cases, such as in Spain with respect to Basque, Galician, and Catalan, suppression has turned into acceptance and even official recognition, but with the increase of mobility and long distance communication shift from regional minority languages to the national language is common and rarely counterbalanced by sufficient efforts of revitalization and preservation.

Less than two centuries ago the situation was different. One of the masterpieces of nineteenth century Russian literature, Tolstoy's "*Vojna i Mir*"² (War and Peace), begins as follows:

Eh bien, mon prince, Gênes et Lucques ne sont plus que des apanages, des *pomest'ja*, de la famille Buonaparte. Non, je vous préviens, que si vous ne me dites pas, que nous avons la guerre, si vous permettez encore de pallier toutes les infamies, toutes les atrocités de cet Antichrist (ma parole, j'y crois) – je ne vous connais plus, vous n'êtes plus mon ami, vous n'êtes plus *moj vernyj rab*, comme vous dites. *Nu, zdravstvujte, zdravstvujte*. Je vois que je vous fait peur. *Sadites' i rasskazyvajte*.³ (Tolstoy 1961: 7).

¹ Some sections in this paper have been presented earlier as a contribution to the Kolokium Antarabangsa "Bahasa dan Masyarakat Serantau", Bangi, Institut Alam dan Tamadun Melayu Universiti Kebangsaan Malaysia, 5 October 2000, entitled "Languages and States in Europe".

² Russian words and expressions have been romanized according to the International system of transliteration.

³ Well, prince, Genoa and Lucca are now no more than private estates, *pomest'ja* [country estates], of the Buonaparte family. No, I warn you, that if you do not tell me we are at war, if you allow yourself to palliate all the infamies and atrocities of this Antichrist (upon my word, I believe he is) – I do not know you in future, you are no longer my friend, no longer *moj vernyj rab* [my faithful slave] as you say. *Nu, zdravstvujte, zdravstvujte* [There, how do you do, how do you do]. I see I am scaring you. *Sadites' i rasskazyvajte* [sit down and tell me everything].

Some eighty words in French and only seven in Russian. That is how Anna Pavlovna Scherer, "*Fräulein* and close to the emperor's wife", greets Prince Vasilij, the first arriving guest to her soirée, the gist of her message being "if you dare to say we are not at war with that infamous Napoleon, I never want to see you again".

What the quotation shows is that at the end of the eighteenth and beginning of the nineteenth century French was the language of the elite and nobility all over Europe. With their exorbitant show-off of wealth and power, the "Sun king" Louis XIV and his successors to the French throne, had become the role model for everyone who thought him or herself to be superior to the common people.⁴

This eighteenth and early nineteenth century francomania did not prevent the Napoleonic wars. The belligerent sentiments of *Fräulein* Scherer may have been "religious", Napoleon's armed (lack of) diplomacy did raise nationalist feelings which had not been there before. His wars and their aftermath resulted in dramatical changes in the political map of Europe, and sowed the seeds of competition among the new European superpowers, which ultimately resulted in the colonialist division of the world among them, and in World War I.

One consequence of the Napoleonic wars was an increase in nationalist emotions and the rise of the conviction that a state could not exist without a (single) "national" language. This conviction is still very much alive, and consciously evoked and exploited by politicians, either or not populist. The fate of Serbo-Croatian is illustrative in this respect. In the nineteenth century, when Croatia was still Austrian and Serbia Turkish, two literary standards were developed. Differences between these varieties were considered insignificant enough⁵ for them to be jointly called Serbo-Croatian (or Croato-Serbian), as long as Croatia and Serbia were both part of the former Yugoslav Federation. The inhabitants of Bosnia-Herzegovina used either the Croatian or the Serbian standard. The whole area, Croatia, Serbia, Montenegro, and Bosnia-Herzegovina, was and still is a complex of (Serbo-Croatian) dialect islands and dialect continua. But when Yugoslavia fell apart Serbo-Croatian followed suit: Croatian and Serbian all at once were considered different languages, and even Bosnia-Herzegovina now has its own Bosnian language. Conscious efforts have since been made to create new lexical differences and to exaggerate old ones.

2 INVENTORY

As can be expected of a global area the size of the Nusantara archipelago, the

⁴ It is still the reason why to this day everyone who can afford it, has the - nearly moral - obligation to buy clothes, make-up, jewellery, and cigarette lighters from Paris-based fashion houses. It may be even the reason why we have so-called *haute-couture* at all.

⁵ The major difference between both varieties was visual, the Croatian standard using the Latin script, the Serbian one Cyrillic. Linguistically the differences consisted of different reflections of a frequent and an infrequent Proto-Slavonic phoneme, a different loan suffix for some infinitives, and a limited set of differing lexical items, among which the words for "train" and "bread" are shibboleths.

national and minority languages of Europe are divided into several unrelated language families. Apart from the *Indo-European* language family which as the name suggests is typically European, there are several other indigenous language families which are well represented, to wit Uralic, Altaic, and the Caucasian language families of which it is still uncertain whether and to what extent they are genetically related. These are Kartvelian (or South Caucasian), Abkhazo-Adyghe (or Northwest Caucasian), Dagestanian (or Northeast Caucasian) and Nakh (or North-Central Caucasian). One should add Semitic, since there are Assyrian communities scattered in Armenia and Georgia, whereas Maltese (with over 300,000 speakers) on the island of Malta is basically a variety of Arabic with an admixture of Italian. The only non-extinct indigenous isolate is Basque in northern Spain and adjacent areas of France with some 700,000 speakers.

Below I shall discuss these four multi-member language families indigenous to Europe at the middle of the twentieth century in somewhat more detail. They are presented as family trees in which branches and sub-branches are indicated numerically. My main sources have been Bray (1951), Comrie (1981), Lewis (2009, henceforth the *Ethnologue*), Stevenson (1989), and Vinogradov (1966a-c, 1967, 1968). Especially for the various lesser known languages of Eastern Europe different names or spellings may be found in the literature. Where my sources vary, I largely stick to Comrie's anglicised form of Russian names. Alternative names are listed in the *Ethnologue*.

Each family tree is followed by a short discussion about origin, spread and status of the individual languages, as well as some typical features of the family in question. Compared with the *Ethnologue* the surveys below are simplified. In some instances sub-branches at the lowest level have been united. Languages for special purposes have been left out. Extinct languages are only mentioned in the discussions. Especially for the Romance and Germanic branches of Indo-European the *Ethnologue* is rather liberal and elaborate in its assignment of language status to dialect groups. On the one hand this suggests a degree of discreteness between dialects which in the presence of the unifying influence of a standard language is no longer as obvious as it may once have been. The apparent effort of the editors of the *Ethnologue* to be exhaustive on the other hand underlines that they are not. For the Netherlands for instance the *Ethnologue* distinguishes besides (the Low Franconian) Dutch, Flemish, and (Western) Frisian, the following (Low Saxon) "languages": Achterhoeks, Drents, Gronings, Sallands, Stellingwerfs, Twents, Veluws, and (Rhenish Franconian) Limburgish. In as far as these names represent linguistic realities (the terms are geographical rather than linguistic), they can be considered dialects of Dutch, although a speaker of standard Dutch may on first hearing have some difficulty in understanding some of them. But more importantly, the *Ethnologue's* degree of detail appears to be arbitrary: it is limited to the northern provinces. Zeeuws for instance is subdivided into nine sub-dialects: Goerees, Flakkees, Schouws, Duvelands, Fluplands, Bevelands, Walchers, Axels and Kezands, whereas for a speaker

of standard Dutch the dialects of the southern province of Limburg and the many Flemish dialects in Belgium are by far the most divergent, but these are not further specified in the *Ethnologue*. Consequently I have economized with regard to this level of specification distinguishing only 116 languages. It should be stressed though that “[a]ll the major languages of Europe have dialects that are mutually incomprehensible without a considerable amount of experience. Americans tend to be baffled by Cockney English and completely floored by broad Lancashire or Scots. A Berliner is lost in the Bavarian Alps, as is a Parisian in Gascony, or a Piedmontese in Sicily” (Haugen 1966: 280).

It should be stressed that a survey like the present one cannot do justice to the dynamics in language use and language choice. All minority languages are under pressure from national and official languages. Some undergo revival, either by conscious efforts or as an effect of political and/or military turmoil, as in the former Yugoslavia. Data on the numbers of speakers, especially for minority languages are often based on estimates only, sometimes from decades ago. Figures from different sources may differ considerably, for one thing because gradation in linguistic proficiency may be evaluated differently, whereas avowed ethnicity may not necessarily imply mastery of the language involved. The figures for the numbers of speakers given below are largely based on or inferred from those given by the *Ethnologue*.

2.1 URALIC LANGUAGES

Finno-Ugric:

- 1 Ugric: HUNGARIAN (Hungary, Slovakia, Rumania, Serbia)
- 2 Non-Ugric:
 - 2.1 Permic: KOMI-PERMYAK, KOMI-ZYRYAN, UDMURT (all Russia)
 - 2.2 Non-Permic:
 - 2.2.1 Lappic: SAAMI (up to eight dialect groups/languages in Russia, Finland, Sweden and Norway)
 - 2.2.2 Balto-Finnic: ESTONIAN (Estonia), FINNISH (Finland), LIV(ONIAN) (Latvia); IZHORA, KARELIAN, VEPS, VOD (all Russia)
 - 2.2.3 Volgaic, Mordva: MOKSHA, ERZYA (both Russia)
Volgaic, Mari: HILL MARI, MEADOW-EASTERN MARI (both Russia)

The Uralic language family, the original homeland of which is thought to have been the area between the Volga river and the Ural mountains, consists of two major sub-groups: the Samoyedic (or Samodic) group and Finno-Ugric. The breaking up of the Finno-Ugric branch into different languages is estimated

⁶ The *Ethnologue* and Comrie have Ingrian for Izhora. The *Ethnologue* mentions Livvi (or Olonets) as another language, but according to Comrie and Soviet linguists it is too closely related to Karelian to be considered more than a Karelian dialect.

to have occurred some four to five thousand years ago. The split of Uralic into Proto-Finno-Ugric and Proto-Samoyedic must be located much earlier. Today there are only four Samoyedic languages left, all spoken in Siberia and all extremely endangered.

The Lappic (or Saami) dialect groups or languages of northern Scandinavia and the adjacent Koli peninsula in Russia may have been Samoyedic in origin, but they have been in contact with Balto-Finnic languages for a very long time and with such linguistic effects that they are now considered to form together with the latter a separate Finno-Lappic sub-branch of the Finno-Ugric languages.

The Balto-Finnic sub-group is the largest in number of languages and consists of two national languages, Finnish and Estonian, respectively with five million and more than one million speakers, and a number of minority languages along the eastern border of Finland and in the hinterland of St. Petersburg: Karelian has probably more than 40,000 speakers, Izhora less than 400. Veps may still have a few thousand speakers, Liv and Vod had already less than 500 speakers each half a century ago, and perhaps between 10 and 20 now. Both Finnish and Estonian have been used in writing for nearly five hundred years.

The Volgaic group consists of two pairs of languages which are rather distantly related to each other: Mordva on the one hand and Mari or Cheremissian on the other. Internally these two branches are each considered to consist of two languages: Moksha and Erzya for Mordva and Hill Mari and Meadow-Eastern Mari for Mari. Only one third of the probably still more than one million Mordva speakers live in their original homeland along the Volga river, south of Kazan', two thirds have moved since the fifteenth century into Siberia and Central Asia. The ancestors of the Mari languages (which still may have about half a million speakers) started to spread along the middle course of the Volga east of Novgorod around the end of the sixteenth century. Contacts with Russian and neighbouring Turkic languages (Chuvashian and Tatar) contributed to the deviating lexicon of the Mari languages.

The Permic languages, Udmurt (with perhaps still half a million speakers), and the closely related Komi-Zyryan, and Komi-Permyak (with respectively slightly over 200,000 and less than 100,000 speakers), are found - roughly speaking - west of the Ural mountains, southwest and north of the city of Perm. In the course of the last millennium Udmurt separated from Komi, and underwent lexical and some grammatical influence from neighbouring Turkic languages. Komi (or Old Permic) was the first Finno-Ugric language with a literary tradition (begun in the fourteenth century, declined after the seventeenth). Certain Komi influence can be observed in the lexicon of the Balto-Finnic languages Karelian and Veps, which dates back to the times when the ancestors of the speakers of these languages were not yet separated by Russian and still inhabited the basin of the Northern Dvina, southeast of Archangel'sk (ninth - fifteenth century).

Also Hungarian, with its thirteen million speakers (ten million in Hungary,

and three million in Slovakia, Serbia, Rumania, and other countries) the main language of the Ugric sub-group, must have been in contact with the Perm languages. The two languages closest related to Hungarian are the Ob'-Ugric languages Mansi (also known as Vogul(ian)) and Khanty (also known as Ostyak), both spoken in Northwest Siberia by only a few thousand speakers, and both characterized by marked dialectal differences. Under pressure from Turkic Siberian tribes the ancestors of the Hungarians had started their westward migration, ending up in the ninth century in the Danube planes which had become depopulated by the annihilation of the Avar⁷ kingdom (568-796) by Charlemagne, king/emperor of the Holy Roman Empire. The Avar, a Turkic tribe had subjugated the Slavic population, which had populated the area after the Germanic Gepids, East Goths, and Langobards. These had come in the slipstream of the Turkic Huns, who reigned in the area from 380 to 453. The Huns themselves were the successors of the Germanic Vandals and Iranian Alanians, who had defeated the Iranian Sarmatans, Yazygs and Roxolans. They in their turn had driven away the Celts (60 BC), who had subjugated the Thracian Ghetes (third century BC). By populating the area after all these predecessors the Hungarians drove a wedge between what later would become the West and South Slavic languages. In the course of history the Hungarian kingdom expanded gradually, to be reduced to one third of its former size after World War I, leaving one third of its speakers in the surrounding countries.

With the exception of the present national languages, Finnish, Estonian, and Hungarian, all the Uralic languages have a bleak future. They rarely live in areas where they are not outnumbered by speakers of Russian or other languages. Bilingualism is the rule and language shift is common.

Widespread characteristics of the Finno-Ugric languages are an elaborate case system (fourteen in Estonian for instance), (various patterns of) vowel harmony, (original) agglutination, which in some languages has developed into a more fusional morphology (such as in Lappic and Estonian). SVO structure is predominant in the western languages, SOV in the eastern ones, the latter probably through contact with Turkic languages. Typologically remarkable is the three-way opposition in length (short, long, extra long), both for vowels and consonants in Estonian.

2.2 CAUCASIAN LANGUAGE FAMILIES

The Caucasian languages belong to four distinct families, namely:

South Caucasian (Kartvelian):

- 1 GEORGIAN (Georgia)
- 2 SVAN (Georgia)
- 3 LAZ (Georgia, Asiatic Turkey), MINGRELIAN (Georgia)

⁷ Not to be confused with the Caucasian Avar (see the next section).

North Central (Nakh):

- 1 BATS (Georgia)
- 2 CHECHEN, INGUSH (both Russia)

Northeast (Dagestanian):

- 1 AVARO-ANDO-DIDO
 - 1.1 AKHVAKH, ANDI, BOTLIKH, CHAMALAL, GODOBERI, KARATA, BAGVAL, TINDI (all Russia)
 - 1.2 AVAR (Russia)
 - 1.3 DIDO, GINUKH, GUNZIB, BEZHTI, KHVARSH (all Russia)
- 2 DARGVA, LAK (both Russia)
- 3 AGUL, ARCHI, LEZGI, RUTUL, TABASARAN (all Russia), BUDUKH, KHINALUG, KRYZ, TSAKHUR, UDI (all Azerbaijan)

Northwest:

- 1 ABKHAZ (Georgia), ABAZA (Russia)
- 2 Circassian: ADYGHE, KABARDIAN (both Russia)

If the present Caucasian (also called Ibero-Caucasian)⁸ languages in the south-eastern frontier area between Europe and Asia are genetically related at all, they must have separated into different branches such a long time ago that convincing evidence of their relationship can no longer be presented. The efforts to relate the Caucasian languages to other language families such as Indo-European are based on even less solid evidence. The search for correspondences with isolates like Basque (in Spain and France) or Burushaski (in Afghanistan) have not yielded any results. It must be assumed that the ancestors of the Caucasians have remained relatively untouched by the streams of migrating Indo-Europeans, sheltered as they were by the inaccessibility of the mountains between the Caspian and the Black Sea. Yet, in the course of history some non-Caucasian languages did become indigenous in the area: the Iranian Ossetian, Tat, Kurdish, and Talysh, the Semitic Assyrian, and the Turkic Azeri, Karachay-Balkar, Nogay, and Kumyk. The southward Russian expansion to the detriment of the Ottoman empire, however, has been the most disruptive: the subjugation of the Caucasian peoples in the course of the nineteenth century led to massive migrations in the area and to the exodus of about 50% of the Northwest Caucasians to Asiatic Turkey (among them the complete Ubykh population). This is the origin of Northwest Caucasian communities in Jordan and Syria (which at the time were still part of the Ottoman Empire). Many of the Circassians who migrated to Anatolia became victim of the Atatürk revolution: they had been too long too loyal to the sultan, and their neutrality vis-à-vis the Anatolian Greek in the Greek-Turkish war

⁸ The term Ibero- refers to the South Caucasian area and should not suggest any relation with the Iberic peninsula, nor with the Basque isolate there.

of 1920-1921 was interpreted as treason. Those Circassians who survived the repercussions have gradually assimilated and shifted to Turkish. Only the oldest generation in rural districts may still be bilingual. The last Ubykh speaker was 80 years old in the early 1980s so the language is now extinct. The Northwest Caucasians who did not migrate to Turkey now live in a number of scattered concentrations separated by Russians. Whereas the unofficial language policy in modern Turkey has often been downright suppressive, the soviet policy towards most minorities was the opposite: orthographies were developed, the languages were used in written and spoken media and in education. Because of the dispersion of the Northwest Caucasian language varieties, which could have been considered dialects of each other (such as Abkhaz and Abaza, and Adyghe and Kabardian), by this policy became separate literary languages.

The differences with and among the other Caucasian languages, however, are real and extremely great. Yet, some general typological observations can be made. Phonologically the Caucasian languages are characterized by their elaborate systems of consonants and consonant clusters. Georgian, for instance, may construct words with word-initial clusters of up to eight consonants.⁹ Ubykh belonged to the small group of languages with more than eighty consonant phonemes: the set of uvular phonemes alone (totalling twenty phonemes) consists of voiceless and ejective stops and voiced and voiceless fricatives, which can all be plain, palatalised, labialised, pharyngealized, or labialized *and* pharyngealized. The complicated consonantal system of especially the Northwest Caucasian languages is matched by a minimum number of vowels. Kuipers proposed an analysis of Kabardian according to which all (phonetic) vowels are predictable realizations of features of the preceding string of one or more consonants, resulting in a pattern of a language without phonemic vowels (Kuipers 1960). Others prefer a bi-vocalic analysis (Smeets 1984: 80; see also Kuipers 1976 for references and comments).

Morphologically the Caucasian languages vary from relatively transparent to extremely complicated. The following one-word sentences are typical for Northwest Caucasian word structure (see Comrie 1981: 219 for Abkhaz; Smeets 1984: 69 for Circassian):

Abkhaz: *j* *-aa* *-sə* *-zə́* *-m* *-k'ə* *-jt'*
 3SG.INAN-hither-1SG(IND.OBJ)-POT-NEG-seize-FIN¹⁰
 'I couldn't get hold of it'

Circassian: *sə* *-qə* *-b* *-de* *-k'°e* *-žə* *-š°ə* *-št* *-ep*
 1SG.SBJ-hither-2SG.OBJ-together.with-go-as.previously-FUT-NEG
 'I will not be able to come back together with you'

⁹ Comrie gives the example *goprckonis* 'he peels us' (Comrie 1981: 200).

¹⁰ Abbreviations used are 1, 3=first, third person, FIN=finite, FUT=future tense, IND=indirect, OBJ=object, POT=potential, SBJ=subject, SG=singular. A consonant followed by an apostrophe is glottalic, a consonant followed by a raised ° is labialized; š is a voiceless alveolo-palatal fricative.

Most Northeast and North Central Caucasian languages have nominal classes from two in northern Tabasaran to eight in Bats. For those languages with the same number of classes the distribution of nominal roots to the various classes may differ. Case systems are characteristic of South Caucasian languages, where Georgian has six basic cases and eleven secondary (derived) ones. The system in the Northwest Caucasian languages is restricted (perhaps only two in Abkhaz/Abaza), but the other North Caucasian languages are extremely rich. Literary Tabasaran for instance has four basic cases and forty two derived ones (all locative) (see Khanmagomedov 1967: 549-550). Verbs are everywhere the most complicated. The examples from Abkhaz and Circassian illustrate the tendency to copy information in the verb through pronominal and locational affixes which may at the same time be expressed in nominal arguments of the verb. These examples also show that morphemes in these Northwest Caucasian languages tend to be consonantal or monosyllabic. As a consequence morphological homonymy and compounding are frequent, sometimes resulting in patterns not unfamiliar to speakers of Malay. From the Circassian roots *ne* 'eye', *psə* 'water', *pe* 'nose', for instance, the following compounds are formed: *ne.psə* 'tear', *psə.ne* 'well, fountain', *ne.pe* 'face', *pe.psə* 'mucus' (see Smeets 1984: 66).

With the exception of Mingrelian all Caucasian languages have some form of ergative syntax. The choice for ergative or some other construction is subject to different restrictions in the various languages, dependent for instance on the semantic subclass of the verb, and/or its particular tense/aspect.

The future of several of the Caucasian languages is uncertain. Only twelve have the status of literary language: Georgian (with some four million speakers), Abkhaz (over 100,000), Abaza (circa 40,000), Kabardian (about 500,000), Adyghe (circa 125,000), Avar (well over 700,000), Lezgi (close to 400,000), Lak (some 150,000), Dargva (about 500,000), Chechen (some 1,600,000), Ingush (over 400,000), Tabasaran (over 130,000).¹¹ All other, non-literary languages had twenty five years ago only a few thousand or even only a few hundred speakers, with the sole exception of Mingrelian (more than 300,000 speakers). Avar is used as the literary language by many of the Dagestani linguistic minorities, and so is Azeri in Azerbaijan. The Mingrelians have embraced Georgian, now the national language of the Georgian republic.

Ethnic tensions in the area are high. Georgia has to struggle with two break-away regions: Abkhazia and South Ossetia. Armenia and Azerbaijan have been at war over each other's enclaves, Nakhichevan and Gornyj Karabakh. The Russian-Chechen war with its terrorist attacks and face saving bombardments and its echoes in neighbouring Dagestan and Ingushetia have not resulted in a tranquil state of peace. Some of these conflicts have been accompanied by actions of ethnic cleansing, in any case by large numbers of refugees. Linguistic consequences may be a matter of time, also in the case

¹¹ These figures for number of speakers are those given by the *Ethnologue*. They are considerably higher than the very rough estimations mentioned by Smeets (1984: 40-45) and Comrie (1981: 197), which were based on figures which were already estimations 25 years ago..

of the minority languages whose speakers are bilingual and/or switching to the language of the local majority.

2.3 ALTAIC LANGUAGES

- 1 Mongolian: KALMYK-OIRAT (Russia)
- 2 Turkic:
 - 2.1 CHUVASH (Russia)
 - 2.2 Southern Turkic: AZERI (Azerbaijan)
GAGAUZ (Moldova), TURKISH (Turkey, Cyprus)
 - 2.3 Kipchak:
 - 2.3.1 NOGAI (Russia)
 - 2.3.2 Ponto-Caspian: CRIMEAN TATAR (Ukraine), KARAIM (Lithuania, Ukraine) KARACHAY-BALKAR (Russia); KUMYK (Russia)
 - 2.3.3 Uralian: BASHKIR (Russia); TATAR (Russia)

The Altaic language family is probably a chain of three separate families: Tungusic (only in Asia), Mongolian (mainly in Asia), and Turkic (in Asia and Europe). The apparent links among these languages may be a result of contact rather than genetic relationship.

Mongolian and Turkic tribes, whether or not under Mongolian command, have a long history of invasions from their homeland along the north and north-western borders of China into central and western Asia and Europe. The first documented raids were those of the Huns. In the second century AD when their country was pre-emptively attacked by the Chinese they started their westward migration. With their superior horsemanship and belligerent disposition they had an ineffective way of conquering the hearts and minds of the peoples they encountered. Yet, by their diplomatic skills they also made allies. The combination of both capacities made them the catalyst of what came to be known in European history as the “wandering of nations” (fifth - sixth centuries AD).

Neither the Huns nor the Turkic Avars (see the paragraph about Hungarian) survived in Europe as an identifiable group with their own language. The same holds for the Turkic Bulgars, who established a kingdom in the sixth century along the northern borders of Greece, but who finally mixed with the local Slavic population, which had replaced the original Thracians, and whose (purely Slavic) language is now known as Bulgarian.

A more lasting linguistic effect had the invasions of the Mongolians under Genghis Khan in the thirteenth century. At the height of their power their realm reached from China to the Danube river, but it soon fell apart under Genghis Khan’s successors. The southern and eastern parts of European Russia became the homeland of the Golden Horde, to which the Russian princedoms of Moscow and surroundings had to pay tribute. Only towards the end of the Middle Ages the Russians succeeded to free themselves from this “Tatar jug”.

But the Tatar rule had lasted long enough to enrich the European cultural heritage with several Turkic languages. The final conquest by the Russians of the Turkic khanates of Kazan' and Astrakhan in the sixteenth century opened the way to the Russian expansion into Siberia and the Caucasus. At the same time it created a region around Astrakhan and the lower current of the Volga river which was sufficiently depopulated for the Mongolian Kalmyk (who at that time were the most powerful tribe in Mongolia) to move in.

From the thirteenth century onwards Osman Turkic tribes originating from Turkmenia started their gradual westward movement, meeting and conquering Iranian, Caucasian, Armenian, Germanic, Greek, and Slavic peoples on their way. In the middle of the fifteenth century they conquered the remnants of the East Roman empire, but were stopped at the walls of Vienna. At the height of their power (in the sixteenth and seventeenth century) their realm reached from Pakistan to Morocco, with Istanbul (the former Byzantium/Constantinople) as their capital and Turkey and the Balkan Peninsula as the core area of their empire. This Turkish Empire shrank to its present proportions in the course of the nineteenth and the beginning of the twentieth century as a result of successful nationalist movements (Greece, Bulgaria, Serbia, and Rumania), internal weakness, and the imperial expansion of the European powers of the time.

Two major steps towards a mono-ethnic Turkish state were the slaughter of 1.5 million Armenians (1915) and the treaty of Lausanne of 1923, which defined the borders between Greece and Turkey that had been contested in the Greek-Turkish war of 1920-1921. The result was a massive population exchange: 1.5 million ethnic Greek were moved from Asian Turkey (Anatolia) to Greece, while 0.8 million Turks moved from northern Greece to European Turkey.¹²

The result of the Turkic migrations since the late Middle Ages is a mosaic of Turkic languages, stretching from China to the Balkan peninsula, with a lot of linguistic pockets in the European part of the former Soviet Union. Only in few cases sharp borders can be drawn between these languages, due to relative isolation for an extended period and/or contact with non-Turkic languages. In Europe Chuvash (with about 1.5 million speakers surrounded by speakers of Russian and the Uralic Mordva) is such an isolated language. The lack of sharp borders and intensive contacts throughout history between most of the other Turkic languages are the reason why the *Stammbaum* model as presented above is only a very impressionistic illustration of the linguistic and historical relationships and contacts.

The most important Turkic language is obviously Turkish, the national language of Turkey, with well over sixty million speakers in Turkey, about

¹² The four million Kurds in East Turkey have been physically and culturally suppressed for years. Use of their (Iranian) language in writing used to be a punishable offence. Only recently under European pressure Turkish policy has become more lenient. But tensions remain. It may be expected that future developments are closely connected with developments in Iraq and Syria.

0.8 million in Bulgaria, and sizable groups in other Balkan countries. Major Turkic languages outside Europe are Uzbek, Kazakh, Turkmen, Kirgizian, and Uygur, each with millions of speakers. Within Europe, Azeri (the national language of Azerbaijan with some six million speakers of which four million monolingual, and an estimated additional three million living in Iran) is the most important. Also widely used in writing and education (though not a national language) is (Volga) Tatar with over six million speakers spread over a wide area in European Russia and Siberia, but with concentrations in the Tatar and Bashkir Autonomous Republics, between Kazan' and the Southern Ural mountains.

Minor Turkic languages are: Bashkir, closely related to Tatar, with about 1.400,000 speakers a minority even in their own autonomous republic; Nogay, with less than 90,000 speakers in Dagestan and neighbouring regions of Russia; Karachay-Balkar, spoken by two related but disjunctive ethnic groups (some 230,000 speakers in total) in the northern Caucasus, where they are outnumbered by Adyghe and Kabardians; Kumyk with perhaps 400,000 speakers one of the written languages of Dagestan; Gagauz, closely related to Turkish, with some 160.000 speakers in Moldavia and the neighbouring regions of the Ukraine, used in writing only since the 1950s.

Of two further (originally) European Turkic languages the present fate is uncertain. Karaim was (once?) spoken by an ethnic group in the Crimea peninsula which adopted Judaism in the past. In the Crimea the original speakers have shifted to Crimean Tatar. More than a generation ago there may still have been between 500 and 1000 speakers in Lithuania and the Ukraine, to where Karaim speakers had migrated some 600 years ago. Their Karaim varieties are close to extinction now with less than 300 speakers in total.

Crimean Tatar was originally spoken (and written) in the Crimea, and maybe still is, albeit by an unknown number of speakers. At the beginning of World War II the whole Crimean Tatar population was forced to leave their homeland and migrate to Uzbekistan, because Stalin suspected them of sympathy for the Germans. Unlike the Caucasian Chechens who underwent the same fate, the Crimean Tatars were not allowed to return to their home country until after the fall of the Soviet Union; according to the *Ethnologue* there must be sizable communities of Crimean Tatars in Turkey, Uzbekistan, Kyrgyzstan, European Russia, Rumania, and Bulgaria.

Typologically the most conspicuous feature of the Turkic (and to some lesser extent the Mongolian) languages is the highly transparent agglutinative structure with suffixes (both derivational and inflectional), combined with a system of progressive vowel harmony, according to which each non-initial vowel assimilates to the previous one in frontness and roundedness. Foreign vocabulary (from Persian and Arabic, and later from Russian), sound-changes, and concomitant articulatory features have caused changes and deviations in this system in individual languages, the most radical being those of Gagauz and Karaim. In these languages vowel harmony was replaced by consonant harmony (all consonants in a word becoming either palatal or plain).

Characteristic syntactic features of the Turkic languages are preposed adjectives and preposed verbal arguments, parataxis or the use of gerunds and participles as the equivalents of complex sentences in other languages.

2.4 INDO-EUROPEAN LANGUAGES

- 1 Indo-Iranian:
 - 1.1 Indo-Aryan, Central Zone:

ROMANY varieties (Balkan, Central Europe, other European countries);
 - 1.2 Iranian: OSSETIAN (Georgia, Russia)

KURDISH (Armenia, Georgia, Azerbaijan), TALYSH (Azerbaijan)

TAT (Azerbaijan, Russia, Armenia)
- 2 Greek: GREEK (Greece, Cyprus, Albania, Ukraine)
- 3 Thraco-phrygian: ARMENIAN (Armenia, Azerbaijan, Georgia)
- 4 Albanian: ALBANIAN (Albania, Serbia, Kosovo, Macedonia, Greece)
- 5 Baltic: LATVIAN (Latvia), LITHUANIAN (Lithuania)
- 6 Slavonic:
 - 6.1 South: BULGARIAN (Bulgaria), MACEDONIAN (Macedonia)

SLOVENIAN (Slovenia), SERBO-CROATIAN (Croatia, Bosnia Hercegovina, Serbia, Montenegro)
 - 6.2 West: LOWER SORBIAN, UPPER SORBIAN (both Germany)

CZECH (Czech Republic), SLOVAK (Slovakia)

POLISH (Poland)
 - 6.3 East: BELARUSIAN (Belarus), RUSSIAN (Russia), UKRAINIAN (Ukraine)
- 7 Celtic:
 - 7.1 British: BRETON (France), WELSH (United Kingdom)
 - 7.2 Gaelic: IRISH GAELIC (Ireland), SCOTS GAELIC (United Kingdom)
- 8 Italic:
 - 8.1 East: RUMANIAN (Rumania), MOLDAVIAN (Moldavia);
 - 8.2 West:
 - 8.2.1 Central section:
 - 8.2.1.1 ISTRO-ROMANCE (Croatia), ITALIAN (Italy, Switzerland)
 - 8.2.1.2 CORSICAN (France), SARDINIAN dialects (Italy)
 - 8.2.1.3 FRIULIAN, LADINO (both Italy), ROMANSH
 - 8.2.2 Western section:
 - 8.2.2.1 SPANISH, GALICIAN (both Spain), PORTUGUESE (Portugal)
 - 8.2.2.2 CATALAN (Spain), PROVENÇAL (France)
 - 8.2.2.3 FRENCH (France, Belgium, Switzerland)

9 Germanic:

- 9.1 North: DANISH (Denmark), NORWEGIAN BOKMÅL (Norway), SWEDISH (Sweden, Finland) ICELANDIC (Iceland), FAROESE (Faroe), NORWEGIAN NYNORSK (Norway),
- 9.2 West: ENGLISH (United Kingdom, Ireland), FRISIAN (Netherlands, Germany) GERMAN (Germany, Austria, Switzerland, Italy), DUTCH (Netherlands, Belgium, France)

The original homeland of the Indo-Europeans is thought to be the plains north of the Caucasus and Central Asia. From there they spread, presumably some 5000 years ago, in all directions, an expansion which was made possible by the domestication of the horse and the invention of the wheel. Human presence in Europe was as is known much older, but apart from a few non-Indo-European roots in some languages and some suspect geographical names there are no linguistic traces left, although Basque is sometimes presented as a candidate. In the course of later history several branches of Indo-European became extinct as well. Two of them are well-documented thanks to preserved libraries: Hittite (East Turkey, disappeared circa 1000 BC) and Tocharian (West China, extinct circa 1000 AD).

All Indo-European languages show different developments of a system of mainly suffixing fusional morphology, nominal inflection with three numbers (singular, dual, and plural) and up to seven cases, and a complicated inflectional verbal paradigm (distinguishing voices, moods, tenses, person and number).

In the following section the various Indo-European subfamilies will be discussed one by one, in as far as they are still indigenous and spoken in Europe today.

2.4.1 INDO-IRANIAN

Most of the early Indo-Iranians moved to the southeast into Persia, Afghanistan, Pakistan and Northern India, that is out of Europe. Some of the Iranian tribes had (later?) moved west, such as the nomadic Scythians whose impressive material culture was at its height from the seventh to the fourth century BC, when they ruled the planes between the Caspian Sea and the lower reaches of the Danube river. Linguistically they left few traces. The same holds for most of their Iranian successors, the various Sarmatian tribes, the only exception being the Alanians, who became the forefathers of the Ossetians (see below).

The only European language belonging to the Indian sub-branch of the Indo-Iranian languages is Romany, the language of the Gypsies. It is assumed that they left northwest and/or central India in the course of the first millennium in several waves. Living as wanderers they moved to nearly all countries of Europe, and from there to America and Australia. Many groups are still living as wanderers, others have settled. Some of the

latter have assimilated (such as those in southern Spain), most are bilingual in their own language and the language in the area they have lived in for some longer period. The many Romany varieties (compare Comrie 1981: 158 and the *Ethnologue* for different lists) are distinguished mainly on the basis of lexical influence from the languages of their host countries. The exact number of Romany speakers is unknown. Large communities are found in most of the former East bloc countries, where many of them live in ghetto-like circumstances and are subject to serious discrimination. In its genocidal effort to free the German Reich from all “inferior forms of human life” the Hitler regime systematically slaughtered more than half a million gypsies in their extermination camps.

Most of the Iranian languages are spoken outside Europe: Persian (in Iran) with its ancient predecessors Old Persian and Avestan, Tajik (in Tajikistan), Pashto (in Afghanistan), Baluchian (in Iran and Pakistan), and many smaller languages (Iran, Afghanistan, Central Asia). The major indigenous Iranian language in Europe today is Ossetian spoken by probably more than half a million speakers in Georgia and neighbouring Russia. It has a written tradition (in Cyrillic script) since the middle of the nineteenth century. Genetically it is classified as a North-East Iranian language, although it is spoken in the far west.

Kurdish is spoken in a number of villages in Armenia, Azerbaijan (Nakhichevan), and Georgia by some 80,000 speakers, although the vast majority of Kurds (maybe over ten million) live in Asia (eastern Turkey, northern Iraq, and the neighbouring regions of Syria and Iran).¹³ In Soviet Armenia local Kurdish used to be written originally in an Armenian-based script, but after World War II in a variety of Cyrillic. Talysh, on both sides of the southeast border of Azerbaijan and Iran, belongs with Kurdish to the North-West Iranian languages. It used to be spoken by less than 100,000 speakers in both countries; bilingualism with Azeri was already common among those in Azerbaijan two generations ago; its present fate is uncertain. Tat, finally, is the language of the descendants of military colonies from south-western Persia, which defended the borders of the Sassanid Empire. It is genetically close to Persian and Tajik and spoken in various pockets throughout the Caucasus (Azerbaijan, Dagestan, Armenia). Dialect differentiation, which corresponds with region and religion, is considerable: Jewish Tat live in North Azerbaijan and neighbouring Dagestan, Muslim Tat in East Azerbaijan, and Christian Tat in Armenia. Like with Talysh, bilingualism with Azeri in Azerbaijan used to be common. The number of speakers today may be less than 15,000.

2.4.2 GREEK

The Greek moved to what is now mainland Greece some 1,500 years BC, and from there to the islands in the Aegean Sea, to Crete, and Cyprus, and

¹³ In many countries of western Europe there are sizable communities of Kurdish refugees who have fled from the Turkish oppression and the anti-Kurd atrocities by the Saddam Hussein regime in Iraq after the first Gulf war.

to the west and north coasts of Asiatic Turkey. Besides, they established colonies along the north coast of the Black Sea, in Sicily and southern Italy, and further north in Naples (Italy) and Marseille (France). A last wave of Greek immigrants into the main land (around 1000 BC) destroyed the early Mycenaean civilization. The "classical" period in Greek civilization which reached its peak in the fifth and fourth century BC, resulted in the cultural hegemony of Athens, with the linguistic effect that the three or four dialect groups which had existed until then converged into a *koine*. In the fourth century BC Alexander the Great (Iskandar Zulkarnain) who started as a king of Macedonia, unified for the first time in history the whole of Greece. He subsequently conquered the Persian Empire and the lands beyond. Moreover, although his own empire did not last long after his early death, the linguistic result of his conquests was that koine Greek became *the* language of culture and learning around the eastern Mediterranean, including Egypt.

The alphabet of the classical language, which was modelled in the ninth century BC after the old Phoenician (Palestinian) alphabet and which had replaced the former syllabic "linear-B" script of the Mycenaean civilization, subsequently became the model of the Cyrillic and Latin alphabets we use today. Koine Greek became the language of the Christian New Testament and early preaching. It was the official language of the East Roman and Byzantine Empire, and developed eventually - not without changes - into "high" literary Greek. This language (called *katharevousa*) which is still preferred for legal and academic writing is for other purposes in competition with the so-called demotic variant of Greek (*demotiki*), which is much closer to the actual spoken language.

Outside Greece Greek is spoken in southern Cyprus, in some villages in southern Albania, and - in southern Ukraine - by the descendents of early nineteenth century refugees from the Ottoman Empire. Since the fall of the iron curtain back-migration has diminished their number (once exceeding 100,000 speakers). During the Ottoman Empire large numbers of Greek lived in Asiatic Turkey, and a somewhat smaller number of Turks in northern Greece. After World War I the Greek tried to make use of the instability of Turkey during the kemalist revolution and shift the Greek-Turkish border further into Anatolia. As mentioned above, their defeat resulted in the massive population exchange over the definitely established borders between the two countries. The more recent division of Cyprus in a northern (Turkish) and southern (Greek) part is the last stage in this process of ethnic purification.

2.4.3 ARMENIAN

Alexander the Great received a Greek education, but his native language may have been Thracian. No traces are left, however, of the Thraco-phrygian tribes who once must have settled in geographic Macedonia (that is northern Greece and the former Yugoslav republic). The Phrygians seem to have moved on from there to the northwest corner of Asiatic Turkey. The only survivors of these tribes today are the Armenians, in the republic of Armenia (well over

three million speakers), and adjacent areas of Azerbaijan, Georgia and Turkey. The genocidal slaughter of Armenians by the Turkish armies in the second decade of the twentieth century¹⁴ caused an exodus of Armenians (which had already begun in the seventeenth century when their country was divided between the Persian and Turkish empires) resulting in substantial Armenian communities in the Middle East and elsewhere in the world. The total number of Armenians is estimated at more than six million.

Armenian has a practically uninterrupted written history since the beginning of the fifth century, when Mesrop Maštoc developed the *grabar* script which is still used today. The differences between Classical Armenian (fifth - eleventh century) and modern Armenian, the national language of the Armenian republic, are considerable. Dialectal differentiation is great.

Armenian deviates from other Indo-European languages by its lack of grammatical gender. Verb-final word order and its agglutinative rather than fusional morphology are ascribed to Turkic influence. Verbal inflection has largely been replaced by analytic constructions.

2.4.4 ALBANIAN

Next to Greece and Armenian, Albanian is the third European branch of Indo-European which consists of a single language.

The ancestors of the Albanians must once have settled in the mountains separating Poland from Slovakia and the Czech Republic. The (non-Slavonic) local names of these mountains (*Karpaty*) and of the neighbouring mountain range (*Beskidy*) are obviously of Albanian origin, the Albanian words for 'rock' and 'high mountain' being *karpë* and *bjeshkë*. The fact that the Albanian lexicon lacks non-borrowed words for concepts related to the sea suggests also that they must have moved to the coast of the Adriatic Sea, only in relatively recent times. It is not certain therefore that they are the surviving descendants of the enigmatic Illyrians, who according to Roman sources inhabited the Dalmatian coasts at the beginning of the Christian era.

Today Albanian is spoken in Albania and adjacent areas: Kosovo (where it is finally thriving thanks to the active and obviously counterproductive oppression by the Milošević regime in the 1980s and early 1990s, only resulting in its independence), northern Greece (where it is less thriving as a consequence of the more passive oppression by the Greek), and Macedonia (where tensions with the Slavic majority have been high and the official policy towards the Albanese is at best ambivalent). A few villages in southern Italy (inhabited by over 200,000 descendants of fifteenth century Albanian mercenaries for the kingdom of Naples and refugees from the Turkish invasion of Albania) still maintain their variety of southern Albanese called Arbëresh. In Albania and adjacent areas two dialect groups are usually distinguished: the northern dialects or Gheg (about two million speakers in northern Albania, Kosovo and Macedonia), and the more homogeneous southern ones or Tosk (over

¹⁴ The number of victims – an estimated 1.5 million – as well as the fact of the genocide is contested by Turkish historians.

three million speakers in southern Albania and northern Greece). Albania is written since the sixteenth century. The national language of Albania is based on the Tosk dialect.

Together with Greece, Macedonian, Bulgarian, East Serbian dialects, and Rumanian, Albanian belongs to the Balkan *Sprachbund*, typical features of which are the development of post-nominal articles, simplification/abolishment of the old case system, replacement of the old inflectional future tense with an analytical construction with the verb 'to want' ('he wants that he goes'), etcetera.

2.4.5 BALTIC LANGUAGES

From the Indo-European homeland the Baltic tribes moved to the northwest, where they came into contact with speakers of Balto-Finnic languages. There may have been a Balto-Slavic proto-language, but most observers now assume that similarities with the Slavonic languages are due to a long history of contact instead of to a direct common ancestor. Several Baltic languages must once have been spoken east of the current Baltic states, but they have disappeared without leaving much more than topographic evidence. In the West, Old Prussian has become extinct centuries ago as a result of German expansion; it is known thanks to a number of religious texts from the seventeenth century.

The surviving Baltic languages, Latvian (with over 1.5 million speakers) and Lithuanian (with over three million speakers) are the national languages of Latvia and Lithuania respectively. The oldest written texts in both languages date from the sixteenth century. The modern literary languages of both Lithuania and Latvia developed towards the end of the nineteenth century. Literary Lithuanian is based on the southern dialects, literary Latvian on the dialects of central Latvia.

Both Latvian and Lithuanian are tone languages: standard Latvian has three phonemic tones on all long syllables (that is syllables with a long vowel, a diphthong or a vowel + liquid nucleus), Lithuanian distinguishes two tones on stressed long syllables only (some dialects, however, seem to have up to five different tones). Lithuanian is identified as the most archaic Indo-European language, especially in its nominal morphology (case system, gender, number, declension types). Latvian has undergone innovations, obviously as a result of contact with German and with Finno-Baltic Liv(onian), which today is moribund, but which once covered a much larger area.

2.4.6 SLAVONIC LANGUAGES

At first the Slavonic (or Slavic) tribes hardly moved. They had settled east of the Carpathian mountains, south of the Baltic tribes, and awaited their time. That time came when the Huns invaded Europe, Germanic tribes moved westwards, and the West Roman empire ended. Some of the resulting open spaces in the West were filled up by Slavonic tribes. Linguistically this western group may have constituted a dialect continuum which in the ninth century was split up by the invasion of the Hungarians into what subsequently

developed as the West Slavonic and South Slavonic languages. Ancestors of this southern group had penetrated the Balkan Peninsula. Some geographical names show that they once even reached southern Greece and Asiatic Turkey. By populating present day Bulgaria and former Yugoslavia they separated the Rumanians from the rest of the Romance world.

Characteristic of all Slavonic languages is the elaborate system of aspectual oppositions (perfective vs. imperfective) in the verbal system, expressed rather unpredictable through prefixation, suffixation, and occasionally suppletion.

The oldest written Slavonic language is Old Bulgarian, the oldest manuscripts of which go back to the second half of the ninth century. It was used in ecclesiastical texts and contexts by the "Slav apostles", SS. Methodius and Cyril and their pupils. S. Cyril was the creator of the Cyrillic script and possibly also of the Glagolitic script which was used for some time in Croatia. What is known as Old Church Slavonic, which is still in vigorous use in the orthodox liturgy from Russia to Serbia, is a more Russified version of this Old Bulgarian.

Modern Bulgarian developed since the sixteenth century. The gap which arose between this older written language and the spoken varieties of Bulgarian was bridged by a number of prominent authors from eastern Bulgaria towards the end of the nineteenth century, resulting in the literary standard and national language of today.

The region where Slavonic was used in writing for the first time, Macedonia the homeland of SS. Methodius and Cyril, was the latest to have its language recognized as a (literary) language. When the Yugoslav Federal Republic was proclaimed by the partisans of Tito who fought the Nazi German occupation and their Croatian puppet regime, Macedonia became one of the constituting republics with its own language (2 August 1944). Thanks to a number of gifted poets the central dialect became the basis of the new Macedonian literary language, which now is the national language of the (former Yugoslav) republic of Macedonia with some two million speakers. Macedonian dialects are also spoken in northern Greece and in a number of villages in Albania and Bulgaria.¹⁵

Both Macedonian and Bulgarian show features typical of the Balkan *Sprachbund* (see the previous paragraph). Macedonian has the typological peculiarity of a three-way deictic opposition expressed both in its demonstratives and in its post-nominal definite articles.

The westernmost South Slavonic language is Slovenian, today the national language of independent Slovenia and spoken by over two million speakers in different dialects. Its oldest source dates from the eleventh century. A flourishing Protestant literature was successfully suppressed in the sixteenth century by the Austrian Catholic Counter Reformation. The modern literary

¹⁵ Bulgarian linguists and politicians have always claimed that Macedonian is a dialect of Bulgarian. Serbian linguists on the other hand have selected "Serbian" isoglosses which would turn most of Macedonian into dialects of Serbian. See Friedman (2000) for the extremely complicated relation between language, ethnicity and political claims in Macedonia.

language developed in the course of the nineteenth century. Slovenian has preserved some archaic features such as a dual (lost in nearly all Slavonic languages), and distinctive tones in stressed syllables (elsewhere only preserved in Serbo-Croatian). Slovenian dialects are also spoken in some villages at the other side of the borders with Austria and also in Italy, where it has the status of "official regional language". There is no sharp linguistic border with Croatia(n).

As described above, the Serbo-Croatian dialect complex between Slovenian and Macedonian has been artificially divided into three national languages after the Yugoslav civil wars. Most Croatian dialects are much more different from standard Croatian than the latter is from standard Serbian or Bosnian. To a somewhat lesser extent this is also true for Serbian dialects vis-à-vis standard Serbian. Together with the other South Slavonic languages Serbo-Croatian has preserved much of the elaborate system of morphologically expressed verbal tenses and moods. The estimated number of speakers of all Serbo-Croatian dialects/languages in the former Yugoslav countries is 18 million (over ten million in Serbia, Montenegro, 5.5 million in Croatia, and 2.5 million in Bosnia-Herzegovina).

The expansion of the West Slavonic tribes to the West once reached the Elbe river in Germany. A Slavonic language, Polabian, was still spoken in the regions of Hamburg in the seventeenth century. The westernmost Slavonic communities today are two enclaves of speakers of Sorbian (or Wendic) along the river Spree east of Berlin and Dresden in Germany. Their dialects became separated by German settlers from the neighbouring West Slavonic languages, Czech and Polish in the thirteenth century. Based on the two dialect groups (Lower Sorbian in the North which is closer to Polish and Upper Sorbian in the South which is closer to Czech) two standard written varieties of Sorbian have developed, differing mainly in sound patterns and phonology. Throughout history the Sorbs have been subject to various forms of repression, most fierce during the Hitler regime, when use of their language became a punishable offence. Yet, they managed to preserve their language, until today, although they are nearly everywhere outnumbered by Germans. The total number of Sorbs is estimated at 90,000 but the number of active speakers, all of whom are bilingual, is much smaller: some 25,000. Both varieties of Sorbian are used in print and to a limited extent in education. Sorbian shares some archaic features with Slovenian, such as a dual.

The major West Slavonic languages are Polish (with some forty million speakers in Poland), Czech (with over ten million speakers in the Czech Republic), and Slovak (with an estimated five million speakers in Slovakia).

In the thirteenth and fourteenth century the kingdom of Bohemia (with Prague as its capital) was the most powerful state of Central Europe. At first Latin was used in writing at the time, but the fourteenth century was the first period of a flourishing Czech literature. In the early fifteenth century Jan Hus (1369-1415) initiated a religious reformation which was the beginning of a second period in the history of literary Czech. He reformed the spelling

and modernized the language using the dialect of Prague as its standard, but in 1415 he was burnt to death at the stake for his religious ideas. His followers produced a much revered translation of the Holy Script (known as the Kralice Bible). For two centuries Czech religious and philosophical writings flourished. But the Catholic counter-reformation finally succeeded in crushing the Hussite and Protestant movements (in the battle of the White Mountain, 1620), and put a temporary end to Czech as a written language. It was replaced by Latin and later, in the eighteenth century by German. In the nineteenth century, however, it was revived again. It is still based on the dialect of Prague, and has been subject to purist cleansing. Most varieties of spoken Czech are rather different from it.

Before the seventeenth century the Slovaks used Czech as their literary language, sometimes with Slovakian traits. After 1620, when Slovakia became part of Hungary, they had to use Latin. Towards the end of the eighteenth century attention began to be paid to the Slovak language, but the modern literary Slovak was created around the middle of the nineteenth century by a group of authors on the basis of central Slovak dialects.

Different from the other West Slavs the Polish have always remained Roman Catholic. From the establishment of the first Polish bishopric (tenth century) the written language was Latin. Yet there are early Polish religious texts from the fourteenth century, and more and also other texts from the fifteenth. The invention of printing coincided with the Polish "Golden Age", the sixteenth century, which resulted in a wealth of Polish publications. In the next century, however, Polish power declined in a series of devastating wars with Sweden, tsarist Russia and German states, and so did its literature. Towards the end of the eighteenth century a revival began, which reached its peak in the nineteenth century when authors like Mickiewicz closed the gap between the stilted literary Polish and the spoken language, resulting in the standard Polish language which is still used today.

Polish traumatic history caused large-scale migration of Poles to other countries: the number of speakers in Poland possibly exceeds forty million, but some six million Poles live in the United States, with a large concentration in Chicago. The large-scale migrations within Poland itself because of its shifting borders have had the result that there is little dialect differentiation. The most deviating dialect is Kashubian, which sometimes is considered a separate language. It is still spoken by perhaps a few thousands of the 100,000 Poles west of Gdynia who consider themselves ethnic Kashubian.

The East Slavonic languages, Russian, Ukrainian, and Belarusian, were still one language when the first East Slavonic state came into being with the rise of the city of Kiev. Built by Vikings as a stronghold on their trade route from the East Sea to the East Roman Byzantine Empire, it soon became the centre of an orthodox Catholic cloister culture with Old Church Slavonic as the language of its scriptures. The invasion of the Mongols (begun in 1237) put an end to this culture and to the supremacy of Kiev. Two centuries later the centre of East Slavonic (Russian) culture had shifted to Moscow. In the

centuries following the Russian defeat of the Tatar khanates (see above), Russians spread to all corners of European Russia, and made Siberia part of their empire. The expansion of Lithuania in the late Middle Ages, however, separated the Ukraine and Belarus from tsarist Russia, as a consequence of which their varieties of East Slavonic began to diverge. Before Lithuanian was used in writing, Belarusian became the official language of the Lithuanian state (fourteenth through seventeenth century). But it lost its status after it became part of tsarist Russia at the end of the seventeenth century and was hardly used in writing again until the communist revolution of 1917. A standardized version was agreed upon only in the 1950s. After the disintegration of the Soviet Union, Belarusian (with about eight million speakers) became the national language of Belarus together with Russian.

After the defeat of the Tatars and the fall of the Byzantine Empire, Moscow was seen by its tsarist rulers as the “third Rome”, the centre of the true Christian belief. It had inherited the orthodox cloister culture of Kiev, with Old Church Slavonic as its major written language. When Tsar Peter the Great opened Russia to the west and to “modernization” (including slavery), a more secular variety of written Russian had to be created. This was still sufficiently different from the spoken varieties of Russia. The great authors of the nineteenth century, however, succeeded in closing the gap and create the flexible standard Russian still used today as one of the largest languages of the world. It is the language of more than hundred forty million speaker in the Russian federation and of many millions in the countries which in 1990 were still parts of the Soviet Union.

The early development of a Ukrainian literary tradition was cut short by the fall of Kiev. Under Lithuanian rule the (spoken) language further diverged from Russian. Only towards the end of the eighteenth century a new literature came up. It was closer to the spoken language than Russian was in the same period. In the nineteenth century written Ukrainian was perceived by the tsarist government as a vehicle of nationalism and a danger to the unity of the state, and its use in printing was forbidden. After the October revolution of 1917 it became the official language of the Ukrainian Socialist Soviet Republic, and after the fall of communism of the independent Ukraine. An estimated forty million people speak the language, albeit often (still) mixed with Russian.

2.4.7 CELTIC LANGUAGES

For the ancestors of the Celtic tribes central Germany had become their temporary homeland. From there they spread between the tenth and fifth centuries BC south to Switzerland, Austria and Bohemia, west into France and Belgium, and beyond, in other words into the Iberic peninsula and to Ireland and the British isles.

In the South they were in contact with Italic tribes, with whom they are assumed to have had a period of common development. The rise of the Roman Empire and the wanderings of Germanic tribes put an end to their omnipresence in central and western Europe. Today the last pockets of Celtic

speakers are found in the western periphery of Europe.

Irish Gaelic used to be the language of Ireland before it was colonized by the British in the seventeenth century. In the early Middle Ages it was the literary language of a religious civilization from where missionaries were sent out to convert the then “heathen” countries of north-western Europe to Christianity.¹⁶ The annexation of Ireland by the British in the seventeenth century secularized the Irish society and ousted Irish as the language of the elite. It remained the language of the poor. About one million of them, however, died in the years 1848-1854 of starvation when the potato harvest failed. Today Irish is taught at schools as the national language of the Irish republic besides English, but it is spoken as a daily language only by some 20 to 30,000 speakers (or even 70,000 according to other estimates) in a number of isolated pockets in the west and southeast of the country.

Scots Gaelic is the language of Celtic invaders from Ireland, who from the fifth century onwards came to replace the language of earlier Celts and of the pre-Celtic Picts, who are known from Roman sources and early Irish records and whose language may have been non-Indo-European. Today Scots Gaelic is still spoken by an estimated 60,000 speakers in the western peripheries of Scotland. As in Ireland the remaining speakers are all bilingual. Shift to English is common. So much so that in the linguistic literature Scots Gaelic has become the prototype of a moribund language. One other Gaelic variety, Manx, ceased to be spoken as a first language when its last native speaker died in 1972; there has been a revival movement, however, in the last decades; it is learned in school as a second language and there are now again some 50 fluent (second language) speakers.

Cornish, an offspring of Old British, which was once spoken in Cornwall (southwest England), became already extinct in 1777. The language most direct related to it is Breton, still spoken by some 300,000 speakers in the rural areas of the French peninsula of Bretagne (Brittany). They are the descendents of Celts from southwest England who had yielded to the pressure of raids from pre-Christian Irish Gaels and of the Germanic Anglo-Saxon invasions into Brittany in the era of the “wandering of nations”. French policy towards national minorities has never been characterized by permissiveness, and the future of Breton is bleak: domains of use are limited, bilingualism is common and so is shift to French. According to the report *Parler breton au XXI siècle* (Speaking Breton in the twenty-first century) Breton lost 28% of speakers between 1997 and 2007, the number decreasing from 246,000 to 172,000 people (see <http://www.nationalia.info/en/news/435>).

Welsh (or Cymraeg) is spoken today by some 0.7 million speakers in Wales and England. At the beginning of the twentieth century half of the population of Wales spoke the language, towards the end of the century this figure had

¹⁶ This is the reason why Indonesian Christians pray before meals. The Dutch were Christianized in the eighth century by Irish monks who did so. Southern Europe, where it is not the custom to pray before meals, has been Christianized through contacts with civil servants, merchants, and mercenaries, who never did.

fallen steadily to some 30%. Increased attention to the language in education and an active movement for its preservation and cultivation has turned the tables in the last decade: the use of Welsh is slowly increasing again, notably also in urban areas.

2.4.8 ITALIC LANGUAGES

The Italic tribes must have had a period of common development with the Celts, before they moved south into Italy around the twelfth century BC. The rise of Rome since the eighth century BC as the central power in Italy caused the extinction of Italic languages other than Latin, such as Osco-Umbrian, and Venetian, as well as of the languages of the non-Indo-European Etruscans¹⁷ in Tuscany, whose civilization reached its peak in the eighth – fifth centuries BC. From these Etruscans the Romans took over their version of the Greek alphabet, which subsequently developed into the Latin alphabet we use today. At an early stage, Latin “fell apart” into two varieties: a cultivated literary and rhetorical one and a daily spoken one, called “vulgar Latin”.

In the second century AD the Roman empire included the whole of North Africa and the Middle East, and united about four fifth of the European population. One of the results of Roman rule and dominance was that nearly all mainland Celtic language speakers shifted to (vulgar) Latin. These spoken varieties became the ancestors of modern “Romance” languages. Meanwhile, a written variety, much more similar to classical Latin, remained the official language of the empire. Latin also became the language of the (Roman) Catholic Church when Christianity was proclaimed the state religion in the fourth century by the emperor Constantine. Long after the Middle Ages Latin remained the language of science: even in the nineteenth century it was still customary at European universities to write PhD dissertations in that language.

The differences between the various Romance languages of today are explained not only by regular linguistic change, but also and foremost by the historical influence of different substrata and adstrata, by geographical isolation, and by political fragmentation.

Proto-Rumanian became separated from the rest of the Romance world as a result of the intrusion of the Slavonic peoples in the Balkan Peninsula (which after centuries caused the extinction of Dalmatian Romance, the last speaker of which died in 1901). The present-day national language was shaped in the course of the nineteenth century, when the Balkan peoples after ages of Turkish rule acquired independence.

Lexical influence of Slavonic languages (Bulgarian and Russian) and of Hungarian in Rumanian is strong. The language also developed features which are characteristic of the Balkan *Sprachbund*, such as post-nominal articles (together with Bulgarian, Macedonian, and Albanese), and simplifications in

¹⁷ So far no living language has been discovered which is genetically related to Etruscan. Although the Etruscan inscriptions can be read, they are not understood. According to a recent theory their roots may have been in Northwest Asiatic Turkey.

its nominal case and verbal tense system (together and parallel with the same languages and with Greece). Some 1000 years ago Rumanian nomads settled in Greece, Macedonia, and the Croatian peninsula of Istria. Their dialects which are highly endangered if not moribund, also called Vlach or Arumun, differ considerably from Rumanian.

That cannot be said of Moldavian. Before World War II Moldavia (or Bessarabia as it was called at the time) was still part of Rumania; but after the war it became part of the Soviet Union, to become an independent state after the fall of communism (1989). Since then it has become a political affront not to subscribe to the opinion that Moldavian is a language of its own. The estimated number of speakers of Rumanian and Moldavian together is well over 23 million.

The fall of the West Roman empire was accompanied by several invasions of Germanic tribes into the Italian peninsula. For ages Italy remained an instable and ever changing mosaic of independent city-states and principalities, contested by foreign powers, and resulting in considerable dialectal differentiation. The varieties of Romance in the islands of Corsica (France) and Sardinia (Italy) may be considered separate languages, the latter characterized by marked archaic features. The same holds for the Rhaeto-Romance varieties of Romance spoken in the Alpine North of Italy and in the Swiss canton of Graubünden. Romansh in Southeast Switzerland has the status of a national language in the country (together with French, German and Italian). It has two written varieties since the seventeenth century, used respectively by the Roman Catholic and Protestant sections of the population. Although protected and used in education, it is yielding to the numerical and cultural overweight of (Swiss) German, its number of speakers being only some 60,000 according to the census of 2000. Ladin is closely related to Romansh and spoken by some 30,000 speakers on the Italian side of the border with Austria. Friulian is spoken in the Italian mountains north of Venice by perhaps half a million speakers; it is more distantly related to Romansh and rather heavily influenced by Venetian Italian. Istrot is an archaic variety of Romance, still used at home by some families in two towns in the Croatian peninsula of Istria. It may constitute the link between Friulian and the extinct Dalmatian Romance.

As a consequence of the political fragmentation of Italy, there was not one centre which could be the natural cradle of a unifying language. It can be said that Italian as we know it is largely the creation of one man, the poet Dante Aleghieri (1265-1321), who was born in Tuscany like the other two great Renaissance authors Boccaccio and Petrarca. The language Dante created on the basis of his native dialect with borrowings from other dialects and Latin became the mould for later authors. Through the ages standard Italian remained the language of an elite, however. Only after Italy was united (1863) and Rome had become part of it, did Italian become the national language of the whole of Italy. The position of the dialects, however, remains strong. Within Europe Italian is also spoken in Switzerland (Tessino), where it is also

one of the national languages.

In the Iberian Peninsula, vulgar Latin varieties ousted the local Celtic dialects and other indigenous languages, with the exception of Basque. After the fall of the West Roman empire Germanic tribes invaded the region, such as the West Goths and the Vandals. The latter, which as their name betrays became notorious for their rather destructive inclinations, raided on to North Africa to be finally destroyed themselves by forces of the East Roman empire. The former established a flourishing kingdom in southern Spain, which still existed when the Muslim armies invaded Europe in the eighth century. Their rule was characterized by freedom of religion (Judaism, Christianity, and Islam) and by flourishing arts and sciences. But lack of a centralized power which was able to unite the whole of the peninsula resulted in changing coalitions and frequent war. The northern alliance, headed by Castile and inspired by a Taliban-like Catholic intolerance and fanaticism succeeded in gradually rolling back the Islamic states, a process which took more than four centuries (from the eleventh century till 1492), and which ousted the Spanish variety of Arabic (Mozarabic). The leading role of the Castilian kingdom in this *reconquista* caused the Castilian variety of Romance to become the model of modern Spanish. It also became the variety which was spread to the Spanish colonies in South and Central America, where it became the national language of all countries, except Belize, most Caribbean islands, the Guyanas, and Brazil. The last decades it is making strong inroads into the United States, where it already has outnumbered English in California. The total number of Spanish speakers in Spain is about 35 million and well over 350 million in the whole world.

The closest ally of Castile in the *reconquista*, the kingdom of Aragon ruled northeast Spain, where the local variety of Romance, Catalan, persisted. Although Aragon extended its power to Sardinia and Naples in Italy, Castile retained its supremacy in Spain. In addition, when towards the end of the seventeenth century the Catalans revolted against Castile, Catalan universities were closed down and the language was henceforth forbidden in printing. In the nineteenth century a literary form of it was reinvented, to be suppressed again under the fascist dictatorship of Franco. Only after his death in 1975 Catalan resurrected as one of the official languages of Spain, in vigorous use in Catalonia (Northeast Spain). In Andorra Catalan has the status of a national language together with French. Its number of speakers in Spain, Andorra and adjacent regions of France is estimated to be more than twelve million. Also in Sardinia there still seem to be (an unknown number of) Catalan speakers.

Under the feudal system of the later Middle Ages the Northwest of Spain (Galicia) and northern Portugal had acquired some independence vis-à-vis Castile. Their contribution in the *reconquista* was directed towards southern Portugal which was conquered in the twelfth century. The linguistic result was that Galician was not completely overrun by Castilian (although it was influenced by it), and that it is recognized today as one of the national languages of Spain like Catalan (and Basque). In Portugal, which became an independent

country – except for a short period in the sixteenth and seventeenth century - Portuguese, a less influenced form of Old Galician, became the national language. It is spoken by more than ten million Portuguese, and by ten times as many Brazilians. In other former Portuguese colonies where mass education has been less advanced, it has official status, but has remained basically the language of local elites.

At the northern border of Spain, Catalan forms a linguistic bridge to Provençal. This latter language is the contemporary remainder of the once widespread *langue d'oc*, which was the variety of Romance in France which was not influenced by the Franks. A bundle of isoglosses, which are not only linguistic but which also pertain to traditional architecture, customs and the like, separates (or separated) the area connected with this *langue d'oc* in southern France from its northern counterpart where *langue d'oïl*¹⁸ is indigenous. In the eleventh and twelfth centuries the southern regions were the cradle of troubadour culture, the culture of itinerant singers and their songs of courtly (and other kinds of) love, which gained an enormous popularity all over western Europe.

In the thirteenth century, however, the cultural importance of southern France and its language declined rapidly, after the crusade of the French from the north against the “heretics” of the town of Albi in southern France. In the process the northerners put an end to the relative independence which the southern nobility, gentry and free farmers had enjoyed before. They became subject to the centralized rule of the French king and exposed to the influence of French, Middle French at first. What remains of Provençal today is a complex of mostly rural dialects which are influenced by French to different degrees. In the Gascony region (Southwest France) it has preserved its purest form, and is still widely spoken, although also here shift to French is common.

French itself is the variety of the former *langue d'oïl* which came up in the Île de France region around and including Paris. It became a prestigious language of learning alongside Latin (which continued to enrich it) after the establishment of the Sorbonne University in Paris (1257). The Île de France variety was proclaimed the national language of France by king François I in 1539. Its modern standards were laid down in the seventeenth century by the works of the poet Malherbe, and subsequently by the *Académie Française*, established in 1635 with the explicit task to guard the purity of the French language. The conservative attitude expected of its members is the reason why the official standard always has been rather slow in following and accepting the changes occurring in the living spoken language. A recurrent hot item in the last decades has been the lexical influence of (American) English. But daily practice is stronger than ideology. Even in the vocabulary of fashion (for a long time nearly a French monopoly) English is making inroads: today fashion magazines discuss the colours of *le tee-shirt* and *le pull*, elaborating on *le must du sexy*.

¹⁸ *Oc* and *oïl* in de names for these two language varieties mean ‘yes’ and are both derived from Latin *hoc illud* ‘that’s it’.

As in most European countries the modern form of the literary standard was shaped by the major authors of the nineteenth century, with in the case of France the acknowledgement of the *Académie Française*. The appeal of French to the Europeans of the eighteenth and early nineteenth century was as suggested in the beginning of this paper the exuberant life style of the French court for the European courts and related elites, but for wider circles it had also been the influence of the philosophers and encyclopaedists of the eighteenth century Enlightenment.

Today French is the first language of well over fifty million speakers in France, four million in Belgium (where it is one of the official languages together with Dutch and German), perhaps 100,000 in Northwest Italy (where it has also official status), and 1.5 million in Switzerland (where it is one of the four national languages). Small communities of first language speakers are found in Europe in Luxemburg, Monaco, and Andorra, and outside Europe in the former and present French colonies in Africa, Oceania, Southeast Asia, the Caribbean etcetera. The largest French speaking community outside Europe (over seven million speakers) is the one of Canada. The total number of first language speakers of French in the world probably exceeds 70 million.

2.4.9 GERMANIC LANGUAGES

The temporary homeland of the Germanic tribes after the dispersion of the Indo-Europeans is thought to have been the southern coasts of the East Sea (northern Poland and Germany and southern Scandinavia). Given early Germanic loanwords in Finnish there must have been contacts with the Balto-Finnic tribes around the Baltic Sea. From there they spread westward into the low countries and southward into southern Germany and Celtic territories. They frequently clashed with Roman armies at the frontiers of the Roman Empire. In the period of the wandering of nations Germanic tribes belonged to the most active wanderers: Galatians ended up in Asiatic Turkey, Langobards through Hungary in northern Italy, Burgundians in eastern France, Batavians in the Netherlands, Franks in the Netherlands, Belgium and northern France, Gepids in Hungary, Vandals in North Africa. Already before the large wandering of nations Germanic tribes had begun their migrations. In the third century Gothic tribes moved from the Danish islands and South Sweden to North Poland and from there further south. One group, the East Goths settled down in the Crimea peninsula in the southern Ukraine, together with the Heruli (who originated from mainland Denmark). The last traces from the east Gothic presence in the Crimea disappeared only in the eighteenth century. Another group, the West Goths helped to destroy the West Roman empire and ruled Italy for some decades before they were beaten by Byzantine forces. They ended up in southern Spain where they established a prosperous kingdom which in 711 fell to the Muslims. From these early Germanic tribes hardly any linguistic traces are left. In most regions they disappeared, leaving at best some names (such as France < Franks), loanwords and in some cases features ascribed to Germanic influence (such as the presence of front rounded

vowels in French which are atypical of Romance languages). The oldest written source of these early Germanic wanderers is the "silver Bible of Uppsala" written by the West Gothic bishop Ulfilas or Wulfila in the fourth century in Latin, Greek and Gothic runic script.

Meanwhile in Scandinavia the runic script was used in inscriptions which date back to the third century. They are written in Old Norse, the language of the early Vikings. This language split into a western (Norway) and an eastern branch (Sweden-Denmark). (Western) Old Norse was transplanted to Iceland where it lives on as Icelandic, the most archaic Germanic language alive, spoken and written by some 300,000 Icelanders. Another daughter language of this Western Old Norse is Faroese, native language of some 50,000 speakers, and the official language of the Faroe Islands between Scotland and Iceland. For both these languages standards were developed in the nineteenth century. Icelandic is the national language of Iceland. Faroese and Danish have that status in the Faroe Islands.

In Norway, however, Danish had functioned as the written norm because of the Norwegian-Danish dynastic union which had united the two countries since the late Middle Ages. On the basis of the Norwegian dialects the Norwegian linguist Ivar Aasen (1813-1896) created a new Norwegian standard language (Landsmål or Nynorsk), which today coexists with the Danish-based Bokmål. Some 20% of the Norwegians use Nynorsk in writing, 80% stick to Bokmål, which in the last century has undergone adaptations and become more Norwegian.

Danish itself and Swedish began to drift apart when in the sixteenth century Protestantism postulated that sermons and religious texts including the Bible should be understood by the community of believers, that is by the common people. In this way the emerging differences between Danish and Swedish became codified as two different languages, which subsequently each developed in its own way.

Today there are over five and a half million speakers of Danish in Denmark and across the southern border in Germany, some nine million of Swedish in Sweden and western Finland, and about five million of Norwegian dialects in Norway, and in some places across the border in Sweden.

Among the old *West* Germanic language varieties an early split occurred when a number of tribes moved south. When the West Roman empire ceased to exist its borders were no longer defended, and the Celtic tribes living there were easily overrun and outnumbered by the Germanic Alamans, Bavarians, and Langobards. The latter ended up in northern Italy (Lombardia), where their language disappeared in the eleventh century. The Alamans moved to southwest Germany, crossed the river Rhine and settled in northeast France (the Elzas) and Switzerland. The Bavarians became the ancestors of the Bavarians (Southeast Germany) and the Austrians of today. The contacts of these Germanic tribes with the Celts who took over their languages brought about the sound changes typical of the later High German (that is the German of the higher countries). In the lower northern regions phonologically more

conservative Germanic dialects continued to be spoken.

One of those dialects was Proto-Frisian, spoken along the North Sea coasts from the western Netherlands to Denmark. The earliest (Old) Frisian texts are charters from the late Middle Ages. In 1498, however, Dutch was declared the official literary language of the later Dutch province of Friesland, where Frisian is still spoken today. In the seventeenth century (Middle) Frisian literature flourished, to decline again in the eighteenth century. A revival movement starting in the nineteenth century finally resulted in the establishment in 1938 of the *Fryske Akademie*, which had to study and document Frisian language and culture and defend its rights vis-à-vis Dutch. Since the middle of the twentieth century Frisian is taught at schools in the Dutch province of Friesland, and officially recognized as a national language besides Dutch. It is the mother tongue (with dialectal differences) of some four to five hundred thousand native speakers, mainly in Friesland, and all of them bilingual in Dutch. Outside Friesland (East) Frisian dialects are still spoken in a few pockets in northwest Germany by at best a few thousand people.

Other Low Germanic tribes were the Franks, the Angels and the Saxons. The Franks who had moved west already in Roman times into the Netherlands, Belgium and northern France became the linguistic ancestors of the Dutch, the Saxons of the speakers of "Low German" dialects in northern Germany. These Low German dialects were quite influential in the late Middle Ages, when the bond of Hanseatic towns (most of them in northern Germany) controlled all trade in the East Sea region. This influence was strengthened by the activities of the German Order, an army of knights raised in the end of the twelfth century, whose aim it was to convert the heathen Baltic peoples to Christianity by force of arms. Their expeditions put an end to Old Prussian and possibly to other Baltic languages as well. They were finally beaten by Lithuania in 1410, but had made Prussia into a German state.

Like in other countries Protestantism and the invention of printing caused the language situation to change dramatically. In Germany the writings and Bible translation of Luther which were printed in more than 100,000 copies became the model for literary German for the centuries to come. It was based on Luther's variety of High German, and built forth upon by the major authors of the nineteenth century. It completely ousted the role of the Low German dialects in writing.

In the course of history large numbers of Germans migrated as labourers, mineworkers, farmers, or as a buffer along former frontiers, to what are now neighbouring countries, such as Poland, Bohemia (Czech Republic), and Rumania. After World War II Germany lost much of its territory in the East, and an estimated fifteen million Germans remigrated to Germany, mainly from Poland and Czech-Slovakia. More Germans followed after the lifting of the Iron curtain: the former Volga German Soviet Republic (migrant farmers from the early eighteenth century) ceased to exist. Also in Rumania the number of Germans who had lived there for eight centuries rapidly declined. While under communist rule there was still a large community of Transylvanian

Germans of more than 300,000 speakers, in 2002 their number had already shrunken to 60,000 (see <http://faroutliers.wordpress.com/2004/03/16/the-vanishing-breed-of-transylvanian-german-writers/>).

In Germany, Austria, and Liechtenstein, High German is the only official language. In Switzerland it shares this status with French, Italian, and Romansh, in Belgium with Dutch and French, in Italy with Italian, French, and Slovenian), in Luxemburg with French and Luxembourgish, the local variety of German. Worldwide the number of speakers of German approaches one hundred million. Another German dialect or German based language (with strong lexical influences from Hebrew and Slavonic languages) is Yiddish, the language of the Ashkenazi Jews. It originates from the German used by Jews who had migrated from southwest Germany and the Elzas to eastern Europe in the Polish Golden Age. In the Belarusian Socialist Soviet Republic their language had official status. Rejecting Zionism they remained in Eastern Europe only to become victims of the Hitler regime and be largely destroyed.

Like modern High German, Dutch too acquired its modern standard shape as a direct result of the Protestant Reformation. In the war with Spain (from 1568 to 1648) for freedom of religion (and taxes), the Belgian provinces which were Dutch (Flemish) speaking were occupied by Spanish forces. Many Flemish Protestants fled to the northern Netherlands. Differences between the then existing dialects were considerable, but in order to make printing profitable and to reach as many readers as possible a language variety had to be sought which would be understood by speakers of most dialects. A state committee for an authorized translation of the Bible was therefore formed in which refugees from Flanders and representatives from the most prosperous and powerful Dutch provinces cooperated. The result was the so called States Bible (1637), which became the model for written Dutch in the following centuries. Although more popular forms of literature with a more folkish language continued to be written, there did evolve a gap between the written standard and the spoken varieties of Dutch in the first half of the nineteenth century. In the second half of the nineteenth century this gap was definitely bridged by probably the major Dutch writer, Eduard Douwes Dekker (alias Multatuli), the author of the anti-colonial roman *Max Havelaar*. Today Dutch is spoken with dialectal differences but with one standard in the Netherlands and northern Belgium by more than twenty three million speakers. Flemish dialects are also spoken in the northwest corner of France, but there the language has no status and is yielding under French pressure.

A final offspring of Low Germanic is – mixed with other elements – today the major language of the world, numerically only second to Chinese (and possibly Hindi, broadly defined), but spoken by far more people as a second language: English. In the time of the wandering of nations Low Germanic tribes from northern Germany and adjacent mainland Denmark, Angels and Saxons, Jutes and some Frisians crossed the North Sea and settled in southern England. The Roman armies had withdrawn with the fall of the West Roman empire and the British Celts, weakened also by raids from Irish Celts, were

too dispersed to organize enough resistance. In the next centuries the Anglo-Saxons were followed by Danes who settled mainly in the eastern province of Northumberland. The mixture of Low Germanic varieties and Old Danish resulted towards the end of the first millennium in Old English. Meanwhile the Scandinavian Vikings had started their raids onto the coastal areas of Europe. They settled on the Scottish west coast, where their language (known as Norn) lived on until the eighteenth century. Their arrival may have been the final blow to the language of the Picts, which possibly was still spoken in the eighth century on some of the Orkney Islands. Vikings had also settled in the French peninsula which hence became known as Normandy (land of the "Nordmen") where they had shifted to (Old) French. In 1066 a Norman army under King William the Conqueror crossed the Channel and defeated the Anglo-Saxon armies in southern England. The French speaking Normans installed themselves as the new nobility and ruling class, reducing the original population to serfdom. After two centuries, however, the nobility had shifted to the language of the majority, enriching it with French lexical alternatives for many daily and less daily concepts. This influence of French/Romance/Latin continued in the following centuries as a corollary of the rise of scientific and academic discourse. Meanwhile the language had simplified its flectional morphology considerably, replacing nominal cases with analytical constructions. The major authors of the Elizabethan era, Shakespeare in the first place, finally made English into the flexible and rich language it is today. Its success as a world language is to some extent due to this character, to some extent also to its role as the language of the British colonial empire, but most of all to the economic, military, political, and cultural power of England's former colony, the United States.

In Europe in the United Kingdom and Ireland English is spoken in various rather deviating dialects by nearly sixty million first language speakers. In the United States by some 225 million with very little dialect differentiation. In the rest of the world (Australia, Canada, New Zealand, South Africa, India, etcetera) there are another fifty to sixty million.

3 EUROPE'S FUTURE?

Lack of space prevents me from discussing the historical changes in the political map of Europe in any detail. Recent history has shown that the political and linguistic map of Europe is still very much subject to change. The emergence of Croatian, Bosnian and Serbian as three separate languages has already been discussed. That the expansion of the European Union in the long run will lead to the emergence of a modern koine does not seem to be likely. The old member states (Belgium, the Netherlands, Luxemburg, France, Italy, and Germany since 1967; Denmark, Ireland and the United Kingdom since 1973, Greece since 1981, Spain and Portugal since 1986, Finland, Austria, and Sweden since 1995) adopted the policy that speaking one's mother tongue better than a foreign language should be no reason for putting anybody at a disadvantage. Consequently, all national languages of these member states

(Danish, Dutch, English, Finnish, French, German, Greek, Italian, Portuguese, Spanish, and Swedish) officially still are official *and* working languages of the Union. This implies an enormous amount of translations, simultaneous and otherwise, between all these eleven languages. Now that Poland, Hungary, the Baltic states, the Czech Republic, Slovakia, Slovenia, Malta, Bulgaria, Rumania and (Greek) Cyprus have joined the Union, the number of official and working languages has only grown. But this has hardly been a matter of political discussion so far. Sensitivities have been too high.

If a language is “a dialect with a navy” it may be predicted that Europe will not come to see a *bahasa Eropa* as long as it does not succeed in creating a truly European supra-national army. In terms of armed forces and global appeal English is of course a strong candidate to become the actual working language of the European Union. Moreover, in practice it is indeed increasingly used as such. As the sociologist Abram de Swaan has put it in this connection on several occasions: “the more languages the more English”. However, it will probably take more than one generation before it can be *officially* proclaimed to be the *bahasa Eropa*. The French will not accept English if French is not acknowledged as the other official language. The Germans would accept English, but if French would acquire official status too, then German should be adopted as the third one, and so forth. As a result the European Union is looking for interpreters Estonian-Greek, Lithuanian-Maltese etcetera.

The tentative conclusion must be that the “Indonesian experiment” is probably unrepeatable. No European language has the quality of Malay: known in some variety to many people as a second language, having a tradition of being used as a means of supra-ethnic communication, not perceived as exclusively advantageous for an identifiable section of the multi-ethnic community, and relatively easy to master up to a reasonably advanced level of sophistication.

BIBLIOGRAPHY

- Amran Halim. 1976. “Fungsi politik bahasa nasional”, in: Amran Halim (ed.), *Politik bahasa nasional 1*, pp. 13-25. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa Departemen Pendidikan dan Kebudayaan.
- Bray, R.G.A. de. 1951. *Guide to the Slavonic languages*. London: J.M. Dent & Sons, New York: E.P. Dutton.
- Comrie, Bernard. 1981. *The languages of the Soviet Union*. Cambridge: Cambridge University Press.
- Friedman, Victor A. 2000. “Observing the observers; Language, ethnicity, and power in the 1994 Macedonian Census and beyond”. Paper presented at the Kolokium Antarabangsa “Bahasa dan Masyarakat Serantau”, Institut Alam dan Tamadun Melayu Universiti Kebangsaan Malaysia, Bangi, 4-5 October.

- Haugen, Einar. 1966. "Semicommunication; The language gap in Scandinavia", *Sociological Inquiry* 36: 280-297. [Reprinted in: Evelyn Scherabon Firchow et al. (eds), 1972, *Studies by Einar Haugen; Presented on the occasion of his 65th birthday - April 19, 1971*, pp. 479-495. The Hague-Paris: Mouton.]
- Khanmagomedov, B. G. K. 1967. "Tabasarskij Jazyk", in: Vinogradov (ed.), *Jazyki Narodov SSSR. Tom čtvrtýj* (E.A. Bokarev ed.). *Iberijsko-Kavkazskie Jazyki*, pp. 545-561. Moskva: Nauka.
- Kuipers, A.H. 1960. *Phoneme and morpheme in Kabardian*. The Hague: Mouton.
- Kuipers, A.H. 1976. "Typologically salient features of some North-west Caucasian languages", *Studia Caucasica* 3: 101-127.
- Lewis, M. Paul (ed.). 2009. *Ethnologue; Languages of the World*. Sixteenth edition. Dallas, Tex.: SIL International. [Online version: <http://www.ethnologue.com>.]
- Smeets, Rieks. 1984. *Studies in West Circassian phonology and morphology*. Leiden: The Hakuchi Press. [PhD thesis, Leiden University.]
- Stevenson, Victor (ed.). 1989. *Atlas van de Europese talen; Geschiedenis en ontwikkeling*. Utrecht: Het Spectrum.
- Tolstoy, L.N. 1961. "Vojna i mir", in: *Sobranie sočinenij v dvadcati tomax. Tom čtvrtýj*. Moskva: Gosudarstvennoe Izdatel'stvo Xudožestvennoj Literatury.
- Vinogradov, V.V. (ed.). 1966a. *Jazyki Narodov SSSR. Tom pervyj* (V.V. Vinogradov ed.). *Indoeuropejskie Jazyki*. Moskva: Nauka.
- Vinogradov, V.V. (ed.). 1966b. *Jazyki Narodov SSSR. Tom vtoroj* (N.A. Baskakov ed.). *Tjurkskie Jazyki*. Moskva: Nauka.
- Vinogradov, V.V. (ed.). 1966c. *Jazyki Narodov SSSR. Tom tretij* (V.I. Lytkin, K.E. Majtinskaja eds). *Finno-Ugorskie i Samodijskie Jazyki*. Moskva: Nauka.
- Vinogradov, V.V. (ed.). 1967. *Jazyki Narodov SSSR. Tom čtvrtýj* (E.A. Bokarev ed.). *Iberijsko-Kavkazskie Jazyki*. Moskva: Nauka.
- Vinogradov, V.V. (ed.). 1968. *Jazyki Narodov SSSR. Tom pjatýj* (P.Ja. Skorik ed.). *Mongol'skie, Tunguso-Man'čžurskie i Paleoaziatskie Jazyki*. Leningrad: Nauka.