LANGUAGES

Hans Christian Luschützky

Department of Linguistics, University of Vienna

hans.christian.luschuetzky@univie.ac.at
http://homepage.univie.ac.at/hans.christian.luschuetzky/

Training Course on Toponymy

Run-up to the 23rd Conference of the

United Nations Group of Experts on Geographical Names

Vienna, March 2006

Speaking about language

As an object of investigation, language is special in that we are forced to use it in order to make statements about it.

Everything we say (or write) about words, phrases and sentences we can only say (or write) by means of words, phrases and sentences.

The consequences of this metalinguistic dilemma are intricate and manifold and cannot be discussed here.

It should be kept in mind, however, that speaking about language is a self-referential act, like shaking one's own hand.

The scope of language

Language, as the most complex of all forms of human communicative behaviour, can be approached and investigated from basically two points of view:

- system and structure (internal aspect)
- function and domain (external aspect)

Both approaches can be refined with respect to the historical (diachronic), geographical (areal) and social dimensions.

The following brief survey will focus on intersections of internal and external aspects of language: variation, glottogenesis, glottodynamics, linguistic stratification and language profiling.

Language system and linguistic structure

Structural analysis of language systems is based on procedures of segmentation and classification (taxonomy).

All human languages are similar with respect to the hierarchical organisation of structural levels:

- textual
- syntactic
- lexical/morphological
- phonological

(notice that writing and orthography, although closely related to language, do not form a structural level of the language system but constitute a separate semiotic system)

All human languages differ with respect to the formal implementation of the subsystems that are constituted according to the structural levels.

Structural levels of the language system

Textual level

utterance: a communicative action made up of one or more sentences (resulting in a text)

text: a sequence of sentences fulfilling criteria of coherence and cohesion

Syntactic level (sentence and phrase structure)

sentence: a syntactic construction with a logical content, made up of phrases, e.g. *The cat was lying on the red carpet* (the logical content is a statement susceptible to truth conditions)

phrase: a minimal syntactic construction made up of free forms, e.g. *on the red carpet* (no logical content, no statement being made)

Lexical/morphological level (word structure)

compound word: a free form made up of simplex words, e.g. eng. *Sunday* (*sun* and *day* occur as free forms). There are various types of compounds: e.g., *Sunday* is a kind of a day (*endocentric compound*), but a *laptop* is neither a lap nor a top (*exocentric compound*)

complex word: a free form containing bound morphemes, e.g. eng. *reconfirmation* (prefixes like *re-* and suffixes like *-ation* do not occur as free forms)

simplex word: a minimal free form (a *lexical* morpheme), e.g. eng. *cat*

morpheme: a minimal meaningful unit, e.g. *re-, con-, firm* and *-ation* in eng. *reconfirmation. Re-, con-* and *-ation* are *derivational* morphemes (adding semantic features to build new words); the verbal ending *-s* in eng. *she says* is an *inflexional* morpheme (adding semantic features that relate to meaning on the phrase or sentence level, here: 3rd person singular indicative present)

Phonological level (sound structure)

syllable: a minimal unit of pronunciation (speech), e.g. [kʰæˀ] (the word *cat* pronounced by a native speaker of American English in spontaneous speech)

phoneme: a minimal distinctive unit, e.g. /k/, /æ/, /t/ in eng. *cat* (versus *rat, cut, cap*)

Language variation

A main feature of language use is *variation* in the selection, arrangement and shaping of the units of speech (*micro-variation*): no two utterances are identical.

Permanent directed micro-variation and mutual tuning processes among speakers on the long run lead to *diachronic change* (*internal factors of language change, evolutive innovations*).

Contact between heterogeneous groups of speakers may cause *borrowing* and also lead to diachronic change (*external factors of language change, adaptive innovations*).

Limited diffusion of innovations caused by internal and external factors of language change leads to *dialect formation* (*macro-variation*).

Since internal and external factors of language change are unavoidable consequences and circumstances of language use, macro-variation occurs in every larger linguistic community, resulting in the rag-rug pattern of detailed dialect maps.

Under favouring conditions (mainly separation and isolation of speaker groups, e.g. due to migration), macro-variation leads to *glottogenesis*, making dialects mutually unintelligible due to accumulation of structural differences.

Apart from its descriptive aspects on the structural levels of the language system, macro-variation is an *historical, social* and *areal* phenomenon, conceivable in terms of *function* and *domain*.

Glottogenesis (the process by which a *dialect* becomes a *language*) can occur also without accumulation of structural differences, by means of social emancipation (*culturalisation,* bottom-up) or political action (*standardisation,* top-down).

Glottogenesis

Culturalisation starts with growing *awareness* among the members of a speech community of the dialectal differences between their own speech and neighbouring vernacular varieties or an established standard language.

A first step in the culturalisation process is *textualisation* (oral poetry and the production of literature).

The next step in the culturalisation of a vernacular variety is *codification* (assessment of the inventory of the language system by means of descriptive lexicography and grammar).

Standardisation starts with the implementation of prescriptive lexicography and grammar for glottodidactic use in primary and secondary education, language courses for foreigners and the like.

A further stage in the standardisation process is *officialisation*, by which the legal status of a language and its use is determined for a given domain.

The final stage of glottogenesis is *medialisation* (the proliferation of information in newspapers and journals, radio, television, and the internet).

Glottogenesis need not necessarily pass through all these stages. The process can come to a stop at any stage or skip intermediate stages. Especially nowadays, medialisation can set in without previous officialisation, standardisation or codification, since textualisation can be implemented directly, and even globally, by means of the Internet or regional/private broadcasting activities.

Glottodynamics

As a countermovement to the multiplication and diversification of varieties of a given language, *dialect levelling* may result from intensive contact among the members of neighbouring or cohabitant groups. Levelling is hardly ever symmetrical; in most cases, a variety with higher prestige supersedes the other(s).

The process of *language obsolescence, attrition* and *decay*, and its endpoint, *language death (extinction)*, is attracting general attention nowadays (*endangered languages*), but has always been at work in the history of mankind.

A countermovement to language decay is *language revival (reculturalisation),* but most of the languages endangered at present have no chance to be ever revived by cultural activists. Only linguistics can save the "memory of mankind" by conserving the accessibility of texts written in *extinct languages* and documenting languages threatened by extinction.

In complex ethnolinguistic settings, collective *bilingualism* and *diglossia* may lead to *language fusion,* resulting in a *mixed language*.

Where communication is needed, but the languages of the respective groups are very different, a third language may be used as a *vehicular language*, or an *auxiliary language* may arise (*lingua franca, pidgin language*). Pidgin languages are typically used in restricted domains (*trade pidgin, plantation pidgin, mine pidgin, sailor's pidgin, garrison pidgin* etc.).

When an auxiliary language becomes the primary language of an upgrowing generation, *creolisation* takes place, whereby the rudimentary pidgin system is enriched with newly developed grammatical categories. Creole languages may run through all stages of glottogenesis and become official languages of states. On the other hand, like any language, they may also fall into oblivion and die out.

Linguistic stratification

When speakers have given up their inherited language and taken over another one of higher prestige (either by free will or under compulsion), *language shift* has taken place.

Language shift may proceed slowly over many generations, leaving traces (*substratum*) of the *recessive language* in the *persisting language*.

Apart from being an object of linguistic investigation as such, substratum is an important source of information for the reconstruction of *prehistoric ethnography*.

Substratum can manifest itself on any level of linguistic structure, but mainly it is the *lexicon* that preserves elements of the recessive language (*relic words*), especially in *proper names*.

Due to their *individualising function*, proper names are felt by speakers to be part of the thing they denote. The feeling of identity of a name and its bearer becomes evident in language *taboo* (naming prohibition).

The more persistent a phenomenon is, the more persistent will be its designation. That is why *toponyms* are usually the most archaic items to be found in a linguistic area.

Historical *flow of migration* and *layers of settlement* are often recoverable only by the study of *toponymic stratification*.

Since *territorial claims* are sometimes based on priority of settlement, toponymy can easily become a *political issue*.

Language profiling

For systematic surveying, every language of a given linguistic area has to be specified in terms of structural and functional features.

A minimal set of features would have to consist of the following:

Name (*glottonym*, including *synonyms, endonyms, exonyms*)

Genetic affiliation (*group, branch, family, stock, phylum*)

Geographic location (*homogeneous area, enclaves, exclaves, scattered population* etc.)

Dialectal diversification (number and relative prominence of main dialects)

Status (*official language, literary language, regional language, minority language*, etc.)

Writing system (kind of *alphabet* or other system; *degree of medialization; literacy rate* of population)

Number of speakers (*primary* and *secondary* speakers) in the order of thousands (M) or millions (M^2)

Contact languages (neighbouring and interspersed, cultural relations)

Example 1: Profile of Macedonian

Name: Macedonian, македонски јазик

Genetic affiliation: South Slavic, Indo-European

Geographic location: Macedonia and neighbouring regions of Albania, Bulgaria and Greece

Dialectal diversification: Northern (Kumanovo-Kratovo Region), Southeastern (around Gevgelija, Strumica and Lake Dorjan), Western (one subdialect in the Veles, Prilep, Kichevo and Bitola Region, another in the Debar-Galchnik Region)
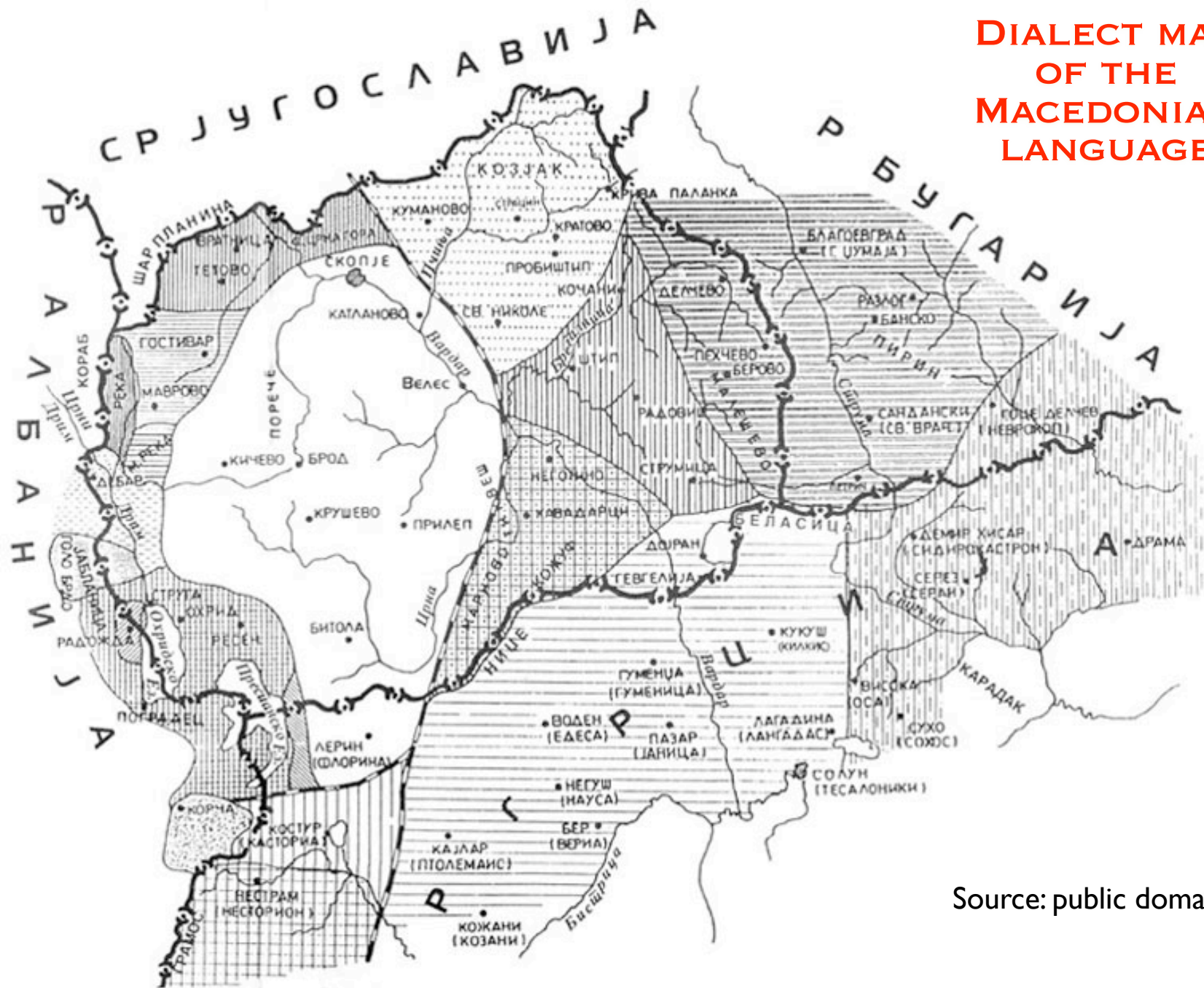
Status: National language of the Republic of Macedonia (proclaimed in 1947), minority language in neighbouring regions of Albania, Bulgaria and Greece

Writing system: Cyrillic alphabet, fully medialised

Number of speakers: 1,3 M^2 in Macedonia (65% of total population), 180 M in Greece, 15 M in Albania, undetermined number of speakers in Bulgaria

Contact languages: Bulgarian, Albanian, Greek, Serbian, Serbo-Croatian, Turkish. Ethnic/ linguistic minorities in Macedonia (with regional official status): Albanian (510 M = 25%), Turkish (78 M = 4%), Romany (60 M = 2,7%), Serbian (35 M = 1,8%), Bosnian (17 M = 8‰), Aromunian (10 M = 5‰), other (21 M = 1%)

Dialect map of the Macedonian language

Source: public domain

Serbia and Montenegro

Albania

SERBIAN

Bulgaria

42°N

Skopje

F.Y.R. MACEDONIA

BLACK SEA

GR

FORMER Y

MA

GHEG ALBANIAN

MACEDONIAN

M.R. = MEGLENO ROMANIAN

TURKISH

TURKISH

BULGARIAN

Turkey

MACEDO ROMANIAN

M.R.

PONTIC

GREEK

30°E

TOSK ALBANIAN

TURKISH

SLAVIC

PONTIC; SLAVIC

Italy

40°N

MACEDO ROMANIAN

GREEK

IONIAN SEA

GREECE

AEGEAN SEA

ARVANITIKA ALBANIAN

Turkey

38°N

ARVANITIKA ALBANIAN

GREEK

Athens

ARVANITIKA ALBANIAN

TSAKONIAN

ARVANITIKA ALBANIAN

TURKISH

36°N

MEDITERRANEAN SEA

GREEK

18°E   20°E   22°E   24°E   26°E   28°E

Multilingualism and place names

The complexity of the *onomastic scenario* in a given region reflects the glottodynamics of its history.

In multiethnic/multilingual communities, toponyms can appear in various versions (*allonyms*) according to the languages spoken there in past and present.

For example, the capital of Macedonia, *Skopje* in the official language of that state (*Скопje*), is also known as *Skoplje* (Croatian; Serbian: *Скопље*), *Shkupi* (Albanian) and *Üsküp* (Turkish).

The Ukrainian city of *L'viv* (spelled *Львів* in Ukrainian) appears as *L'vov* (*Львов*) in Russian, *L'voǔ* (*Львоў*) in Belorussian, *Lwów* in Polish, and as *Lemberg* in German and Yiddish (לעמבערג).

For the Ukrainian City of *Chernivtsy* (spelled *Чернівці* in Ukrainian), the following allonyms are in use: Russian *Chernovcy* (*Черновцы*), Polish *Czerniowce*, Romanian *Cernăuţi*, German *Czernowitz* (Yiddish: טשערנאָוויץ).

For toponyms with greater intercultural relevance, *exonyms* appear in addition to the local allonyms, thus *Skopje* is *Skópia* (Σκόπια) in Greek, and *L'viv* is *Ilyvó* in Hungarian, *Liov* in Romanian, *Lavov* in Croatian (Serbian: *Лавов*), *Léopol* in French, and *Leopoli* in Italian.

Exonyms may continue historical versions of a toponym. For example, the *endonym* of the capital of Serbia (and of former Yugoslavia) is *Beograd* (*Београд*), but its exonyms still display the syllable-final /l/ that has been changed to /o/ in Serbian (and Croatian), e.g. German *Belgrad*, English and French *Belgrade*, Greek *Veligradi* (Βελιγράδι).

Allonymic variation is a matter of degree: most of the examples cited above show various degrees of *weak allonymy*: the lexical basis is modified with respect to writing, phonology or morphology only.

In *strong allonymy*, the lexical basis of the name is different: either fully or partly translated, like older German *Griechisch-Weißenburg* and older Hungarian *Nándorfehérvar* for *Beograd*, or with no relation to the meaning of the name, like *Stalinabad* (literally: "Stalin-City") for *Dushanbe* (literally: "Monday" – referring to the weekly market) during the Soviet period.

Glottogenesis by accumulation of structural differences is an evolutive process that can be represented schematically in terms of a *family tree* (*dendrogram* or *cladogram*).

The pedigree model of glottogenesis has been developed in historical and comparative linguistics since the beginning of the 19th century and is still prevailing, in spite of serious shortcomings.

Genetic classification is the only firmly established scientific method of language classification.

(Picture by Thomas C. Moore, taken from *Scientific American*, German edition)
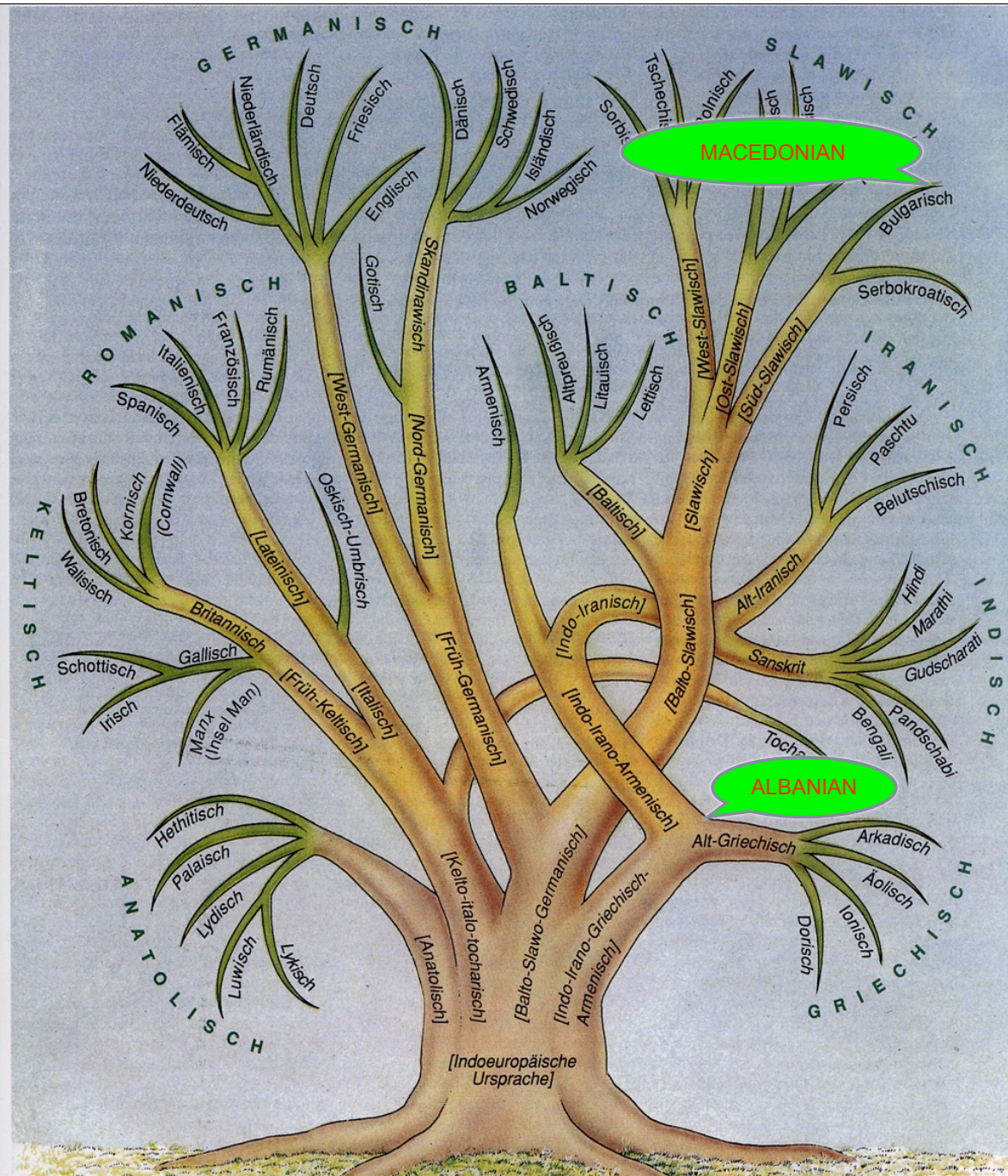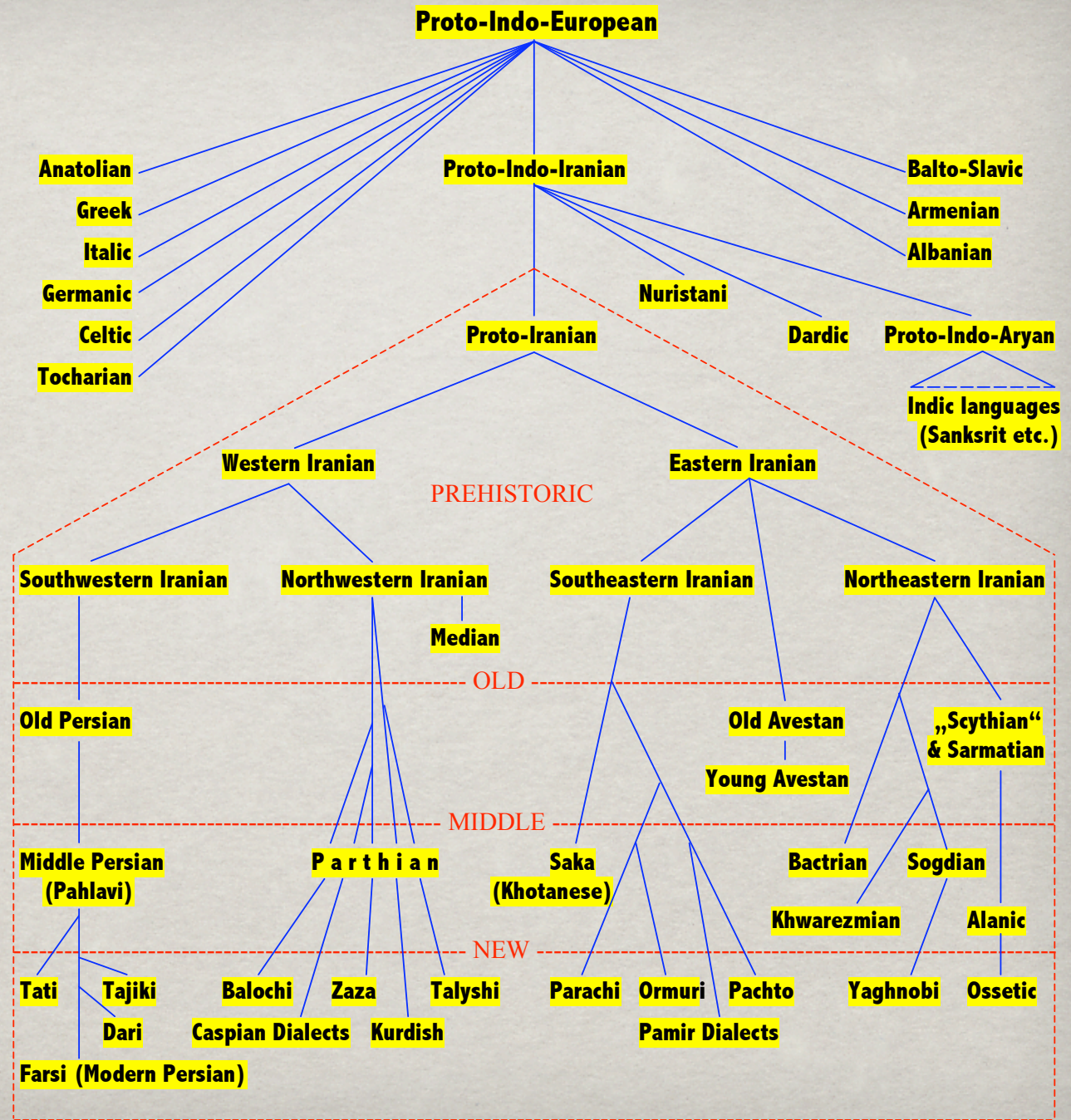
Iranian languages are attested since the 6th century BC (Old Persian inscriptions of the Achaemenid empire, in cuneiform script).
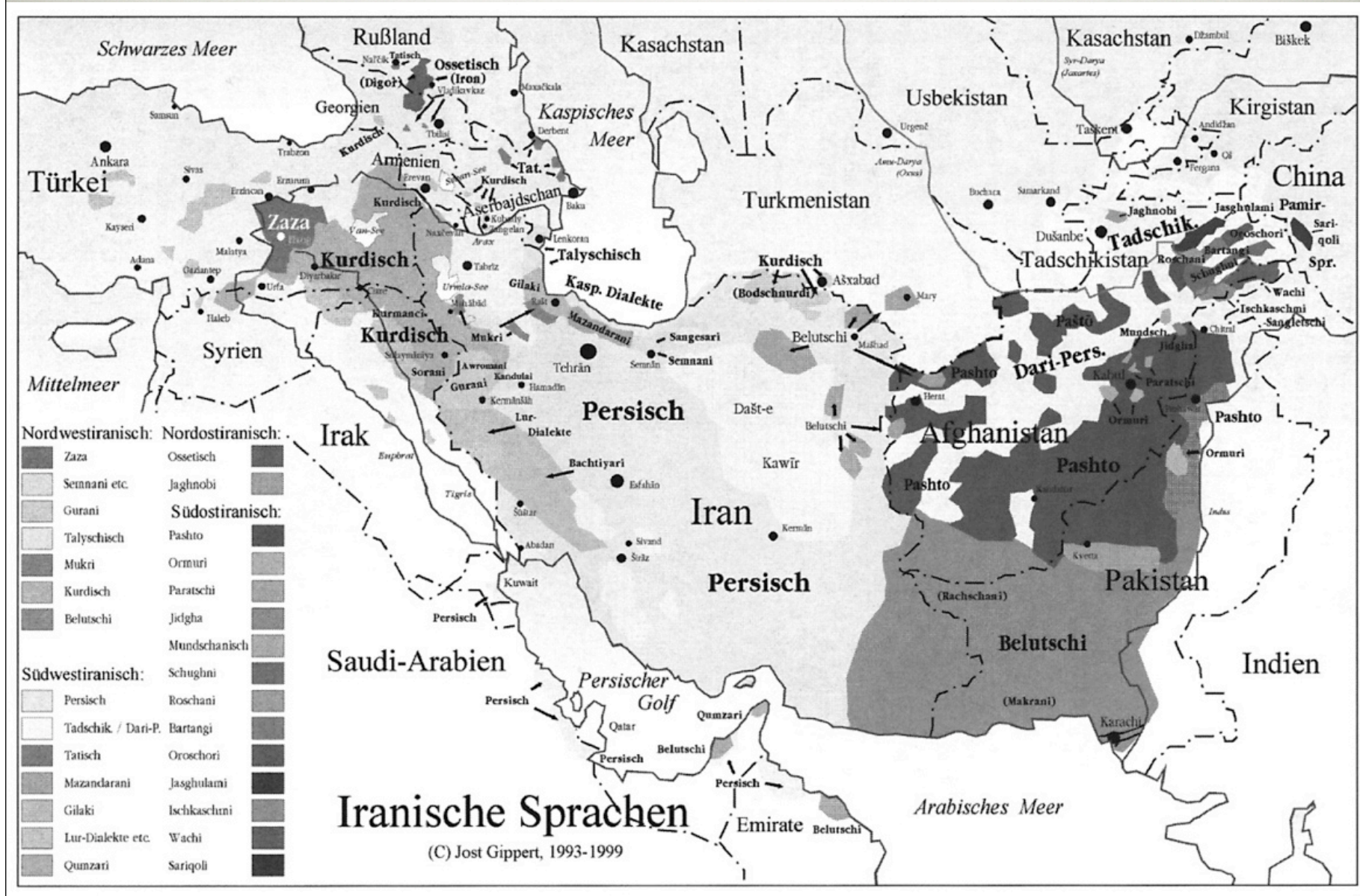
Avestan, the language of Zoroastrianism, was spoken in the early 1st millennium BC, but the written texts have been compiled, after many centuries of oral tradition, in the beginning of the Middle Iranian period (3rd to 9th centuries AD).

For Middle Persian, the Aramaic script and some of its offsprings have been used.

Today, most of the Iranian languages are written with (modified) Arabic script.

Proto-Indo-European

Anatolian

Greek

Italic

Germanic

Celtic

Tocharian

Proto-Indo-Iranian

Balto-Slavic

Armenian

Albanian

Nuristani

Proto-Iranian

Dardic

Proto-Indo-Aryan

Indic languages (Sanksrit etc.)

Western Iranian

PREHISTORIC

Eastern Iranian

Southwestern Iranian

Northwestern Iranian

Median

Southeastern Iranian

Northeastern Iranian

OLD

Old Persian

Old Avestan

„Scythian" & Sarmatian

Young Avestan

MIDDLE

Middle Persian (Pahlavi)

Parthian

Saka (Khotanese)

Bactrian

Sogdian

Khwarezmian

Alanic

NEW

Tati

Tajiki

Balochi

Zaza

Talyshi

Parachi

Ormuri

Pachto

Yaghnobi

Ossetic

Dari

Caspian Dialects

Kurdish

Pamir Dialects

Farsi (Modern Persian)

GEOGRAPHICAL DISTRIBUTION OF THE IRANIAN LANGUAGES

GEOGRAPHICAL DISTRIBUTION OF THE IRANIAN LANGUAGES

Northwestern • **Southwestern** • **Northeastern** • **Southeastern**

Iranische Sprachen

(C) Jost Gippert, 1993-1999

Example 2: Profile of Yaghnobi

Name: Yaghnobi, Yagnob, yaɣnobī́ zivók

Genetic affiliation: Northeastern Iranian, Indo-European

Geographic location: Yaghnob valley of the Zarafshan Region, Tajikistan; population resettled in the 1960s to Southern Tajikistan

Dialectal diversification: Western and Eastern, differing mainly in phonetics

Status: Minority language; diglossia (Yaghnobi being used in the private domain and Tajiki in public)

Writing system: Unwritten (for linguistic purposes a Latin-based transcription is being used); Tajiki and Russian are in use as literary languages

Number of speakers: 3 M (estimated, sources vary between 2 and 6 M)

Contact languages: Tajiki, Russian
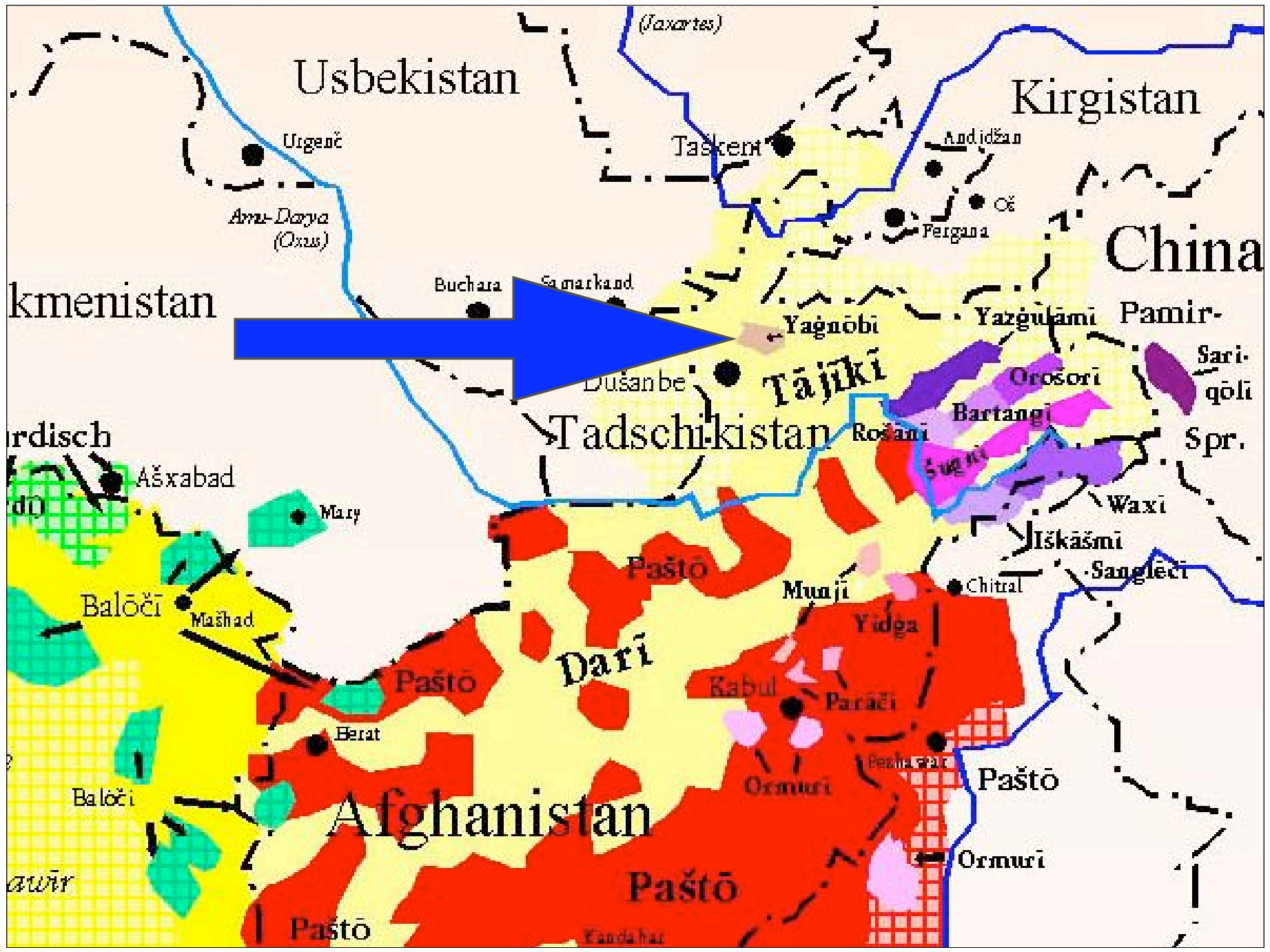
Example 3: Profile of Balochi

Name: Balochi, Baluchi, بلوچی

Genetic affiliation: Northwestern Iranian, Indo-European

Geographic location: Southeastern Iran (province of Sistan va Baluchestan, parts of Khorasan), Western Pakistan (province of Balochistan, partly in the provinces of Sindh and Punjab), Southwestern Afghanistan, a small part of Southern Turkmenistan, small regions in Oman (Masqat) and the UAE (Dubai)

Dialectal diversification: Western, Eastern and Southern dialect groups, treated as separate languages in some sources

Status: Minority language

Writing system: Arabic (modified, Urdu version in Pakistan), low medialisation and literacy rate

Number of speakers: 7,1 M^2 (Western dialects: 1,8 M^2; Southern dialects: 3,5 M^2; Eastern dialects: 1,8 M^2)
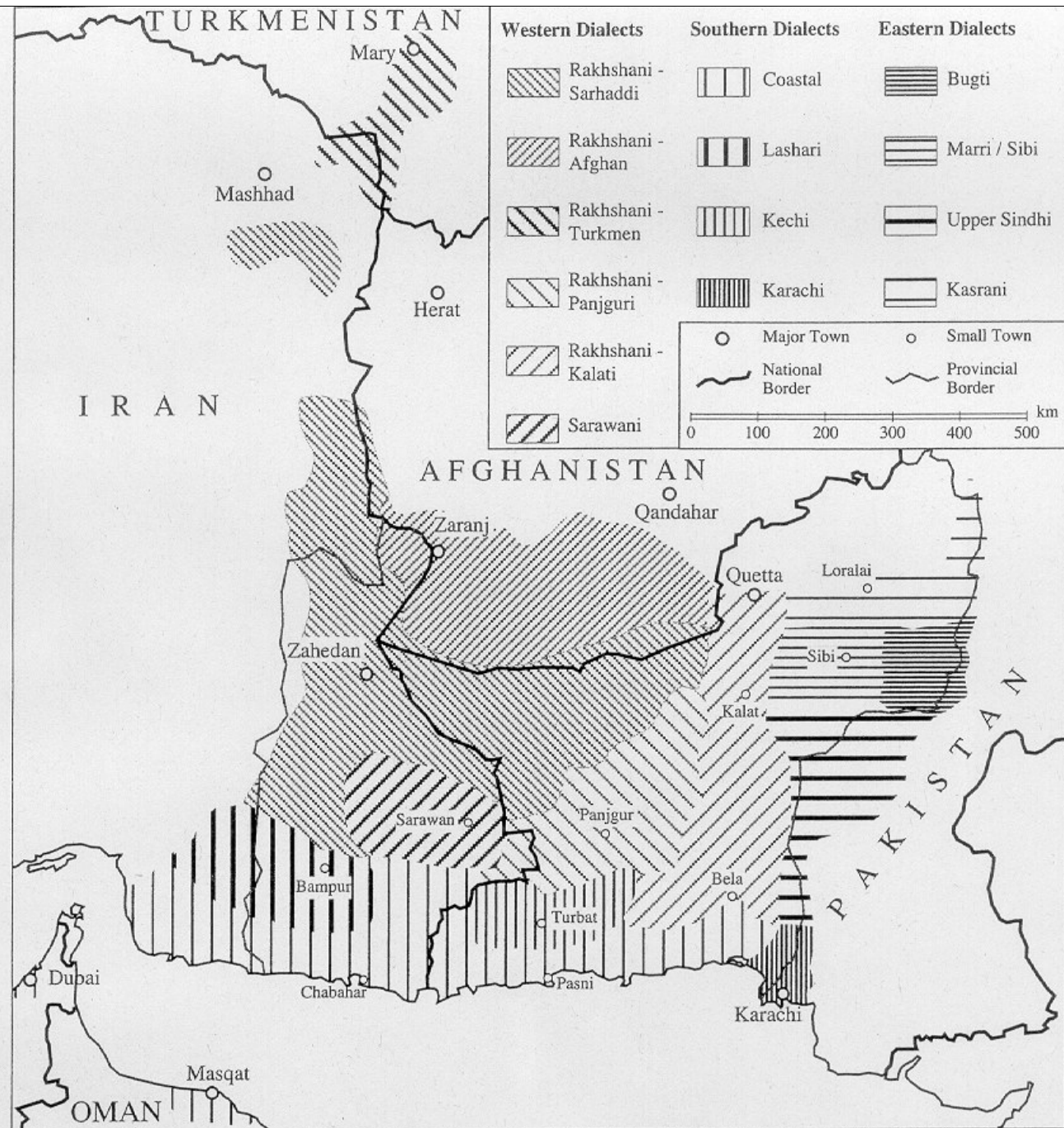
Contact languages: Farsi, Dari, Turkmen, Pashto, Urdu, Brahui, Saraiki, Sindhi, Punjabi, (Gulf) Arabic

Dialect areas include scarcely, only periodically or unpopulated areas, and also represent areas where a minority of the population speaks Balochi. Borders between dialect areas are often overlapping, and migration results in the presence of speakers of different Balochi dialects within one area.
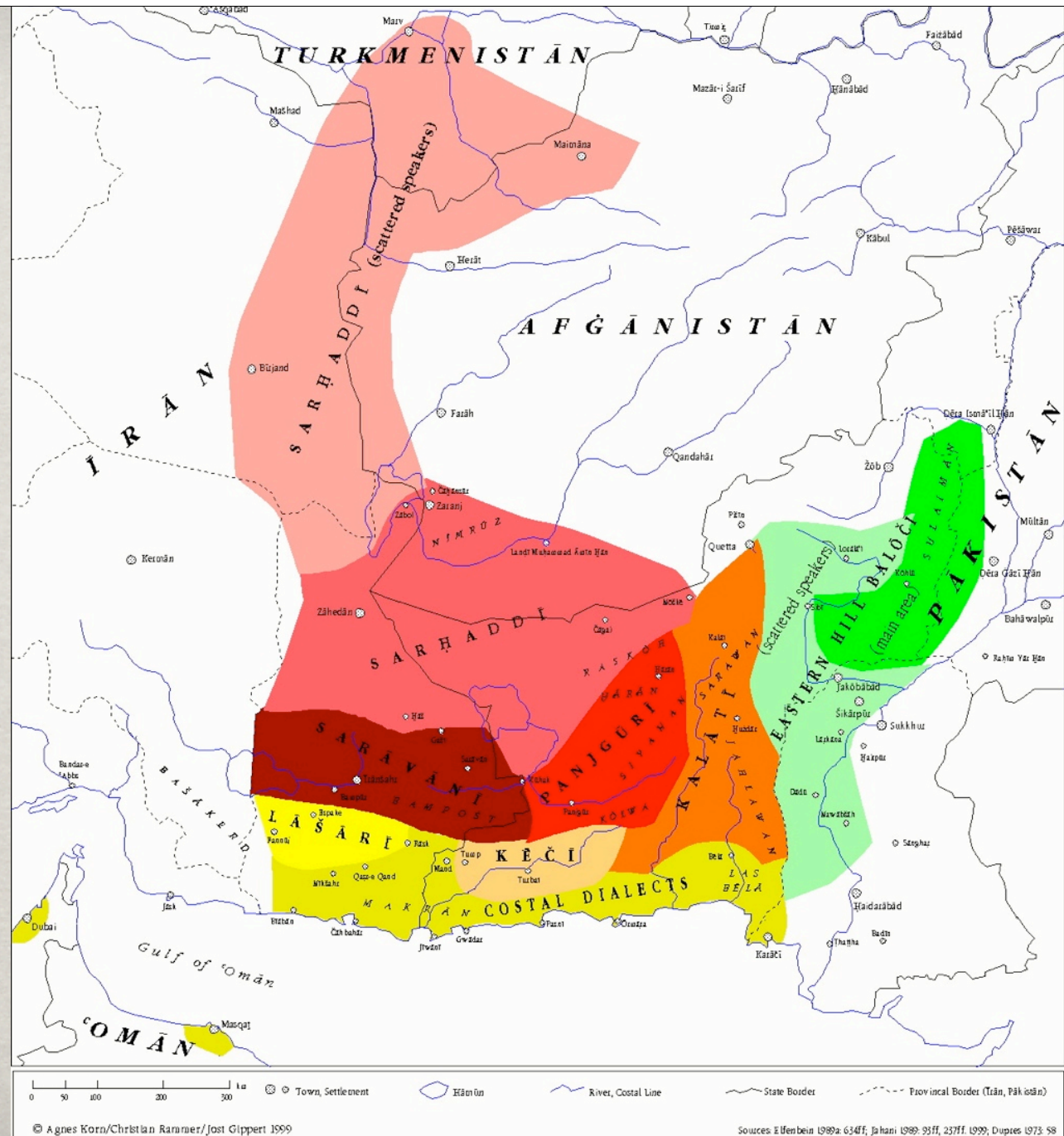From:
The Baloch and Their Neighbours. Ethnic and Linguistic Contact in Balochistan in Historical and Modern Times (eds.: Agnes Korn / Carina Jahani). Wiesbaden, Reichert, 2003.

TURKMENISTAN

Mary

Mashhad

IRAN

Herat

AFGHANISTAN

Qandahar

Zaranj

Quetta   Loralai

Zahedan

Sibi

Kalat

Sarawan   Panjgur

Bampur

Bela

Turbat

PAKISTAN

Dubai

Chabahar   Pasni

Karachi

Masqat

OMAN

Western Dialects

- Rakhshani - Sarhaddi
- Rakhshani - Afghan
- Rakhshani - Turkmen
- Rakhshani - Panjguri
- Rakhshani - Kalati
- Sarawani

Southern Dialects

- Coastal
- Lashari
- Kechi
- Karachi

Eastern Dialects

- Bugti
- Marri / Sibi
- Upper Sindhi
- Kasrani

- ◯ Major Town
- ◦ Small Town
- National Border
- Provincial Border

km
0   100   200   300   400   500

**APPROXIMATE
LOCATION OF
BALOCHI DIALECTS
(II)**

Dialect areas can be determined precisely for settled populations only.

In case of populations with nomadic way of life, linguistic cartography is pushed beyond its limits: instead of a map, a movie would have to be made in order to display the spatio-temporal distribution of ethnic groups and their respective language varieties.
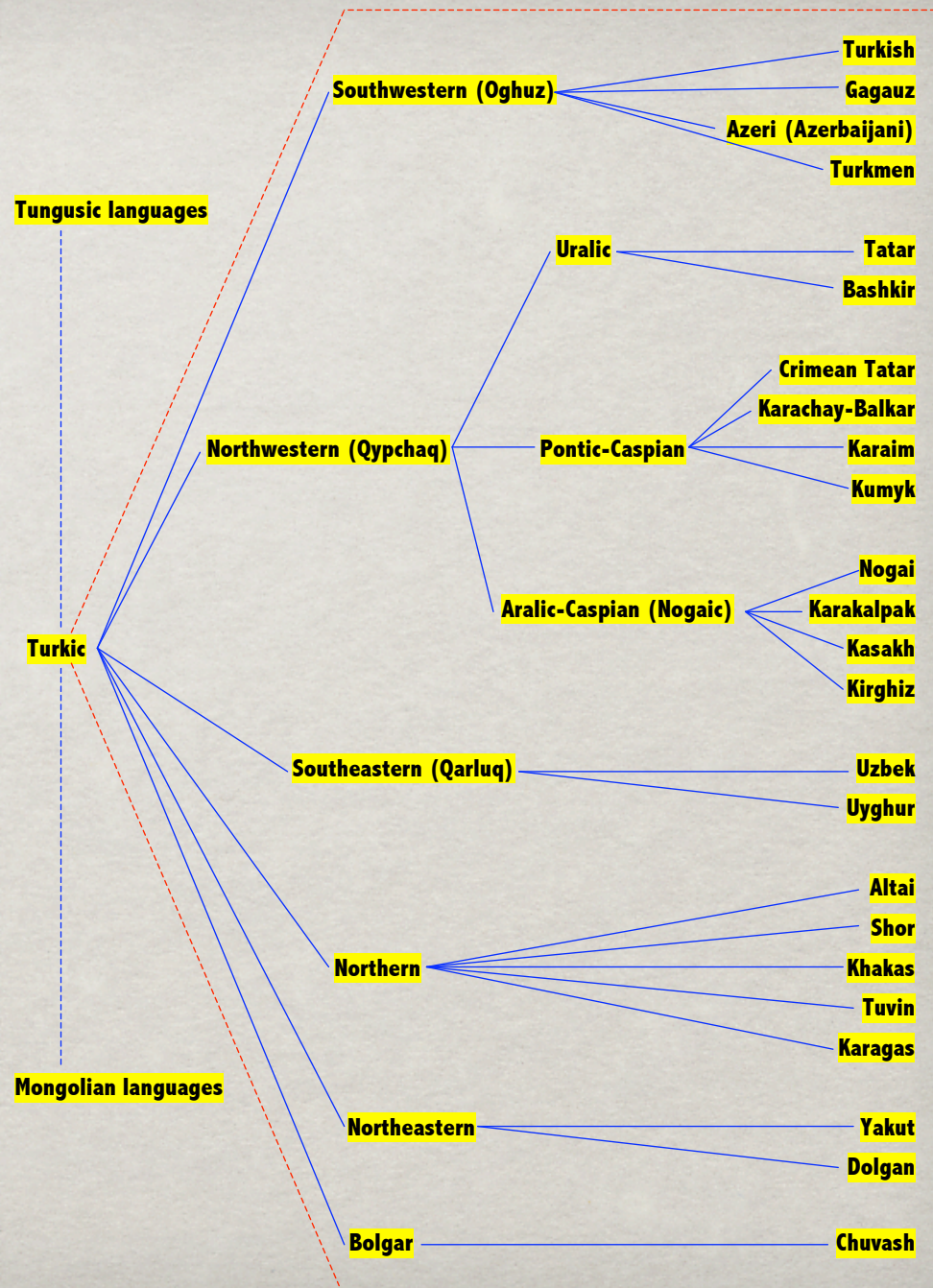

© Agnes Korn/Christian Rammer/Jost Gippert 1999

Sources: Elfenbein 1989a: 634ff; Jahani 1989: 93ff, 237ff. 1999; Dupree 1973: 58

Family-tree of the Turkic languages

The Turkic, Mongolian and Tungusic languages are usually subsumed under the common label "Altaic", although the genetic relationship between these three linguistic stocks is still beyond proof.
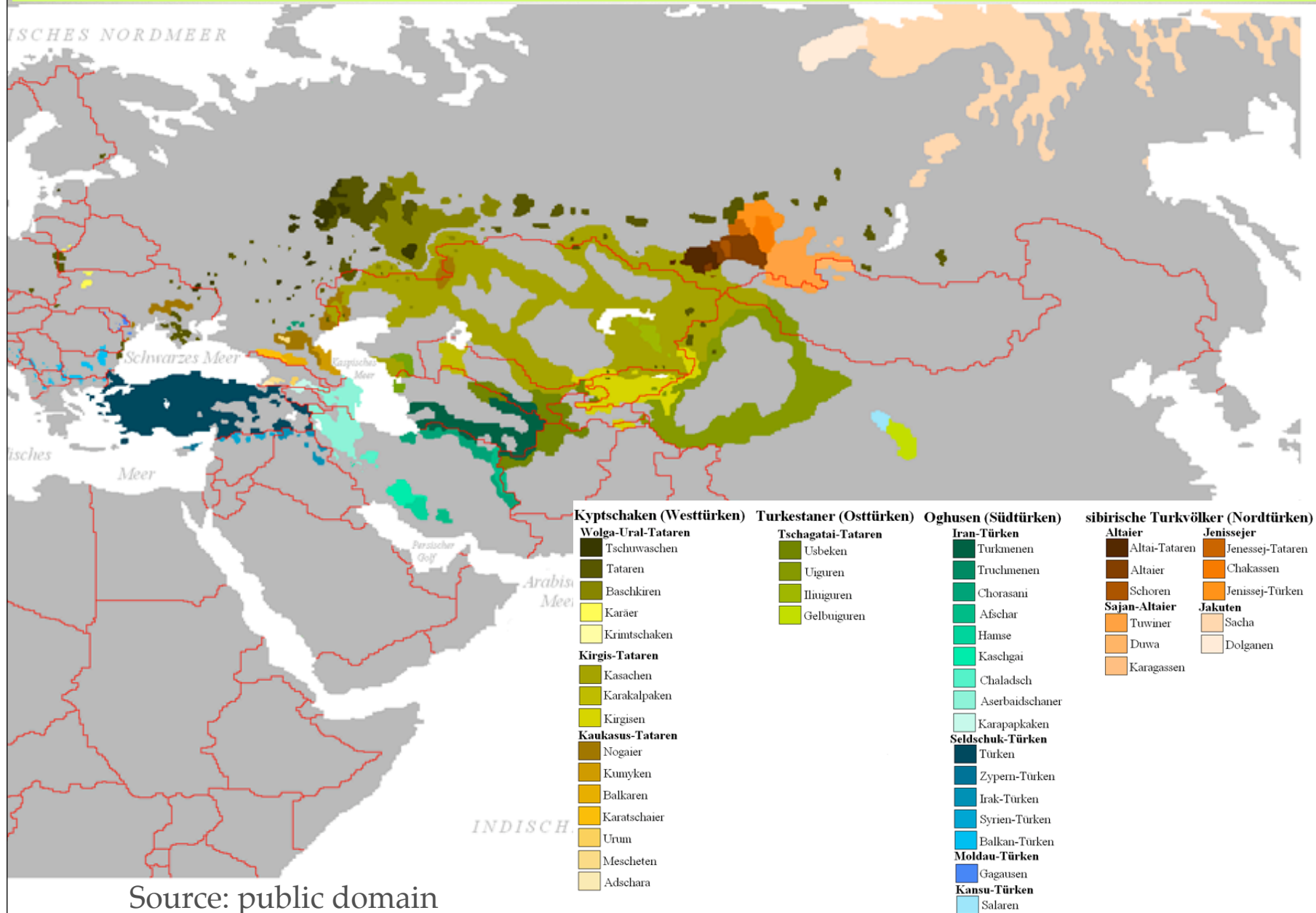
Turkic languages are attested since the 8th century AD (Old Turkic "runic" inscriptions of the Orkhon-Yenissey region and manuscripts of the Turanian region, as well as inscriptions and manuscripts in Uyghur, Sogdian and Manichaean script).

For many centuries, Turkic languages were written with the Arabic script. In Turkey, the Latin alphabet was introduced in 1928.

During the Soviet period, most of the Turkic languages were written with modified Cyrillic alphabets. Since 1990, a turn towards Latin-based orthographies has set in.

Tungusic languages

Mongolian languages

Turkic

- Southwestern (Oghuz)
  - Turkish
  - Gagauz
  - Azeri (Azerbaijani)
  - Turkmen
- Northwestern (Qypchaq)
  - Uralic
    - Tatar
    - Bashkir
  - Pontic-Caspian
    - Crimean Tatar
    - Karachay-Balkar
    - Karaim
    - Kumyk
  - Aralic-Caspian (Nogaic)
    - Nogai
    - Karakalpak
    - Kasakh
    - Kirghiz
- Southeastern (Qarluq)
  - Uzbek
  - Uyghur
- Northern
  - Altai
  - Shor
  - Khakas
  - Tuvin
  - Karagas
- Northeastern
  - Yakut
  - Dolgan
- Bolgar
  - Chuvash

GEOGRAPHIC DISTRIBUTION OF TURKIC LANGUAGES

Source: public domain

Example 4: Profile of Kyrgyz

Name: Kyrgyz, Kirghiz, Kara-Kirgiz, кыргыз тили

Genetic affiliation: Aralic-Caspian branch of Northwestern Turkic

Geographic location: Kyrgyzstan, Northern and Eastern Tajikistan, Afghanistan (Pamir Valley), China (Xinjiang), Eastern Uzbekistan, Southeastern Kazakhstan

Dialectal diversification: Northern and Southern

Status: Official language of Kyrgyzstan (together with Russian); minority language elsewhere

Writing system: Cyrillic (Arabic in China), fully medialised, high literacy rate

Number of speakers: 4 M² (3,6 M² in Kyrgyzstan, 180 M in Uzbekistan, 120 M in China, 70 M in Tajikistan, 30 M in Afghanistan, 20 M in Kazakhstan)

Contact languages: Russian, Dungan, Chinese, Tajiki, Kazakh, Uzbek

Languages of Kyrgyzstan and Tajikistan