

The background of the slide is a photograph of a stone wall, likely the Great Wall of China, winding across a lush green hillside. The scene is shrouded in a thick, white mist or fog, which obscures the distant parts of the wall and the surrounding landscape. The sky is overcast and grey. The overall mood is serene and atmospheric.

Hungarian Academy of Sciences (MTA)
21 November 2023

**New horizons on the typology of East and
Southeast Asian languages – and more
particularity of Sino-Tibetan and Sinitic
languages**

Alain PEYRAUBE
CNRS, Paris, France



Outline

- Classification of the languages of the world – An overview
- Sino-Tibetan and Sinitic languages
- Macro-families of ST and Sinitic
- Conclusion



Languages of the world – An overview

- Linguists traditionally group the 6000 to 6500 languages of the world – almost half of which will disappear in the course of the century - into 400 to 500 families, of very unequal size.
- Some of these, such as the Austronesian family, have more than 1200 languages, others have only one: Basque, for example, or Burushaski, well-known examples of linguistic isolates.



Languages of the world

- Europe, on the other hand, is rather poorly represented.
- The ALE (Atlas linguarum Europae) officially recognizes only 90 languages, grouped into 17 branches forming six families (phyla), of which Indo-European (IE) is by far the most important (12 branches according to Contini 2000).



Languages of the world

12 most widely spoken languages in the world (in millions of speakers)

- 1. Chinese 897, 2. Spanish 489, 3. English 371, 4. Hindi-Urdu 329, 5. Arabic 290, 6. Portuguese 283, 7. Bengali 282, 8. French 267,
- 9. Russian 154, 10. Punjabi 148, 11. Japanese 128, 12. Hausa 86



Languages of the world

Indo-European (10 or 12 branches)

- 1. Germanic languages, 2. Celtic languages, 3. Slavic languages, 4. Baltic languages (Balto-Slavic languages), 5. Romance languages,
- 6. Greek, 7. Armenian, 8. Albanian, 9. Indic, 10. Iranian (Indo-Iranian), 11. Tokharian, 12. Anatolian



Languages of the world

Eleven Romance (Latin) languages

- 1. French, 2. Spanish, 3. Italian, 4. Portuguese, 5. Romanian
- 6. Occitan, 7. Catalan, 8. Galician, 9. Rhaeto-Romance (Rhaetian), 10. Corsican, 11. Sardinian



Languages of the world

- Hungarian, (or Magyarul), remains the only major language in Central and Western Europe... that doesn't come from Europe. Compared to the Indo-European languages - from which French, English, Spanish, German and the Balto-Slavic languages are derived - Hungarian belongs to the Finno-Ugric language family, as do Finnish and Estonian.
- Hungarian is an agglutinative language, which means that it adds affixes - prefixes, suffixes or infixes - to words to express nuances and grammatical relationships.
- It is also characterised by its vowel harmony.



Languages of the world

- Until now, linguists have tended to try to reconstruct proto-languages (the oldest of which are thought to date back no further than 6000 BCE) for the language families in which they specialise. Few have attempted in the past to compare different language families. This methodological bias has changed considerably since the 1990s..
- Although, hypotheses of large groupings were regularly put forward throughout the 20th century, but they were never really taken seriously, such as that of Trombetti (1905) who thought that all languages should belong to a single family.



Languages of the world

- Also at the beginning of the 20th century, the Danish linguist Pedersen (1903) suggested linking Indo-European (IE) to other families such as Uralic (Finno-Ugric and Samoyed languages), Altaic, Eskimo-Aleut and Semitic, within a macro-family that he called Nostratic.
- However, it was not until the early 1960s that this proposal was seriously taken up and developed by the Russian linguistic school (Dolgopolski and Illitch-Svitytch), which also included Kartvelian, Dravidian and Chukchi-Kamtchatkian in Nostratic.
- However, this nostratic thesis has never gained acceptance in the international linguistic community.



Languages of the world

The same was true of Austric, which originally (Schmidt's hypothesis in 1906) included Austronesian and the Mon-Khmer or Austroasiatic languages. The existence of an Austric family is still controversial, as is Benedict's (1942) proposal to group Tai-Kadai and Austronesian into an Austro-Tai family.



Languages of the world

The situation changed somewhat in the early 1980s. Greenberg's (1963a) proposal to reduce the diversity of African languages to four macro-families was accepted by a large majority of specialists, after having been the subject of fierce criticism by Africanists.

This was also the period when the Russian linguists Starostin and Nikolaïev (Starostin 1989) presented evidence in favour of similarities between Proto-Caucasian, Proto-Sino-Tibetan and Proto-Ienisean.

In a second phase, the Caucasian family was linked to the Na-Dene family, which had been discovered by Sapir but to which no one had previously paid much attention, were linked (Ruhlen 1994): Dene-Caucasian macro-family.

At the end of the 1980s, the American linguist Bengston (1991) was also convinced that Basque and Burushaski, as well as Sumerian, could be related to Dene-Caucasian.



Languages of the world

- Finally, in 1987, Greenberg reduced the two hundred or so independent families of the Americas (including about sixty for the languages of North America, which, according to Chafe 1962, grouped 213 languages) to just three families, Amerind/Amerindian, Althabaskan and Eskimo-Aleut, the last two of which, moreover, belonged to larger groups.
- Later, like the Russian linguists who supported the nostratic approach, Greenberg thought that IE was clearly related to other Eurasian families, such as Uralic, Altaic and Eskimo-Aleut, and he spoke of a Eurasiatic macro-family.



Languages of the world

- On the basis of these various proposals, in 1992 Ruhlen identified a dozen macro-families that would encompass the 6,000 to 6,500 languages of the world.
- The twelve macro-families are as follows: Khoisan, Nilo-Saharan, Nigero-Kordofanian, Afroasiatic, Kartvelian, Dravidian, Eurasiatic, Dene-Caucasian, Austric, Indo-Pacific, Australian and Amerindian. They are grouped into major geographical areas:



Languages of the world

Africa (4)

- Khoisan (South Africa, Tanzania)
- Nigero-Kordofanian, which occupies a vast area of sub-Saharan Africa and is made up of two main branches: Kordofanian (southern Sudan) and Niger-Congo (with hundreds of Bantu languages: Zulu, Swahili, Mbundu, etc.).
- Nilo-Saharan, a group of languages spoken in northern Central and East Africa, with several sub-families, including Nilotic.
- Afroasiatic, which includes Semitic (Arabic, Hebrew), Chadic (Hausa, etc.), Berber, Old Egyptian, Omotic (Kafa, Mocha,) and Cushitic (Afar, Somali).



Languages of the world

Southeast Asia and Oceania (3)

Austriac is the only macro-family present in Southeast Asia. It includes Austroasiatic (composed of the Munda languages of northern India and the Mon-Khmer languages of Vietnam and Cambodia), Miao-yao (in southern China and Vietnam), Kra-Dai or Tai-kadai (in Thailand and Laos) and Austronesian (Taiwan, Malaysia, Indonesia, Philippines, Madagascar, New Zealand, Tahiti, etc.).

Three macro-families are found on the Oceania continent: Austronesian (a sub-family of Austriac, see above), **Indo-Pacific** (in Papua New Guinea, which has almost 800 languages) and **Australian** (over 200 languages).



Languages of the world

Americas (1)

Amerind/Amerindian, a single macro-family divided into eleven sub-families and including, among others: in North and Central America, Almosan, Algonquin, Uto-Aztecan, Central Amerindian, etc.; in South America, Andean, Arawak, Macro-Tucano, Macro-Caribbean, etc.

Many other languages of the Americas are grouped into two other macro-families: Eskimo-Aleut (in Alaska, a sub-family of Eurasian) Na-Dene (a subfamily of Dene-Caucasian, which includes languages of the Athabaskan family and other languages of the southern coast of Alaska).



Languages of the world

Eurasia (4)

- Dravidian (South India): Tamil, Brahoui ,
- Kartvelian (Georgia)
- Eurasiatic, which includes Indo-European (Anatolian, Romance languages, Germanic languages, Tokharian, etc.), Uralic (some twenty-five Finno-Ugric languages [Finnish, Hungarian, Lappish, Estonian], as well as Samoyed languages), Altaic (divided into three branches: Turkic, Mongolian and Manchu-Tungusic), a Korean-Japanese-Ainu group, Chukchi-Kamchatkian (in Northern and Eastern Siberia), Eskimo-Aleut (Greenlandic).
- Dene-Caucasian



Languages of the world

- **Dene-Caucasian,**
which includes:
 - Basque (Pyrenees),
 - Caucasian (including Chechen),
 - Burushaski (mountains of North Pakistan),
 - Ienisean (the Ket spoken in central Siberia),
 - Sino-Tibetan, Na-Dene.



Languages of the world

This classification is far from being the subject of even a vague consensus among linguists. While Greenberg's four macro-families of African languages are now accepted by a large majority of specialists, the same cannot be said of his new hypotheses on Eurasiatic, Dene-Caucasian or Amerindian.

They have been heavily criticised by Indo-Europeanists, but also by Americanists who continue to believe that linguistic change is so rapid that after about 6,000-7,000 years all traces of earlier relationships have been erased by incessant phonetic and semantic erosion.



Languages of the world

- Despite the controversy and lack of certainty, however, the concept of the ‘unifiers’ (the American school of Greenberg and Ruhlen, and the Russian school of Dolgopolski and Starostin) is paradoxically gaining increasing support.
- Some have even suggested going a step further and grouping macro-families. Relationships could be established between Amerindian and Eurasian, as suggested by Ruhlen (1992, 1997). The idea that Austric, Caucasian-Dene and even Indo-Pacific might be related is also not far-fetched either. This could lead to a single proto-language.



Languages of the world

- The greatest challenge to linguistic typology would, of course, remain the organisation of this single proto-language into sub-macro-families.
- At present, the existence of macro-families, let alone a single proto-language, is far from being accepted.
- A study of the situation as it stands in East Asia and Southeast Asia clearly shows the variety of possible groupings and the difficulty of having indisputable arguments in favour of one of them.



East Asia and Southeast Asia

- This vast geographical area is considered to encompass five major language families:
- Sino-Tibetan
- Austronesian
- Austroasiatic
- Tai-Kadai or Kra-Dai
- Miao-yao (Hmong-Mjen).



Sino-Tibetan

- Sino-Tibetan (ST) is dated to around 4500 BCE and is generally divided into two branches that diverged 6000 years ago, at the time when archaeologists identify the formation of what they call "early China" (Chang 1986, Wang 1996) : the Sinitic languages and the Tibeto-Burman (TB) languages.
- The term Sino-Tibetan was coined and introduced by Przyluski in 1931 to replace the the Indo-Chinese theory.



Sino-Tibetan

- Although certain genealogical affinities between Chinese and Tibeto-Burman are often undeniable (a common lexicon is clearly shared in both branches, comprising a large part of the basic vocabulary and showing regular phonological correspondences), it must be acknowledged that work in the field of Sino-Tibetan linguistic comparison is still at a stage that does not allow the Sino-Tibetan hypothesis to be unambiguously validated.



Tibeto-Burman

- TB groups together some 250 languages (65 million speakers), including Tibetan, for which the first written documents date from the 7th century, Burmese (for which the first attested texts date from the 12th century), Lolo, Jingpo and others.
- The connection between Burmese and Tibetan dates from the 18th century.
- However, a Han document (ca. 2nd century) contains a lexicon of 155 words in a language called *bailangge*, clearly Tibeto-Burman: 110 of these words are considered to be ancient Burmese and 40 of them ancient Tibetan.



Tibeto-Burman

- These languages are spoken in a region that includes the Himalayas of Pakistan, India, Nepal and Bhutan, as well as the Tibetan plateau down to Sichuan and Yunnan provinces in China, the Assam valley, Burma (Myanmar) and northern Thailand to northwestern Vietnam.
- Without going into the details of the current discussions and controversies about Sino-Tibetan, let us mention just one of the competing hypotheses, that of van Driem (2001), who claims that there is no reason to consider the Sinitic languages as not belonging to Tibeto-Burman, in which case Sino-Tibetan simply does not exist.
- He suggests that the Sinitic languages are a branch of Tibeto-Burman, or even a sub-sub-sub-branch, according to the diagram he proposes below:



Tibeto-Burman

Western (Brahmaputran et al.)

Eastern

North (Sino-Bodic)

Northwest (Bodic)

Bodish

Himalayan

Northeast (Sinitic)

South

Great South (Burman, Karen)

Lolo-Burman

Karen

Central (Qiang, Xifan)



Tibeto-Burman

- The typological diversity of this group of languages is considerable. There are tonal and atonal languages, monosyllabic and polysyllabic, isolating and inflected, verb-final and verb-medial, and so on.
- Benedict's (1972) original classification, which differs from van Driem's, still has many adherents. He divided the Tibeto-Burman languages into seven groups: Tibetan-Kanauri; Bahing-Vayu; Abor-Miri-Dafla; Kachin (Jingpo); Burman-Lolo; Bodo-garo; and Kuki-Naga.
- More recently, Thurgood (2003) also distinguished seven distinct groups: Lolo-Burman; Bodo or Tibetan group; Sal languages (Bodo, Konyak,); Kuki-chin-naga group; rung branch (rGyalrong, dulong, kiranti, qiang languages); Karenic group; other small sub-groups (for tani, hrusish, idu-digaru, miju-kaman languages).



Tai-kadai

- Tai-Kadai (or Daic) comprises the Tai or Thay-Yay languages, according to the terminology of Ferlus 1990, which clarifies that of Haudricourt 1972, and the Kadai languages.
- They cover a vast arc of Southeast Asia, stretching from the provinces of Guizhou, Guangxi and Yunnan in China, northern Vietnam and northeastern Burma to the upper Brahmaputra valley in Assam, India. The family includes Thai, of course, whose oldest written documents date from the late 13th century, but also Lao, Zhuang, Buyi and the Kam-Sui languages (Kam, Sui, Mulao, Maonan, etc.) of southern China.



Tai-kadai

- At the beginning of the 19th century, the Tai languages were considered part of a large group known as Indo-Chinese, which by the end of the century had been reduced to two branches, one of which was Sino-Tai.
- It was not until the middle of the 20th century that the Tai languages ceased to be associated with Chinese. It was undoubtedly the absence of affixal morphology and the presence of tones that led to their genealogically association to Chinese.
- Kadai languages (Benedict 1975) such as Kelao, Lati, Laqua, Laha and Li (Hlai) share several features with Austronesian languages.



Austroasiatic

- The Austroasiatic family, identified around 1880, comprises about 130 languages. It covers the whole of continental Southeast Asia, including the Malay Peninsula, part of southern China, eastern India and the Nicobar Islands.
- It is divided into two branches: the Munda languages of northern India and the Mon-Khmer languages, spoken mainly in Vietnam and Cambodia. Today, these two branches have relatively few typological features in common. In fact, their morphosyntactic types are diametrically opposed. Munda presents the SOV (subject - object - verb) constituent order of the Indian linguistic area, while Mon-Khmer shows an inverted image known as SVO or even VSO.
- According to Gérard Diffloth (2005), Austroasiatic could date from 4,000 to 5,000 years BC, with its origins in Southeast Asia, perhaps even in eastern India, but certainly not in China.



Austronesian

- Austronesian is thought to date back to around 5,000 years BC. It is one of the oldest families identified, which includes more than 1,200 languages, or about 20% of the world's languages.
- However, there are no written records of any of these languages before 670. These languages cover an extremely wide area, from Madagascar and Mayotte in the west to New Zealand, Melanesia, Polynesia, Micronesia and Easter Island in the east, in the Pacific.



Austronesian

- Robert Blust (1997) points out that Austronesian has ten branches, nine of which, with only 26 languages are found in Taiwan, the Formosan languages spoken by the island's aborigines, who make up 2% of the population and are the cradle of all Austronesian languages: Ayatal, Saisiat, Bunun, Tsou, Rukai, Paiwan, Ami, Puyuma and Yami.
- The tenth branch, known as Malayo-Polynesian, is thought to comprise more than 1,100 languages, from Madagascar to Eastern Polynesia. According to Laurent Sagart (2005), there are only six branches, all of which are attested in Taiwan, and only one of which (Malayo-Polynesian) is also present elsewhere.



Miao-Yao

- The Miao-yao (or Hmong-Mjen) family comprises about thirty languages, which can be divided into two groups: the Hmong or Miao languages (Hmong from Guizhou, Yunan and Indochina, Hmn from western Guizhou, Qo Xiong from western Hunan) and the mjen or yao languages (mjen, mun, tsao min).
- Finally there is also the family of Altaic languages in China.



Altaic languages

As we have already seen in the section on Europe, the Altaic languages comprise three groups: Turkic, Mongolian and Tungusic-Manchu. These Altaic languages (the name comes from Altai ‘gold’, a mountain in western Mongolia) are spoken by populations that form a continuum from the Balkans to the Pacific Ocean.

In China, they are spoken in the north in Manchuria (Heilongjiang province), in Inner Mongolia and in the north-west in the provinces of Gansu, Qinghai and Xinjiang. The Mongolian languages are Oirate, Santa or Dongxiang, Bonan, Monguor, Eastern Yugur or Yellow Yugur; the Turkic languages are Uyghur, Kazakh, Kyrgyz, Salar and Western Yugur; and the Tungusic-Manchu languages are Xibe. Cf. Peyraube (2015).



Macro-families

- The reason why we have singled out this geographical zone, 'East Asia and South-East Asia', from the large group of two continents, Asia and Oceania, is that various groupings into macro-families have been proposed over the last century.
- Almost every combination has been tried. Some proposals are now obsolete (e.g. the Sino-Tai family), but over the last twenty years archaeological, genetic and anthropological evidence has accumulated some evidence for linking several East Asian and Southeast Asian language families.



Macro-families

- Four or five scenarios for several macro-families are currently under discussion:
- Austro-Tai
- Austric
- Sino-Caucasian
- Sino-Austronesian
- or even Sino-Tibetan-Austronesian.



Austro-Tai

- As far as the Austro-Tai are concerned, Schlegel (1901) had already noted the relationship between the Tai languages and the Austronesian languages. But it was Benedict, in 1942, who actually proposed the existence of an Austro-Tai macro-family consisting of two branches: Tai-Kadai and Austronesian, to which Miao-yao was sometimes added later.
- This hypothesis is still defended today by Ostapirat (2005), who has found regular correspondences between the common lexicon of Tai-Kadai and Austronesian. It has also been criticised, notably by Thurgood (1994), who argues that Austronesian vocabulary in Tai languages has been borrowed. No date is given for this Austro-Tai, but the macro-family could date back to 6000 or 7000 BCE.



Austric

- Austric is another hypothesis put forward long ago by Schmidt (1905-1906), who grouped the Austroasiatic and Austronesian families together.
- It has now been taken up by Blust (1996) and Lawrence A. Reid (1994), who have shown a number of similarities between the two families, both lexically and morphologically. Others add the Miao-yao and Tai-Kadai languages (Ruhlen 1997), but Diffloth (2005) argues that these similarities are far from convincing. Since he places Proto-Austroasiatic far to the south of the Southeast Asian subcontinent, or even in India itself, this makes the Austric hypothesis rather implausible.



Dene-caucasian

After identifying a large number of lexical items in common between Sino-Tibetan, North Caucasian and Yenisean (Ket language), which do not belong to the 'East Asian and South-East Asian' zone, Starostin proposed, at the beginning of the 1980s, the existence of a macro-family, Sino-Caucasian (Starostin, 1989). Later, John D. Bengston (1991) added Basque, Bourouchaski and Na-dene to make up the macro-family now called Dene-Caucasian by Ruhlen (1997).



Sino-austronesian

- Finally, in the early 1990s, Sagart took up a suggestion from the late 19th century that Chinese and Austronesian might be related languages (Sagart, 1993).
- In fact, he notes regular phonetic correspondences as well as a vocabulary and morphology common to these languages, which would tend to prove that they are genealogically related.
- He therefore assumes the existence of a macro-family, Sino-Austronesian, divided into two branches: the Sinitic languages and the Austronesian languages. Tibeto-Burman is left out of the equation, a position that will come in for some criticism.



STAN

- He later went back on this hypothesis, recognising the integrity of Sino-Tibetan (with its two branches, Sinitic languages and Tibeto-Burman languages) and proposed a new macro-family: Sino-Tibetan-Austronesian (Sagart, 2005). The Tai-Kadai languages are also finally included in this macro-family, which is thought to date from around 6,500 BCE, at the same time as the domestication of cereals in the Yellow River valley.



PEA (Proto-Eastasian)

- The links presumed to exist between the various language macro-families, in the hypotheses mentioned above, have naturally led scholars to envisage the possibility of the existence of a common ancestral language.
- Straosta (2005) took this a step further by calling this macro-macro-family Proto-Eastasian (PEA).
- It is thought to have been spoken in central China, around the Han River and the Yellow River, between 10,000 and 8,000 BCE. It comprises just two branches: Sino-Tibetan-Austronesian (including Tai-Kadai) and Austroasiatic. There is therefore no need to recognise an Austric macro-family..



PEA

This hypothesis of a single macro-macro-family for East and Southeast Asia remains pure speculation for the time being. However, it is still being discussed in the context of typological studies, or even areal linguistics, which tend to use syntactic criteria rather than phonetic or lexical criteria.

A number of syntactic properties have been identified that are common to most of the languages in this vast area (Matisoff, 1991; Enfield, 2002; Chappell and Peyraube, 2006). These grammatical features, however, serve to characterise a zone, a linguistic area. They cannot be considered as solid criteria for a properly genealogical classification of languages.



Sinitic languages

- It is generally considered that there are 10 Sinitic languages in China, namely the 7 major traditional dialect groups (**Wu**, **Min**, **Kejia**, **Yue**, **Xiang**, **Gan** and **Mandarin**), to which have been added since the late 1980s **Jin**, which is spoken in Shanxi, **Pinghua**, which is spoken in Guangxi, and **Hui**, which is spoken in Anhui. (Atlas of the languages of China, 1987, 2008 and 2012).
- **Guanhua** (Mandarin) is further subdivided into 7 or 8, depending on the author, dialect groups (Li Rong 1985, Liu Xunning 1995).



Mandarin

Guanhua (Mandarin) is in turn divided into eight groups:

- **Beijing Mandarin**, spoken in the regions of Beijing, Hebei, Inner Mongolia, Liaoning and Tianjin, the most important of which is the Beijing dialect, which forms the phonological basis of Standard Chinese
- **Northeastern Mandarin**, spoken in the provinces of Liaoning, Jilin and Heilongjiang. It is a dialect of Mandarin close to Beijing Mandarin
- **Mandarin Jilu**, spoken in Hebei and Shandong provinces.



Mandarin

- **Jiaoliao** Mandarin, spoken in the Jiaodong (Shandong province) and Liaodong (Liaoning province) peninsulas, as well as in certain coastal areas of Heilongjiang province.
- **Zhongyuan** or Central Plain Mandarin, spoken in parts of Henan and Shaanxi provinces in the Yellow River valley, and as far east as southern Xinjiang (East Turkestan) in northwest mainland China.



Mandarin

- **Lanyin Mandarin**, spoken in the province of Gansu (capital Lanzhou) and in the autonomous region of Ningxia (capital Yinchuan), as well as in Xinjiang.
- **Jianghuai Mandarin** or Lower Yangtze Mandarin, spoken in the provinces of Jiangsu and Anhui, on the north bank of the Yangtze, as well as in certain regions on the south bank, such as Jiangsu including Nanjing, Jiangxi including Jiujiang.
- **Southwest Mandarin** or Upper Yangtze Mandarin, spoken between the Yangtze valley and the eastern Tibetan plateau: city of Chongqing, provinces of Hunan, Hubei, Sichuan and Yunnan.



Mandarin

- Of these eight dialect groups, the first 4 can be identified as part of Northern Mandarin, which was the basis of Putonghua or Standard Mandarin, whose official definition, promulgated in 1955, is as follows:
- It has as its standard pronunciation the Peking pronunciation, as its basic dialect the Northern dialect, and as its grammatical model the exemplary literary works written in the modern colloquial. (translation by Ramsey 1987)



Mandarin

- Classification of the 7 Sinitic languages (Ramsey, 1987 : 87)
- Classification of the 10 Sinitic (Language Atlas of China 1987/2008, 2012)
- Classification of the 8 Mandarin varieties (Li Rong 1985, Liu Xunning 1995)





Mandarin

- It was only around 1850 that the common language of China really shifted from the Mandarin of Nanjing to the Mandarin of Peking, as documentary accounts and grammars of the official Chinese language by Edkins (1857) or Wade (1867) indicate.
- In 1909, the Qing dynasty adopted the Beijing dialect as its *guóyǔ* 'national language'.
- To come back to all these classifications,



Different systems of classification

- There are different systems for classifying languages. There are three basic approaches:
- the genealogical approach, which leads to the construction of trees (*Stammbaum* model),
- the areal approach, which focuses on geographical contacts and leads to reasoning in terms of extension and overlap of areas,
- the typological approach, which looks for similarities between linguistic structures, while also highlighting the significant divergences (Szulmajster-Celnikier 1998).



Different systems

- The **genealogical approach** is still considered superior to the others (Greenberg 2001).
- The so-called areal classification has led to revisions of certain reconstructions. For example, the unity of the Uralo-Altaic (Uralic and Altaic languages) is no longer recognised. The same is true of Chamito-Semitic (Semitic languages and certain Saharan and East African languages, including Egyptian).



Different systems

- In the **typological approach**, languages are classified into types on the basis of similarities in linguistic structure, usually syntactic and semantic. We can thus distinguish between languages with verbo-nominal distinction (IE); languages with ergative construction, formally marking the agent (Basque or South Caucasian languages); languages with accusative construction, formally marking the patient, etc. Cf. Hagège (1982, pp. 39-40).



Different systems

- The old classification into agglutinative languages (made up of identifiable radicals and affixes), inflectional languages (with conjugated verbs and declensional nouns) and isolating languages (often monosyllabic and tonal) is a classification by type.
- Developed in the 19th century, this typological approach has evolved considerably. It has shown that, in many places, great genealogical diversity is accompanied by common typological features, in line with the results obtained by the areal classification method.



Languages and genes

- Population genetics has come to lend a hand to linguists in their endeavor to classify languages.
- As soon as geneticists began to systematically analyze DNA in the 1980s, data accumulated considerably, at the same time as biochemical and molecular techniques developed, as well as theoretical models of evolution necessary for the interpretation of genetic material in a historical context



Languages and genes

- Correlations between genetic distance - a central concept in population genetics - and linguistic distance were then attempted. And close correspondences were found between the genetic classification of populations and that of macro-families of languages, as proposed by Greenberg and Ruhlen:
- Greenberg et al. 1986 on American populations, Excoffier et al. 1987 on sub-Saharan Africa, Sokal et al. 1988 and Barbujani et al. 1990 on Europe, Poloni et al. 1997 on several African and European populations, and above all Cavalli-Sforza et al. 1988 who has constructed a differentiation tree of forty-two human populations from different continents.



Languages and genes

- Other studies have subsequently refuted the existence of indisputable correlations between the genetic classification of populations and the classification of languages.
- Cavalli-Sforza himself acknowledged that the correlation between genetic distance and linguistic distance is "extremely weak, if not zero" (Cavalli-Sforza et al. 1992, p. 5623).
- Today we also know of cases in the Caucasus where correlations are not good.



Languages and genes

- The first concerns Armenians and Azeris. They speak different languages (Armenian is an Indo-European language and Azeri is a Turkic, and therefore Altaic, language) but are nevertheless very close genetically.
- The second case is the opposite: Chechens and Ingush speak very similar languages (belonging to the North Caucasus branch of languages) but are very different from a genetic point of view (Nasidze et al. 2001).
- But what specifically is the situation in East Asia and South-East Asia?



Languages and genes

- For a while, we thought that the numerous population genetics studies in East Asia would favour one scenario over another, or even support the PEA macro-macro-family hypothesis. However, this is not the case.
- Today, we have to recognise that as these studies have accumulated and diversified (moving from the analysis of classic markers to that of mitochondrial DNA and the Y chromosome), the landscape has become increasingly blurred and it is not possible today to choose a macro-family hypothesis with any certainty (Peyraube, 2007).



Languages and genes

- The first major studies carried out by Chinese geneticists clearly concluded that the Northern Han and Southern Han have marked genetic differences. See Chu et al (1998), Su et al (1999).
- Others, more recent and more sophisticated, conclude that the differences in languages observed between northern and southern Chinese are simply due to cultural phenomena without any genetic implications. See Ding et al (2000), Yao et al (2002).



Conclusion

- As Renfrew (2000) points out, the debate on the existence or not of macro-families, not to mention the even more hypothetical existence of a single 'mother language', is not about to be resolved.
- Until we have shown that these macro-phyla share several hundred primordial lexemes, there will be no evidence.
- However, it is increasingly likely that the historical depth for reconstructing language families or proto-languages is no longer limited to 6000 BCE.



Conclusion

- The most recent and undoubtedly most promising research, whatever the contradictory hypotheses and lack of certainty to which it has so far led, is unique in that it has become truly interdisciplinary.
- Linguists have gone beyond their strict field of competence and have felt the need to work with geneticists, archaeologists, palaeo-anthropologists and palaeodemographers in an attempt to achieve the 'new synthesis' that Renfrew has been calling for since the early 1990s.



References

ALE Atlas Linguarum Europaea

Atlas des langues de Chine ((1987, 2008, 2012)

Barbujani G. et R. Sokal (1990), "Zones of sharp genetic change in Europe are also linguistic boundaries", *Proceedings of the National Academy of Science* 87 : 1816-1819.

Benedict (1942), "Thai, Kadai, and Indonesian: A new realignment in Southeast Asia", *American Anthropologist* 44: 576-601.

Benedict (1972)

Benedict (1975)

Bengston J.D. (1991), "Notes on Sino-caucasian", V. Shevoroshkin éd., *Dene-Sino-Caucasian languages*. Bochum : Studienverlag Brockmeyer.

Blust R. (1996), "Beyond the Austronesian homeland : the Austric hypothesis and its implications for archaeology", W. Goodenough éd., *Prehistoric settlement of the Pacific*. Philadelphia : American Philosophical Society. 117-140.



References

Blust R. (1997), "Subgrouping of the AN languages: consensus and controversies". Communication au 8ème colloque international de linguistique austronésienne. Taipei.

Cavalli-Sforza L.L., A. Piazza, P. Menozzi et J. Mountain (1988), "Reconstruction of human evolution : bringing together genetic, archaeological and linguistic data", *Proceedings of the National Academy of Sciences* 85 : 6002-6006.

Cavalli-Sforza L.L., E. Minch et J. Mountain (1992), "Coevolution of genes and languages revisited", *Proceedings of the National Academy of Sciences* 89 : 5620- 5624.

Cavalli-Sforza L.L. (2000), *Genes, Peoples, and Languages*. Farrar, Strauss & Giroux.



References

Chafe W.L. (1962), "Estimates regarding the present speakers of North American Indian Languages", *International Journal of American Linguistics* 28 : 162-171.

Chang K.C. (1986), *The Archaeology of Ancient China*. New Haven : Yale University Press.

Chappell & Peyraube (2006)

Chu J.Y., W. Huang, S. Kuang, J. Wang, J. Xu, Z. Chu, Z. Yang, K. Lin, P. Li, M. Wu, Z. Geng, C. Tan, R. Du et L. Jin (1998), "Genetic relationship of populations in China", *Proceedings of the National Academy of Sciences* 95 : 11763-11768.

Comrie B. (1989), *Language universals and linguistic typology*. Oxford : Basic Blanckwell. 2nd edition.

Comrie B. (2000), "Is there a single time depth cut-off point in historical linguistics?", C. Renfrew, A. McMahon et L. Trask eds., *Time depth in historical linguistics*. Cambridge: McDonald Institute for Archaeological



References

Contini M. (2000), "Vers une nouvelle linguistique historique : l'ouvrage de Mario Alinei, *Origini delle lingue d'Europa*", *Dialectologia et Geolinguistica* 2000-8 : 13-35.

Diffloth G. (2005), "On the high cost of Austric". Communication au symposium sur les perspectives d'une phylogénie des langues d'Asie Orientale. Périgueux.

Ding Y.C., S. Wooding, H. Harpending, H. Chi, H. Li, Y. Fu, J. Pang, Y. Yao, J. Yu, R. Moyzis et Y. Zhang (2000), "Population structure and history in East Asia", *Proceedings of the National Academy of Sciences* 97 : 14003-14006.

Dolgopolsky A. (1998), *The Nostratic macrofamily and linguistic paleontology*. Cambridge: McDonald Institute for Archaeological Research

Edkins (1857)



References

Enfield N. (2002), Linguistic epidemiology : Semantics and grammar of language contact in Mainland Southeast Asia. London : Routledge.

Excoffier L, B. Pelligrini, A. Sanchez-Mazas, C. Simon et A. Langaney (1987), "Genetic and history of Sub-Saharan Africa", Yearbook of Physical Anthropology 30 : 151-194.

Ferlus (1990)

Greenberg J.H. (1963a), Languages of Africa. Bloomington : Indiana Research Center in Anthropology.

Greenberg J.H. (1963b), "Some universals of grammar with particular reference to the order of meaningful elements", J.H. Greenberg éd., Universals of Language. Cambridge : MIT Press.

Greenberg et. al. (1986)



References

Greenberg J.H. (1987), *Languages in Americas*. Stanford : Stanford University Press.

Greenberg J.H. (1992), "Preliminaries to a systematic comparison between biological and linguistic evolution", J.A. Hawkins et M. Gell-Mann éd., *The Evolution of human languages*. Redwood City (CA) : Addison-Wesley Publishing Company. 139-158.

Greenberg J.H. (2001), "The methods and purposes of linguistic genetic classification", *Language and Linguistics (Taiwan)* 2-2 : 111-135.

Greenberg J.H. *Indo-European and its closest relatives : the Eurasiatic language family*. Stanford : Stanford University Press.

Greenberg J.H., C. Turner, et S. Zegura (1986), "The settlement of the Americas : A comparison of the linguistic, dental, and genetic evidence", *Current Anthropology* 27 : 477-497.



References

Hagège C. (1982), *La structure des langues*. Paris : PUF.

Haudricourt (1972)

Li Rong (1985)

Liu Xuning (1995)

Matisoff J. (1991), "Areal and universals dimensions of grammaticalization in Lahu", E. Traugott et B. Heine édés., *Approaches to grammaticalization*. Amsterdam : John Benjamins. 383-453.

Nasidze I., G. Risch, M. Robichaux, S. Sherry, M. Batzer et M. Stoneking (2001), "Alu insertion polymorphisms and the genetic structure of human populations from the Caucasus", *European Journal of Human Genetics* 9 : 267-272.



References

- Ostapirat W. (2001), "Kra-dai and Austronesian : notes on some phonological correspondences". Communication au symposium sur les perspectives d'une phylogénie des langues d'Asie Orientale. Périgueux.
- Pedersen H. (1903), "Türkische Lautgesetze", Zeitschrift der Deutschen Mo
- Peyraube A. (1986), "Recent issues in Chinese historical syntax", J. Huang et A. Li éd., New Horizons in Chinese Linguistics. Dordrecht : Kluwer. 161-214.
- Peyraube A. (2002), "L'évolution des structures grammaticales", Langages 146 : 46-58.
- Peyraube (2007)
- Peyraube (2015)



References

- Poloni E.S., O. Semino, G. Passarino, A. Santachiara-Benerecetti, I. Dupanloup, A. Langaney et L. Excoffier (1997), "Human genetic affinities for Y-chromosome P49a,f/TaqI haplotypes show strong correspondance with linguistics", *American Journal of Human Genetics* 61 : 1015-1035.
- Przyluski (1931)
- Ramsey (1987)
- Reid L. (1994), "Morphological evidence for Austric", *Oceanic linguistics* 33-2 : 323-344.
- Renfrew C. (2000), "At the edge of knowability : towards a prehistory of languages", *Cambridge Archaeological Journal* 10-1 : 7-34.
- Ringe D. (1999), "Language classification: Scientific and unscientific methos", B. Sykes éd., *The Human Inheritance : Genes, Language and Evolution*. Oxford : Oxford University Press. 45-74.



References

Ruhlen M. (1992), "An overview of genetic classification", J.A. Hawkins et M. Gell-Mann édés., *The Evolution of human Languages*. Redwood City (CA): Addison-Wesley Publishing Company. 159-189.

Ruhlen M. (1994), *On the origins of languages : Studies in linguistic taxonomy*. Stanford: Stanford University Press.

Ruhlen M. (1997), *L'origine des langues*. Paris : Belin.

Sagart L. (1993), "Chinese and Austronesian : evidence for a genetic relationship", *Journal of Chinese Linguistics* 21-1 : 1-62.

Sagart L. (2001a), "The Evidence for Sino-Austronesian". Communication au symposium sur les perspectives d'une phylogénie des langues d'Asie Orientale. Périgueux.



References

Schlegel G. (1901), "Review of Frankfurter's Siamese Grammar", *T'oung Pao* 2 : 76-87.

Schmidt W. (1905), *Grundzüge einer Lautlehre der Mon-khmer Sprachen*.

Sokal R.R., N. Oden et B. Thomson (1988), "Genetic changes across language boundaries in Europe", *American Journal of Physical Anthropology* 76: 337-361.

Sokal R.R., N. Oden et B. Thomson (1988), "Genetic changes across language boundaries in Europe", *American Journal of Physical Anthropology* 76: 337-361.

Starosta S. (2001), "PEA: A scenario for the origin and the dispersal of the languages of East and Southeast Asia and the Pacific". Communication au symposium sur les perspectives d'une phylogénie des langues d'Asie Orientale. Périgueux.



References

Starostin S. (1989), "Nostratic and Sino-Caucasian", V. Shevoroshkin ed., Explorations in language macrofamilies. Bochum: Studienverlag Dr. Brockmeyer. 42-66.

Su B., J. Xiao, P. Underhill, R. Deka, W. Zhang, J. Akey, W. Huang, D. Shen, D. Lu, J. Luo, J. Chu, J. Tan, P. Shen, R. Davis, L. Cavalli-Sforza, R. Chakraborty, M. Xiong, R. Du, P. Oefner, Z. Chen et L. Jin (1999), "Y-chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age", American Journal of Human Genetics 65 : 1718-1724.

Szulmajster-Celnikier A. (1998), "Eloge de la prudence méthodologique", La Recherche 306 : 76-81.

Thurgood G. (1994), "Tai-kadai and Austronesian : the nature of the historical relationship", Oceanic linguistics 33-2 : 354-368.



References

Thurgood (2003)

Trombetti A. (1905), *L'Unità d'origine del Linguaggio*. Bologne.

Van Driem G. (1997), "Sino-Bodic", *Bulletin of the School of Oriental and African Studies* 60-3 : 455-488.

Van Driem G. (2001), *Languages of the Himalayas*. Leiden : Brill.

Wade (1987)

Wang (1996)