On the internal classification of Indo-European languages: survey[*]
Václav Blažek

The main purpose of the present study is to confront most representative models of the internal classification of Indo-European languages and their daughter branches.

0. Indo-European
0.1. In the 19th century the tree-diagram of A. Schleicher (1860) was very popular:

After the discovery of the Indo-European affiliation of the Tocharian A & B languages and the languages of ancient Asia Minor, it is necessary to take them in account. The models of the recent time accept the Anatolian vs. non-Anatolian ('Indo-European' in the narrower sense) dichotomy, which was first formulated by E. Sturtevant (1942). Naturally, it is difficult to include the relic languages into the model of any classification, if they are known only from several inscriptions, glosses or even only from proper names. That is why there are so big differences in classification between these scantily recorded languages. For this reason some scholars omit them at all.

0.2. Gamkrelidze & Ivanov (1984, 415) developed the traditional ideas:

[*] Previously unpublished. Reproduced with permission. [Editor’s note]
0.3. Vladimir Georgiev (1981, 363) included in his Indo-European classification some of the relic languages, plus the languages with a doubtful IE affiliation at all:

```
Indo-European
  Northern
    Balto-Slavic
    Germanic
      Celtic
        Ligurian
        Italic & Venetic
  Western
    Germanic
      Celtic
        Ligurian
        Italic & Venetic
    Messapic
    Siculian
  Central
    Greek & Macedonian
  Eastern
    Daco-Mysian & Albanian
    Indo-Iranian
  Southern = Aegean
    Thracian
    Pelasgian
  Southeast = Anatolian
    Palaic
    Hittite; Lydian; Etruscan-Rhaetic; Elymian
    Luwian; Lycian; Carian; Eteocretan
```

0.4. Eric Hamp proposed his original model of the Indo-European disintegration, including the relic idioms, based on specific isoglosses in phonology, morphology and lexicon (1990):

```
Indo-European
  Indo-Hittite
    Anatolian
      Indo-Aryan
      Nuristanic
    Asiatic Indo-European
      Iranian
      Armenian
    Pontic South Indo-European
      Greek
      Macedonian
      Slavic
    Residual
      Baltic
      Thracian
      Dacian
      Albanian
      Prehellenic
        Pelasgic
        Germanic
      Tocharian
      Illyrian
      Messapic
      Phrygian
      Venetic
      Italic
      Celtic
```
0.5. One of the last applications of cladistics was presented by D. Ringe, T. Warnow & A. Taylor (2002, 87):

Anatolian
Tocharian
Celtic
Italic
Germanic
Albanian
Armenian
Greek
Indo-Aryan
Iranian
Slavic
Baltic

0.5. The absolute chronology is available only thanks to glottochronology. The most recent result of Sergei Starostin (Workshop on the chronology in linguistics, Santa Fe 2004) applies his own model of the ‘recalibrated’ glottochronology, where all borrowings are excluded before any calculation and the coefficient of changes is empirically recounted to 5% per millennium instead of 14% postulated by Swadesh.
1. Indo-Aryan

The only attempt to apply glottochronology for several modern Indo-Aryan languages in confrontation with Sanskrit was realized by S. Starostin and his team (database 2004):

![Indo-Aryan Tree Diagram](image)

2. Dardic

The only tree-diagram was constructed by S. Starostin's team (database 2004):

![Dardic Tree Diagram](image)

Note: Concerning the internal classification and depth of divergence of the Nuristani languages, there are no available results.
3. Iranian
The only attempt to construct the tree-diagram for the Iranian languages was realized by S. Starostin and his team (Santa Fe 2004).

4. Anatolian
With exception of Hittite no Anatolian language allows to apply glottochronology for our limited knowledge of their lexical corpora. That is why the existing classifications are based on combinations of phonological, morphological and lexical isoglosses. In the recent time these three alternative models of the internal classification of the Anatolian languages were proposed.

4.1. N. Oettinger 1978, 92 (supplemented on the basis of personal communication in 2001):

4.2. R. Werner (1991, 17)

5. Greek
The most detailed scheme classifying the Greek dialects was proposed by A. Bartoněk on the basis of phonology and morphology (1987, 104; 2003, 494):

<table>
<thead>
<tr>
<th>2000</th>
<th>1800</th>
<th>1600</th>
<th>1400</th>
<th>1200</th>
<th>1000</th>
<th>800</th>
<th>600 B.C.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: AM Asia Minor, C. Central, E. East, I. Insular, W. West.

Note: Greek can be classified as one of the Hellenic languages, together with Phrygian / Brygian, ancient Macedonian, and perhaps also Messapic, if the hypothesis of M. Huld (1995, 147-55) is accepted. Unfortunately, the lexical corpora do not allow any quantification.

6. Paleo-Balkanian
Extremely poor data and their ambiguous interpretations lead to various hypotheses. The present author finds as probable following: In Prehellenic = Pelasgian the Lautverschiebung operated; the language was of the centum-type rather than of satem-type. If Thracian & Bithynian were satem-languages with Lautverschiebung, their closer relation with Armenian is expectable. Albanian is a descendant of Illyrian, both the satem-languages. The change *ĝ* > b in Dacian indicates more probably the centum-type.
7. Italic

For ancient Italic languages the application of glottochronology is not possible, because their lexicons are very limited, naturally with exception of Latin. For this reason the following hypothetical classification is based only on unique common features in phonology, morphology and sporadically in lexicon. The result does not reflect any grade of a mutual relationship.

Italic

North

*kw > kv/qu

Venetic

Faliscan

Latin

North

Latino-Faliscan

South Picenian

Umbrian

Sabine

Vestinian

Aequian

Marsian

Pre-Samnite

Osco-Umbrian

Umbro-Sabine

Marrucinian

Paelignian

Volscan

Oscan

South

*kw > p

Osco-Sabellian

Ausonio-Sicilian

Ausonian & Auruncan

Siculian

Elymian

7A. Romance

On the other hand, the lexical material of the Romance languages served for determination of the basic constants in glottochronology. Let us confront several models of their disintegration:


Proto-Romance

East Romance

Balkano-Romance

Dalmatian

Italian

Sardinian

Italo-Dalmatian

Rhaeto-Romance

French

Occitlanian

Gallo-Romance

Catalanian

Spanish

Galician

Portuguese

West Romance

Ibero-Romance

Rumanian.

7A.3. Merritt Ruhlen 1987, 326:
It is natural that glottochronology was also applied for Romance languages. Let us confront two attempts from the recent time:

Embleton 1986, 142:

![Tree diagram for Romance languages](image1)

Starostin (Santa Fe 2004):

![Tree diagram for Romance languages](image2)

8. In the area between Italic and Celtic there were at least two relic languages which could form a closer unit in the genealogical sense: **Ligurian** and **Lusitanian**, former reconstructed on the basis of proper names attested by classical authors in northern Italy, latter known from several inscriptions written in the Latin alphabet, discovered in south Portugal and Spain.

9. **Celtic**

There are two alternative models of disintegration of the Celtic languages.

9.1. The first model has to reflect the opposition between the insular and continental languages. It is defended e.g. by W. Cowgill (1975) or P. Schrijver (1995, 463).

9.2. The alternative and more traditional model is based on the *q/p*-isogloss in the reflexes of the Indo-European labiovelar *k̞*. The figures for living languages (plus Cornish) and the age of the divergence of Goidelic vs. Brythonic were calculated by S. Starostin and his team (Santa Fe 2004). The positions of other nodes indicated by question marks represent only rough assessments:
10. Germanic

The best summarization of various ideas concerning the classification of the Germanic languages is the study of W. Mańczak (1992; cf. also Blažek & Pirochta 2004).

10.1. J.Ch. Adelung (1806) divided the Germanic languages into two branches:

- Scandinavian
  - non-Suevic: Frisian, Frankish, Saxon, Anglosaxon
- Suevic
  - Langobardic, Alamanic, Suevic, Gothic, Burgundian, Vandalic

10.2. Similarly J. Grimm (1819) operated with the binary classification. For some of the tribal dialects he supposed the transit character, viz. Frisian & Anglian (1-2), Frankish (2-3), Quadic & Marcomanic (3-4). Alternatively he assumed the opposition of East Germanic vs. others.

10.3. Applying his original method based on the lexicostatistic analysis of parallel texts, Mańczak (1992) formulated a similar conclusion. He ordered the languages decliningly according their relationship with Gothic: the closest has to be Old High German, further Old Saxon, finally Scandinavian languages.

10.4. Another model of the binary classification was presented by K. Müllenhoff (1898):
10.5. The most frequent model divides the Germanic languages into three branches: East, North and West. The author of the following classification is J. Schmidt (1860):

- **North**
  - Low German in a wider sense
    - Saxon
    - Old Saxon
  - Frisian
  - Anglosaxon
  - Dutch
  - High German
    - Plattdeutsch
    - Hochdeutsch

- **West**
  - Saxon
  - Old Saxon
  - Low German
    - Plattdeutsch
  - High German

- **East**
  - Gothic

10.6. F. Maurer (1943) tried to depict the development from the tribal Germanic dialects to the languages of the late middle age and present time, including the convergent processes:

- **Germanic**
  - Hessenic
  - Weser-Rhine
  - Frankish
  - Saxon
  - Anglosaxon
  - Angelian
  - Frisian
  - Nordic
  - Scandinavian

10.7. E. Schwarz (1951) assumed that c. 200 B.C. the Germanic language continuum was already divided into the North zone, generating the later Scandinavian languages and Gothic, and the South zone, where the later German dialects were formed. About 4 cent. later the third, transit zone, cristalized, developing in the languages of Angels and Frisians.
10.8. The most detailed scheme of the development of the Germanic languages was proposed by T.V. Toporova (2000), inspired by Maurer and Schwartz:


10.9. E. Antonsen (1975) assumed the opposition of the east and northwest branches:
10.10. H. F. Nielsen (2000) returned to the traditional idea, identifying in Old Runic a direct ancestor only of the Scandinavian languages:

```
Northwest Germanic

Old Runic

West Germanic

North Sea

Old High German

Old Saxon

Old Frisian

Old English

Old Norse
```

10.11. Sheila Embleton (1986, 117) used for her classification of the Germanic languages her modification of glottochronology:

```
+100 n. l. +300 +600 +900 +1200 +1500 +1800

Swedish

Danish

Norwegian

Faeroese

Islandic

English

Frisian

Vlamish

Afrikaans

Dutch

Yiddish

Low German

High German

Gothic
```

10.12. The most recent attempt to classify the Germanic languages was published by Starostin & Burlak (2001, 82-105). They applied Starostin’s recalibrated glottochronology for 7 literary living languages and Gothic:

```
-100 +100 +300 +500 +700 +900 +1100 +1300 +1500

Swedish

Danish

Nynorsk

Icelandic

English

Dutch

High German

Gothic
```

```
+30 +60 +90

73%

80%

70%

-80

93%

94%

89-91%

94%

98%

80%

94%

98%

94%

98%
```

```
+1000 +1200 +1500

89-91%

+1000

+1200

+1500
```

```
+70

73%

+1000

+1200

+1500
```

```
98%

Swedish

Danish

Nynorsk

Icelandic

English

Dutch

High German

Gothic
```
11. Baltic
According to tradition, the Baltic languages are divided into the west part represented by Old Prussian, from c. 1700 extinct, and eastern part, represented by the living languages, Lithuanian and Latvian. But the Baltic dialectology was much more complex a millennium ago. The following model was proposed by V. Mažiulis (1981):

```
North periphery
  Zemgalian
  Selian
  Couronian
  Latvian

Central
-140

Baltic

Lithuanian

South periphery
  Yatvingian
  Prussian
  Galindian
```

The first serious application of the classical glottochronology was used by Lanszweert (1984, xxxii-xxxvii), who has found 58.6% for Prussian vs. Lithuanian and 55.2% for Prussian vs. Latvian. The result of Girde尼斯 & Mažiulis (1994, 9) are lower: 68% Lithuanian vs. Latvian, 49% Lithuanian vs. Prussian, 44% Latvian vs. Prussian. Starostin (Santa Fe 2004 and p.c., June 2005) dated the separation of Lithuanian and Latvian to 80 B.C., Lithuanian and the ‘Dialect of Narew’ to 30 B.C., Latvian and the ‘Dialect of Narew’ to 230 B.C. The position of Prussian in his calculations is rather strange, it has to be closer to Slavic than to Baltic. Novotná & Blažek (forthcoming), calculating the synonyms too, have reached the following results:

```
-1400    -1000   -600   -200   +200   +600

Latvian   84.8%

-600 76.3% +600 Lithuanian

56% / 58% +190 ‘Dialect of Narew’

-830 / -730

46.7% -1400 Prussian

Common Slavic
```

The double-result 58/56% for Prussian vs. other Baltic languages reflects the calculation without / with the ‘Dialect of Narew’. The score 43% between Prussian and the ‘Dialect of Narew’ (Poganiške gwary z Narewu; see Zinkevičius 1984) in confrontation with 62% and 55.2% for Prussian vs. Lithuanian and Prussian vs. Latvian respectively, excludes the identification of the ‘Dialect of Narew’ with the historical Yatwingians, known from the Middle Ages, if their language had to be connected with the other Baltic idioms of the southern periphery, including Prussian. Regarding this big difference, it seems better to accept the explanation of Schmid (1986) who identified in the ‘Dialect of Narew’ a strong influence of Northeast Yiddish, spoken in the big cities of Lithuania and Latvia, hence the hybrid East Baltic-German idiom.
12. Slavic

12.1. According to the traditional model the Slavic languages are divided into three parts (cf. e.g. J. & B. Grimes 1996, 58):

- East Slavic
  - Russian
  - Belorussian
  - Ukrainian & Rusyn
- Lechitic
  - Polish
  - Kašubian
  - Pomerian Slovincian
  - Polabian
- Slouth Slavic
  - Lower Sorbian
  - Upper Sorbian
- Czech-Slovak
  - Czech
  - Slovak
  - Slovenian
  - Serbo-Croatian

12.2. The classification of the Slavic languages by Starostin (Santa Fe 2004), using his recalibrated glottochronology, is revolutionary in both topology and chronology:
12.3. Using the principles of Starostin’s recalibration of glottochronology, Novotná & Blažek (2005) proposed another model of the internal grouping of the Slavic languages:

<table>
<thead>
<tr>
<th>Year</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>81</td>
<td>Russian</td>
</tr>
<tr>
<td>83</td>
<td>Ukrainian</td>
</tr>
<tr>
<td>85</td>
<td>Belorussian</td>
</tr>
<tr>
<td>87</td>
<td>Polish &amp; Kashubian</td>
</tr>
<tr>
<td>89</td>
<td>Polabian</td>
</tr>
<tr>
<td>91</td>
<td>Lower Lusatian</td>
</tr>
<tr>
<td>95</td>
<td>Upper Lusatian</td>
</tr>
<tr>
<td>97</td>
<td>Slovak</td>
</tr>
<tr>
<td>1070</td>
<td>Czech</td>
</tr>
<tr>
<td>1220</td>
<td>Slovenian</td>
</tr>
<tr>
<td>1300</td>
<td>Serbo-Croatian</td>
</tr>
<tr>
<td>1300</td>
<td>Makedonian</td>
</tr>
<tr>
<td>1630</td>
<td>Bulgarian</td>
</tr>
<tr>
<td>520-600</td>
<td>Lower Lusatian</td>
</tr>
<tr>
<td>720</td>
<td>Upper Lusatian</td>
</tr>
</tbody>
</table>

REFERENCES:
Müllenhoff, K. 1900: Deutsche Altertumskunde, IV. Berlin.
Schleicher, August. 1860: Die Deutsche Sprache. Stuttgart.