Languages of Europe: a typological perspective

Standard Average European

Lecturer: Luigi Talamo

Department of Language Science and Technology
Credits

The present material was originally made by Jun. Prof. Annemarie Verkerk for her course on ‘Languages of Europe – a typological perspective’. Used here with permissions and some modifications/adaptations.
Sprachbund

Language contact:
- Lexicon, grammatical characteristics and other linguistic features may spread between languages when speakers of different languages are in contact for longer periods. This occurs most often when speakers are multilingual in one or more geographically close languages.

Sprachbund or linguistic area:
- A group of languages that shares no systematic sound correspondences, i.e. are not related or only remotely, but have a significant amount of shared features in grammatical structure and word-formation.

KvdA2011 p. 307
Krasnoukhova 2019: The languages in the Balkan Sprachbund can be divided into core languages (i.e., sharing many of the Balkan characteristics) and periphery languages (i.e., sharing the Balkan characteristics to a lesser degree). According to Olga Mišeska Tomić, among the core languages are Macedonian, Bulgarian, Romanian, Aromanian, Megleno-Romanian, Albanian, and Modern Greek. Among the peripheral languages are Balkan Romani dialects, Serbian, and Judeo-Spanish, with the Balkan Turkish as a probable instigator for the Sprachbund development.
Krasnoukhova 2019: However, the core languages are not always in the core for each specific characteristic, illustrating that clear-cut boundaries of the area are difficult to draw.

Three main structural areas: case markers/determiners, pronominal clitics and subjunctive constructions.
List of linguistic areas (not exhaustive)

- Balkan
- Ethiopian highlands (Ferguson 1970)
- Meso-America (Campbell et al. 1986)
- Sepik River Basin in New Guinea (Thomason 2000)
- Pacific Northwest (Thomason 2000)
- Mainland South-East Asia (Enfield 2005)
- India/South Asia (Emeneau 1956)
- Caucasus (Chirikba 2008)
- East Nusantara (Klamer et al. 2008, Austronesian and Papuan languages spoken in eastern Indonesia and East Timor)
- Eurasia (Dryer 1989, as extending from Turkey to Japan and from South India to northern Russia to western Siberia)
Standard Average European (SAE)

Origin of the phrase:
„Since, with respect to the traits compared, there is little difference between English, French, German, or other European languages with the POSSIBLE (but doubtful) exception of Balto-Slavic and non-Indo-European, I have lumped these languages into one group called SAE, or “Standard Average European”. (Whorf 1956: 138, KvdA2011 p. 291)

The idea that there is such a thing as a European linguistic area has been the topic of several papers:
Standard Average European (SAE)

Bechert et al. 1990, *Toward a typology of European languages.*
Haspelmath 1998, *How young is Standard Average European?*
van der Auwera 1998, *Phasal adverbials in the languages of Europe*
König & Haspelmath 1999, *Der europäische Sprachbund*
Haspelmath 2001, *The European linguistic area: Standard Average European*
Heine & Kuteva 2006, *The Changing Languages of Europe*
Helmbrecht 2006, *Typologie und Diffusion von Höflichkeitspronomina in Europa*
van der Auwera 2011, *Standard Average European*
Cysouw 2011, *Quantitative explorations of the worldwide distribution of rare characteristics, or: the exceptionality of northwestern European languages*
Drinka 2019, *Areal factors in the development of the European periphrastic perfect*
Eurocentrism

- Discovery of the European linguistic area is relatively late; only after the grammatical properties of languages around the world were described, did comparative linguists realize how peculiar, in some ways, the European languages are.

- Dahl (1990: 3): „One of the greatest problems that the universal study of human language has had to cope with has indeed been the **European bias**: most linguists have been speakers of European languages, and the other languages that they have known or had access to information about have more often than not been European. As Bell (1978) notes, even linguists who have an ambition to widen their perspective mostly end up with a European or even Indo-European bias in their data bases. This would of course not be so problematic if it were not the case that European languages are much more like each other than languages are in general.“

A nice paper against Eurocentrism is “Gil, David - Escaping Eurocentrism: Fieldwork as a Process of Unlearning”, which you can find on the repository.
Standard Average European (SAE)

Definition of Sprachbund or linguistic area:

A group of languages that shares no systematic sound correspondences, i.e. are not related or only remotely, but have a significant amount of shared features in grammatical structure and word-formation.

- “In the case of SAE, three entire branches of Indo-European (Romance, Germanic and Balto-Slavic) belong to the linguistic area. However, here too it is clear that we are not dealing with a genealogical grouping, because nobody ever proposed a branch of Indo-European that consists of precisely these three families.” Haspelmath (2001: 1492)

- In addition, the SAE features are innovations rather than retentions from Proto-Indo-European.

- “Standard Average European may even appear as an “exotic language” (Dahl 1990).” Haspelmath (2001: 1492)
Europeanisms (SAE features)

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

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

Haspelmath (2001: 1493)
Haspelmath 2011: The map combines nine features of § 2.:
1. definite and indefinite articles;
2. relative clauses with relative pronouns;
3. ‘have’-perfect;
4. participial passive;
5. dative external possessors;
6. negative pronouns and lack of verbal negation;
7. relative-based equative constructions;
8. subject person affixes as strict agreement markers;
9. intensifier- reflexive differentiation.

The languages in the core (French and German) show the SAE value in all nine of these features. The languages in the next layer (Dutch, other Romance, Albanian) show eight features, the next layer (English, Greek, Romanian) shows seven features, and so on. In this map, the resulting picture is actually very clear, because the SAE area with at least five SAE features stands out from the remaining languages, which have at most two SAE features.
### Standard Average European (SAE)

- **core (8 or 9 features):** Dutch & German, French, Spanish, Portuguese, Italian, Albanian, Sardinian. Also, *Charlemagne Sprachbund*
- **just outside the core (7 features):** English, Romanian, Greek
- **further removed but within SAE (6 features):** Icelandic, Faroese, Norwegian, Swedish, Czech, Polish, Latvian, Lithuanian, Slovene, Serbo-Croatian, Ukrarian, Russian and other Slavonic languages
- **marginal:** Hungarian, Balto-Finnic (5 features or less), Maltese
- **outside SAE (max 2 features):** Celtic, Basque, Turkish, etc.

Charlemagne Sprachbund is a term proposed by Van der Auwera: it refers to the historical fact that Dutch/German and French are the offspring of the languages spoken in the kingdom of Charlemagne, who himself spoke a Rhenish Franconian dialect (incidentally, of the same dialect family of Saarländisch!).
The major Standard Average European features

All the features covered in the rest of Haspelmath (2001) are **syntactic** or **morphosyntactic**:

„Perhaps phonologists have not looked hard enough, but at least one major recent study of word prosody in European languages has not found any phonological evidence for Standard Average European (van der Hulst et al. 1999, especially Maps 1-4) (but cf. Pisani 1969). A few generalizations are discussed by Ternes (1998), but he finds that in most respects European languages are unremarkable from a world-wide perspective.“

„Perhaps the only features worth mentioning are the relatively large vowel inventories (no 3-vowel or 4-vowel inventories) and the relatively common consonant clusters (no restriction to CV syllables). In these respects, European languages are not average, but they are by no means extreme either.“

Haspelmath (2001: 1493)

No phonetic Europeanisms, just relatively large vowel inventories (i.e. more than 4 vowels) and barely unrestricted consonant cluster (i.e. not only CV syllables).
The major Standard Average European features

1. Definite and indefinite articles
2. Relative clauses with relative pronouns
3. ‘Have’-perfect
4. Nominative experiencers (has been refuted)
5. Participle passive
6. Anticausative prominence
7. Dative external possessors
8. Negative pronouns and lack of verbal negation
9. Particles in comparative constructions
10. Relative-based equative constructions
11. Subject person affixes as strict agreement markers
12. Intensifier-reflexive differentiation

A dozen of Europeanisms.
Grammatical properties of languages from a typological perspective

- Word order (sentential, adnominal)
- Morphology (fusion, exponence, flexivity)
- Nominal categories (gender, number, definiteness, demonstratives, case)
- Verbal categories (tense, aspect (perfect), mood, evidentiality)
- Clausal (alignment (nominative experiencers), person marking, passive, causative, negation, question marking, non-verbal predication, comparatives)
- Complex clauses (relative clauses, other multi-clause sentence types)
Definite and indefinite articles

Lyons (1999: 2ff) *Definiteness*: „So the question we are concerned with is: What is the difference in meaning between the car and a car, between the greedy child and a greedy child, between the hibiscus I planted last summer and a hibiscus I planted last summer? Many traditional grammars would give answers like the following: The indicates that the speaker or writer is referring to a definite or particular car etc., not just any. But apart from being rather vague, this answer is quite inaccurate. If I say I bought a car this morning, I am not referring to just any car; the car I bought is a particular one, and is distinguished in my mind from all others. Yet a car is indefinite. There is in fact no general agreement on what the correct answer is, but two major components of meaning have been much discussed“:

- **Familiarity and identifiability;** the is used when referent is familiar to both hearer and speaker;
- **Uniqueness and inclusiveness;** the signals that there is just one entity satisfying the description used.

Four semantic (external) definitions for definiteness.
Map 107.1: Definite and indefinite article
Definite and indefinite articles

https://wals.info/combinations/37A_38A#2/25.5/148.5
- To the west: Celtic languages don’t have indefinite articles;
- To the east: Uralic and Slavic languages lack both definite and indefinite articles;
- To the southeast: Turkish and Caucasian languages have no definite articles;
- Having both definite and indefinite articles is relatively uncommon across the world’s languages.
Relative clauses with relative pronouns

Characteristics of SAE relative clause (Haspelmath 2001: 1494):
1. The relative clause is **post-nominal**: NRel,
2. There is an **inflecting** relative pronoun: der vs. den;
3. This pronoun **introduces the relative clause**, 
4. The relative pronoun functions as a **resumptive**, i.e. it signals the head’s role within the relative clause.

Der Mann, [der mich begrüßt hat], war ein Deutscher.
man.NOM REL.NOM me greet.PTCP has be.3SG.PST one German
‘The man who greeted me was a German.’

Comrie 2005, WALS chapter on ’Relativization on Subjects’

What’s so special about SAE relative clauses?
relative clause with introducing relative pronoun
- - - - only particle relative clause

Map 107.2: Two relative clause types in Europe
Values
Relative pronouns: SAE-style relative pronouns
Non-reduction: ‘the head noun appears as a full-fledged noun phrase within the relative clause’ (https://wals.info/chapter/122)
Pronoun retention: ‘the position relativized is explicitly indicated by means of a resumptive personal pronoun’: something like ‘*The man who he greeted me is German’ but with ‘he’ not mandatory in a simple sentence.
Gap: the relative clause is juxtaposed without external marking.
Tense and aspect: ‘Have’-perfect

- **Tense** and **aspect** are **grammatical categories** of verbs;
- “tense is grammaticalisation of location in time” & “aspect is grammaticalisation of expression of internal temporal constituency” (of events, processes etc.)
- **Tense**: opposition between past-present-future tense.
- **Aspect**: imperfective-perfective contrast: opposition between a form which is used (almost) exclusively for single completed events in the past (perfective) and a form which is used for everything else (imperfective).

*Imperfective vs. perfective vs. perfect in English*
I am doing something: imperfective
I did something: perfective (event completed)
I have done something: perfect (the event is completed but has consequences on the present)
Map 107.3: ‘Have’-perfects in Europe
Tense and aspect: ‘Have’-perfect

Aspectual category „to express events that took place before the temporal reference point but which have an effect on or are in some way still relevant at that point“:

<table>
<thead>
<tr>
<th>Language</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>I have read the book</td>
</tr>
<tr>
<td>Dutch</td>
<td>Ik heb het boek gelezen</td>
</tr>
<tr>
<td>German</td>
<td>Ich habe die Zeitschrift gelesen</td>
</tr>
<tr>
<td>Spanish</td>
<td>He leído la revista</td>
</tr>
<tr>
<td>Albanian</td>
<td>Kam lexuar revistën</td>
</tr>
<tr>
<td>Danish</td>
<td>Jeg har læst magasinet</td>
</tr>
</tbody>
</table>

Dahl and Velupillai 2005, WALS chapter „The Perfect“

External definition of perfect (different from perfective!).
Diachronic sources for the perfect (or lack thereof)
From possessive: SAE-style have perfect, which derives from a verb meaning possession
From finish, already: grammaticalization of other verbs or adverbs
Other perfect: other sources
No perfect: lack of perfect
How does the semantic role of experiencer is coded? From a semantic perspective, it is something that is between agent and patient...

If the experiencer is coded like agent i.e., with the nominative: generalizing type.
If the stimulus is coded like agent i.e., with the nominative: inverting type.

Of course, this may vary even within a single language, so Bossong has taken 10 verbs across languages.
Low scores indicate that the language has few verbs in which the experiencer is coded as the passive. This ranges from English, which has zero experiential predicates with inversion marking (out of the 10 verbs investigated by Bossong) and Lezgian, which has only experiential predicates with inversion marking.
Nominative experiencers

This claim has been refuted very recently by Maria Zielenbach:
- Typological study of the WALS 100-language sample;
- 22 relevant languages
- Out of these 22, 19 pattern similarly to SAE, i.e. are of the mostly generalizing type;
- To have nominative experiencers is thus the default among the world’s languages, not a feature special for SAE.
Participial passive

Form:
**Auxiliary verb + past participle:**

- active voice
- passive voice

**English:**
- John loves Mary
- Mary is **loved** by John

**German:**
- Jan liebt Marie
- Marie wird von Jan **geliebt**

Function:
Haspelmath (1990: 27):

„A construction is called passive if:

(i) the active subject corresponds either to a non-obligatory oblique phrase or to nothing; and
(ii) the active direct object (if any) corresponds to the subject of the passive; and
(iii) the construction is somehow restricted vis-à-vis another unrestricted construction (the active), e.g. less frequent, functionally specialized, not fully productive.”

Check Haspelmath’s passive definition against Croft’s definition of basic (slide no. 52 of the introductory course).
Map 107.5: Participial passives in Europe
Passive Voice

Hasepmlath (1990), The Grammaticization of Passive Morphology’

Typological study of passives in 80 languages:
31/80 have a passive; out of these 31:
• 25 have a passive stem affix;
• 6 have a auxiliary verb (participle) passive construction;
• 1 has a passive particle;
• 3 have an extrainflectional passive affix;
• 2 have differential subject person markers for the passive;
• 1 has an alternate passive stem affix.

Across world’s language, only six languages have a SAE passive type.
Passive Voice: passive particle

Tahitian (Austronesian, Malayo-Polynesian: Potsdam & Polinksy 2012)

‘Ua hōhoni te ma’o ‘i te tāvana.
PFV bite DET shark ACC DET chief
‘The shark bit the chief.’

‘Ua hōhoni hia te tāvana nā/e te ma’o.
PFV bite PASS DET chief by/by DET shark
‘The chief was bitten by the shark.’
Passive Voice: participial passive

Haspelmath (1990: 29): „Auxiliary verb. This expression type is found in the present sample only in the Indo-European languages Latin (esse 'be'), Baluchi (bu- 'be'), Danish (blive 'become,' være 'be') and Maithili (jā- 'go,' par- 'fall'). In each case the auxiliary is combined with a passive participle of some sort. Although this expression type is certainly not confined to Indo-European, the distribution in the Gramcats sample confirms Dryer's (1982:55) observation that "the use of a copula plus an adjectival in passive clauses is rare outside Indo-European. In most languages, the passive is formed by adding a passive affix to the verb."“

Adjectival = past participle
Anticausative prominence: Inchoative-Causative alternations

Haspelmath (1993: 90): „An inchoative/causative verb pair is defined semantically: it is a pair of verbs which express the same basic situation (generally a change of state, more rarely a going-on) and differ only in that the causative verb meaning includes an agent participant who causes the situation, whereas the inchoative verb meaning excludes a causing agent and presents the situation as occurring spontaneously.“

The stick broke (inchoative) vs. The girl broke the stick (causative)
The snowman melted (inchoative) vs. The sun melted the snowman (causative)
The pants tore (inchoative) vs. Rosalind tore her pants (causative)

Haspelmath (1993: 90)

External definition for inchoative/causative action pair.
Map 107.6: Percentage of anticausative pairs
Bold = basic form

In the SAE type (Anticasusative alternation) the lexical form of the verb already encodes the causative meaning: in order to have an inchoative meaning, we add syntactical/morphological marking, in this case the Italian si clitic.

Inchoative-Causative alternations

Haseplamth (1993: 91): recognises three types:

1. causative alternation: the inchoative verb is basic and the causative verb is derived;
   - French: Le soleil fait fondre la neige; The sun melts the snow’ vs. La neige *fondre; The snow melts’

2. anticasusative alternation: the causative verb is basic and the inchoative verb is derived;
   - Italian: Il sole *scioglie; The sun melts the snow’ vs. La neve si scioglie; The snow melts’

3. non-directed alternation: neither the inchoative nor the causative verb is derived from the other
   1. equipollent alternations: both are derived from the same stem which expresses the basic situation, by means of different affixes, different auxiliary verbs, or different stem modifications, as in Hindi-Urdu *Surru honaa; begin (intr’); Surru karmaa; begin (tr’)
   2. suppletive alternations: different verb roots are used, as in Russian *goret; burn (intr’); žeč; burn (tr’)
   3. labile alternations: the same verb is used both in the inchoative and in the causative sense, as in English break, melt, tear
<table>
<thead>
<tr>
<th>Language</th>
<th>A</th>
<th>C</th>
<th>E</th>
<th>L</th>
<th>S</th>
<th>A/C</th>
<th>% non-dir.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Russian</td>
<td>31</td>
<td>23</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>46.00</td>
<td>26</td>
</tr>
<tr>
<td>German</td>
<td>31</td>
<td>14.5</td>
<td>0</td>
<td>4</td>
<td>11.5</td>
<td>1</td>
<td>29.00</td>
</tr>
<tr>
<td>Greek</td>
<td>31</td>
<td>13.5</td>
<td>0</td>
<td>0</td>
<td>16.5</td>
<td>1</td>
<td>27.00</td>
</tr>
<tr>
<td>Romanian</td>
<td>30</td>
<td>24</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>24.00</td>
<td>17</td>
</tr>
<tr>
<td>French</td>
<td>31</td>
<td>20.5</td>
<td>2</td>
<td>0</td>
<td>7.5</td>
<td>1</td>
<td>18.25</td>
</tr>
<tr>
<td>Lithuanian</td>
<td>31</td>
<td>17.5</td>
<td>6</td>
<td>6</td>
<td>0.5</td>
<td>1</td>
<td>2.92</td>
</tr>
<tr>
<td>Hebrew</td>
<td>31</td>
<td>20.5</td>
<td>7.5</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2.75</td>
</tr>
<tr>
<td>Arabic</td>
<td>31</td>
<td>17</td>
<td>8.5</td>
<td>3</td>
<td>1</td>
<td>1.5</td>
<td>2.60</td>
</tr>
<tr>
<td>Georgian</td>
<td>31</td>
<td>9</td>
<td>4.5</td>
<td>15.5</td>
<td>0</td>
<td>2</td>
<td>2.60</td>
</tr>
<tr>
<td>Armenian</td>
<td>31</td>
<td>16</td>
<td>8.5</td>
<td>5.5</td>
<td>0</td>
<td>1</td>
<td>1.88</td>
</tr>
<tr>
<td>Swahili</td>
<td>31</td>
<td>11</td>
<td>11</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>Finnish</td>
<td>31</td>
<td>12</td>
<td>13.5</td>
<td>0.5</td>
<td>5</td>
<td>1.5</td>
<td>0.88</td>
</tr>
<tr>
<td>Udmurt</td>
<td>31</td>
<td>10.5</td>
<td>12.5</td>
<td>4.2</td>
<td>2.5</td>
<td>1</td>
<td>0.84</td>
</tr>
<tr>
<td>Hungarian</td>
<td>31</td>
<td>7</td>
<td>9</td>
<td>12</td>
<td>0</td>
<td>3</td>
<td>0.78</td>
</tr>
<tr>
<td>Ligurean</td>
<td>31</td>
<td>8</td>
<td>12</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>0.66</td>
</tr>
<tr>
<td>Hindi-Urdu</td>
<td>31</td>
<td>7.5</td>
<td>14</td>
<td>7.5</td>
<td>2</td>
<td>0</td>
<td>0.54</td>
</tr>
<tr>
<td>Turkish</td>
<td>31</td>
<td>9</td>
<td>17.5</td>
<td>2.5</td>
<td>0</td>
<td>1</td>
<td>0.51</td>
</tr>
<tr>
<td>Mongolian</td>
<td>31</td>
<td>6</td>
<td>22</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0.27</td>
</tr>
<tr>
<td>Indonesian</td>
<td>31</td>
<td>0</td>
<td>14</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0.04</td>
</tr>
<tr>
<td>English</td>
<td>31</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>25</td>
<td>3</td>
<td>94</td>
</tr>
<tr>
<td>Japanese</td>
<td>31</td>
<td>3.5</td>
<td>5.5</td>
<td>20.5</td>
<td>0.5</td>
<td>1</td>
<td>71</td>
</tr>
</tbody>
</table>

**Total:**

|     | 636| 243| 164.5| 128.5| 69| 319|

**Abbreviations:**

A = anticausative alternation  
C = causative alternation  
E = equipollent alternation  
L = liable alternation  
S = suppletive alternation  
A/C = ratio of anticausative to causative pairs  
% non-dir. = percentage of non-directed pairs  

Hasepalmath (1999: 101)
External possession

Haspelmath (1999: 109): „In an external-possession construction, a possessive modifier does not occur as a dependent constituent of the modifier NP, but NP-externally as a constituent of the clause.”

Die Mutter wusch dem Kind die Haare.
the mother washed the:DAT child the:ACT hairs
‘The mother washed the child’s hair.’

Internal NP-possession:

Die Mutter wusch das Haar des Kindes.
the mother washed the:ACC child the:GEN hairs
‘The mother washed the child’s hair.’
In far-west European languages such as English we cannot say "The mother washes the hair to the child"
External possession

The dative external possessive (Haspmath 2001: 1498):
- is found in Romance, Continental West Germanic, Balto-Slavic, Hungarian and Balkan languages (Greek, Albanian);
- Celtic & English and in Turkish & Lezgian have no external possessives;
- Eastern Indo-European languages (Kurdish, Hindi-Urdu, Persian) have no external possessives;
- **Dative external possessors seem to be very rare outside Europe** (only case mentioned in the paper is Ewe).
Negative pronouns and lack of verbal negation

**Two types:**

*V + NI (verb + negative indefinite), such as German:*

niemand fährt Fahrrad
Nobody drives bike

’nobody rides a bike’

*NV + NI (negated verb + negative indefinite), such as Polish:*

nikt nie jeździ na rowerze

nobody NEG rides by bicycle

’nobody rides a bike’

Plus the subtype of V+NI, where the negative particle appears under certain circumstances:

**Italian**

Nessuno guida una bici
Nobody drives a bike

Non guida nessuno una bici
NEG drives nobody a bike

’nobody drives a bike’
Map 107.8: Languages lacking verbal negation with a negative indefinite
Negative pronouns and lack of verbal negation

- The V + NI type is found „in French (if we disregard the particle ne), Occitan and all Germanic languages, as well as (in the mixed variety) in Ibero- and Italo-Romance and Albanian (but not in Romanian or other Balkan languages).” (Hauselmant 2001: 1498)

- The eastern European languages (Balto-Slavic, Finno-Ugrian, Turkic, Nakh-Daghestanian), with the exception of Georgian, have the NV + NI type;
- The Celtic languages have the NV + NI type;
- The Indic and Iranian languages have the NV + NI type;
- While languages outside of Europe may also have the V + NI type; **the NV + NI type is by far the most common pattern.**

Again, the SAE value is rare across world’s languages.
Six types of comparative constructions are recognized for world’s languages: only two types are attested in Europe, with the prevalence of the particle comparative type. In this type the marker of comparison (Cuzzolin & Lehmann 2004:1212) is a particle. The other type, the locative comparative, employs locative markers as markers of comparison: for instance, Estonian employs the elative case markers, which means ‘from’.

**Particles in comparative constructions**

**particle comparative**: Hungarian (SAE)

*István* *magasa-bb* *mint* *Peter*
*István.NOM* *tall-more* *than* *Peter.NOM*

‘*István is taller than Peter.’

**locative comparative**: Estonian (non-SE)

*kevad* *on* *sügis-est* *ilusam*
*spring* *is* *fall-ELA* *more.beautiful*

‘The spring is more beautiful than the fall.’

(in locative comparatives, the standard NP is invariably constructed in a case form which also has a locational/adverbial function)
Map 107.9: Comparative types in Europe
Feature 121A: Comparative Constructions

Values
- Lexical: 79
- Exceed: 53
- Compared: 34
- Inflate: 22

This feature is described in the text of chapter 12. Comparative Constructions to Lex Sassen. You may combine this feature with another one. Start typing the feature name or number in the field below.

Legend: Icon size: Show/Hide Labels

Map of the world with various markers indicating the feature's distribution across different locations.
Relative-based equative constructions

Haspelmath (2001: 1499): „In Europe, many languages have an equative construction that is based on an adverbial relative-clause construction.”

<table>
<thead>
<tr>
<th>Language</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>as pretty as you</td>
</tr>
<tr>
<td>German</td>
<td>so hübsch wie du</td>
</tr>
<tr>
<td>Portuguese</td>
<td>tão bonita como você</td>
</tr>
<tr>
<td>Czech</td>
<td>tak hezké jako ty</td>
</tr>
<tr>
<td>Hungarian</td>
<td>olyan csinos mint te</td>
</tr>
</tbody>
</table>

non-SAE languages of the world have rather different equative constructions, including the Eastern Indo-European languages.

The standard marker used in equative constructions has its diachronic source in the same item used for adverbial relative-clause constructions.
This is true for many European languages and rare across world’s languages.
Romance languages use forms deriving from Latin quo modo < Pt. como, It. Come, etc., while in Germanic forms are derived from the demonstrative, which was also used for adverbial relative-clause constructions.
Non-SAE languages use special equative standard markers (Caucasian languages), non-demonstrative adjectival markers (Celtic languages) or words meaning ‘equal’ (Scandinavian languages).
Subject person affixes as strict agreement markers: overt marking of subject pronouns

'Referential-agreement: pro-drop languages' – Spanish
Trabaj-as todos los días
work-2SG.PRS all DEF days
'you work every day'

'Strict-agreement: non-pro-drop languages' – German
du arbeit-est jeden Tag
2SG work-2SG.PRS every day
'you work every day'

Pro-drop: we may omit the subject
Non-pro drop: the subject is compulsory
In Spanish, the indexical marker –as refers to something not overtly expressed, while in German the indexical marker –est refers to something expressed i.e., the overt and compulsory subject.
We have four types:

1. Subject pronouns are compulsory and subject agreement markers are found on verbs;
2. Subject pronouns are compulsory and subject agreement markers are not found on verbs;
3. Subject pronouns are optional and subject agreement markers are found on verbs;
4. Subject pronouns are optional and subject agreement markers are not found on verbs.

The SAE-type is the first type, but many EU languages (as well as world languages) belong to the third type.
Feature 101A: Expression of Pronominal Subjects

This feature is described in the text of Chapter 101: Expression of Pronominal Subjects by Matthews & Dryer.

Values
- Obligatory pronoun in subject position: 83
- Subject affix on verb: 407
- Subject affix on variable head: 12
- Subject pronoun in different position: 87
- Optional pronoun in subject position: 81
- Voided: 13

Legend:
- Show/Hide Labels
- GeoBabel
Intensifier-reflexive differentiation

- **Intensifiers**: words that “characterize a noun phrase referent as central as opposed to an implicit or explicit periphery”: English *himself*, German *selbst*, French *même* and Russian *sam* (Haspelmath 2001:1501)

- **Reflexives**: words that must be referred (anaphora) related to another words (antecedents): English *himself* but German *sich*.

In many European languages **intensifiers** are differentiated from **reflexives**, as in the German examples:

*Sich* Muggel wie Sie sollten diesen freudigen, freudigen Tag feiern!

Mr. Dursley gab *sich* einen kleinen Ruck.

- **Across world’s languages, intensifiers and reflexives use the same form.** (however there are still not world-wide studies)
Incidentally not West Germanic languages such as Dutch and English.
More Standard Average European features

- Verb fronting in polar interrogatives
- Comparative marking of adjectives
- “A and-B” conjunction
- Comitative-instrumental syncretism
- Suppletive second ordinal

These are perhaps marginal, but for two of these, nice maps from WALS can be found.
Verb fronting in polar interrogatives

Polar interrogatives: yes/no questions
Verb fronting: the verb appears at the beginning of the sentence, in a reversed order with respect to the normal (declarative) order: VS vs. SV
The allegedly SAE value is identified with yellow dots: it seems that WALS map confirm the Europeanism, as the pattern is attested only in one language outside Europe i.e., Malay. However some ‘core’ SAE languages such as French and Italian do not display the SAE value.
Comitative-instrumental syncretism

- Syncretism: two or more functions are formally expressed by the same (case) marker.

Remember? We have seen examples of this in slide no. 12, but according to Croft this syncretism was not exclusive feature of European language!
More Standard Average European features

- Lack of an alienable/inalienable opposition in adnominal possession
- Lack of an inclusive/exclusive opposition in first person non-singular pronouns
- Lack of reduplicating constructions
- Discourse pragmatic notions such as topic and focus are expressed primarily by sentence stress and word order differences
- SVO basic word order at the level of the clause
- Tend to have just one converb
- Usually have a special construction for negative coordination
- Have a large number of characteristic properties in the area of phasal adverbials
- “Preterite decay”
Lack of an alienable/inalienable opposition in adnominal possession

Haspelmath (2001: 1503): „In Nichols’s (1992) world-wide sample, almost half of the languages show such an opposition, but no European language does (1992: 123). More generally, this opposition is rarer in the Old World and common in the New World, but in Europe it is even less common than in Africa and Asia.“

**Mesa Grande Diegueño (Cochimi-Yuman):**

1. ?-atal\textsuperscript{y}
   
   1sg-mother ‘my mother’

2. ?a-n\textsuperscript{y}-ewa:
   
   1sg-ALIENABLE-house ‘my house’

*Nichols and Bickel (2005)*
Feature 59A: Possessive Classification

This feature is classified in the text of chapter 59, Possessive Classification, by Johanna Nichols and Batjak.

Values

- No possessive classification: 135
- Two classes: 94
- Three to five classes: 30
- More than five classes: 4

You may combine this feature with another one. Start typing the feature name or number in the field below.

Nichols and Bickel 2005
Lack of an inclusive/exclusive opposition in first person non-singular pronouns

Haspelmath (2001: 1503): “Lack of an inclusive/exclusive opposition in first person non-singular pronouns. Again, this opposition is commonest in the New World and in the Pacific region, but in Europe it is even rarer than in Africa and Asia, as was shown by Nichols (1992: 123).”

Malagasy (Austronesian) 1st person pronouns:
_isika _we, I and you’ (including the addressee)
_izay _we, I and some others, but not you’ (excluding the addressee)
Feature 39A: Inclusive/Exclusive Distinction in Independent Pronouns

This feature is described in the text of chapter 39, "Inclusive/Exclusive Distinction in Independent Pronouns," by Michael Cy pass (2010).

Values

- No ‘we’ 2
- ‘we’ the same as ‘I’ 11
- No inclusive/exclusive 313
- Only exclusive 5
- Inclusive ‘we’ alone 63

Legend:
- Icon size: Small
- Show/Hide Labels

Map of the world with markers indicating the presence of the feature across different regions.
Lack of reduplicating constructions

Haspelmath (2001: 1503): „Lack of reduplicating constructions. I have no systematic evidence to back up the claim that this is a characteristic feature of European languages, but reduplication is so common across languages that its almost total absence in the core European languages becomes striking. (Interestingly, reduplication existed in older Indo-European languages at least in one construction, the perfect, but even here it was lost entirely by the Middle Ages.)“

Rubino (2005): „Full reduplication is the repetition of an entire word, word stem (root with one or more affixes), or root. Examples are Nez Perce (Sahaptian; northwestern United States) full word lexical reduplication: těːmʊl ‘hall’ vs. temultéːmʊl ‘rain and snow mixed’ (Aoki 1963: 43), or Tagalog full root reduplication, shown here with the verbalizing prefix mag-, where the reduplicant isip is identical to the base isip ‘think’: mag-isip ‘to think’ vs. mag-isip-isip ‘to think about seriously.’“
Feature 27A: Reduplication

This feature is described in the text of chapter 27 by Carl Rubin.

You may combine this feature with another one. Start typing the feature name or number in the field below.

Submit
SVO basic word order at the level of the clause

Haspelmath (2001: 1504): „SVO basic word order at the level of the clause. This feature is of course found elsewhere in the world, but in Europe it correlates particularly well with the other SAE features. The Celtic languages in the west have VSO order (except for Breton, which is also otherwise more SAE than Irish and Welsh), and the eastern languages have SOV word order. Interestingly, Balto-Finnic (Finnish, Estonian, etc.) and (less unequivocally) Hungarian have SVO word order, whereas the eastern Uralic languages have SOV. Similarly, the eastern Indo-European languages tend to show SOV word order. (See Dryer 1998 for more on word order in the languages of Europe.)“
Feature 81A: Order of Subject, Object and Verb

This feature is described in the field of chapter in "Order of Subject, Object and Verb" for Michael B. Ogden, etc.

You may combine this feature with another one. Start typing the feature by name or number in the field below.

Legend:
- SOV
- SVO
- VSO
- VOS
- OVS
- OSV
- OVS
- No specific order

Values:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SOV</td>
<td>86</td>
</tr>
<tr>
<td>SVO</td>
<td>48</td>
</tr>
<tr>
<td>VSO</td>
<td>46</td>
</tr>
<tr>
<td>VOS</td>
<td>21</td>
</tr>
<tr>
<td>OVS</td>
<td>11</td>
</tr>
<tr>
<td>OSV</td>
<td>4</td>
</tr>
<tr>
<td>No specific order</td>
<td>109</td>
</tr>
</tbody>
</table>
Isogloss: geographic boundary of a linguistic feature. The boundaries between dialects are usually consist of bundles of isoglosses. Sometimes they diverge, such as with the Rhenish fan, where several different dialects of West Central German are found.

The Rhenish fan: 1 Dutch (West Low Franconian),
2 Limburgian (East Low Franconian),
3 Ripuarian Franconian,
4 & 5 Mosel Franconian,
6 Rhenish Franconian
Haspelmath 2011: The map combines nine features of § 2.:
1. definite and indefinite articles;
2. relative clauses with relative pronouns;
3. ‘have’-perfect;
4. participial passive;
5. dative external possessors;
6. negative pronouns and lack of verbal negation;
7. relative-based equative constructions;
8. subject person affixes as strict agreement markers;
9. intensifier- reflexive differentiation.

The languages in the core (French and German) show the SAE value in all nine of these features. The languages in the next layer (Dutch, other Romance, Albanian) show eight features, the next layer (English, Greek, Romanian) shows seven features, and so on. In this map, the resulting picture is actually very clear, because the SAE area with at least five SAE features stands out from the remaining languages, which have at most two SAE features.
Degrees of membership in SAE

- core: German and French
- just outside the core: Dutch, Spanish, Portuguese, Italian, Albanian, Sardinian, English, Romanian, Greek
- further removed but within SAE: Icelandic, Faroese, Norwegian, Swedish, Czech, Polish, Latvian, Lithuanian, Slovene, Serbo-Croatian, Ukrainian, Russian and other Slavonic languages
- Marginal: Hungarian, Balto-Finnic, Maltese
- Outside SAE: Celtic, Basque, Turkish, etc.
The origin of Standard Average European

Which contact situation gave rise to SAE? And who borrowed what from whom? Haspelmath (2001: 150ff) gives five scenarios:

(i) retention of Proto-Indo-European structures and assimilation of some non-Indo-European languages to Indo-European language structure;
(ii) influence from a common substratum of a pre-Indo-European population in Europe;
(iii) contacts during the great transformations at the transition from late antiquity to the early Middle Ages in Europe;
(iv) the official language (Latin) and the common European culture of the Middle Ages;
(v) the common European culture of modern times, from the Renaissance to the Enlightenment."
The origin of Standard Average European

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

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

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

„The second possibility, a pre-Indo-European substratum in Europe causing the SAE features, would be extremely difficult to demonstrate, but it might be worth pursuing. It is intriguing to note that the geographical space occupied by SAE languages coincides fairly precisely with the area of the Old European hydronymy, i.e. the homogeneous layer of river names discovered by Hans Krahe (see Vennemann 1994 for recent discussion).“

What is a substratum? -&gt; -&gt; -&gt;
The origin of Standard Average European

- **stratum** (Latin for 'layer')/strate: language that influences/is being influenced by another language
- **substrate**, 'below layer': the influencing (source) language has lower prestige as the language that is being influenced (target); in many cases, the substrate language is unknown, but traces of it can still be seen in the prestige language adopted by its speakers;
- **superstrate**, 'above layer': the influencing (source) language has higher prestige as the language that is being influenced (target); for example the Greek and Latin borrowings for scientific terms in many European languages.

Normally, the substratum is the ‘conquered’ language
The origin of Standard Average European

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

This era seems to be the right time-frame for many of the SAE features to evolve;

Language contact must have been pervasive at this time of great migrations.
The origin of Standard Average European

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

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

Double negation and non-resumptive pronouns are indeed widely attested in vernacular languages.
The origin of Standard Average European

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

„The fifth possibility must be rejected because a time depth of 300-500 years is not sufficient to account for grammatical commonalities of the kind discussed above. If lexical similarities between the European languages are discussed [...] then the last several centuries are the appropriate time frame for explaining the historical links, but the basic syntactic structures common to SAE languages must be older.“
References


Thomsen, Sarah (2000), "Linguistic areas and language history", in: Gilkes, Elly; Neuhoff, John; Schaefer, Ios (eds.), Language in Context, Amsterdam: Rodopi, pp. 311-327.