

Structural Challenges for Typological Databases

Michael Cysouw
Martin Haspelmath



Max Planck Institute
for Evolutionary Anthropology
Department of Linguistics

Collaborative Data Collection

- Collection of structured information for a wide variety of languages

- Collection of structured information for a wide variety of languages
- that allows for different kinds of collaboration
 - ▶ hierarchically organised research projects
 - ▶ more open, wiki-style collaboration
 - ▶ of both specialists and comparativists
 - ▶ of both single-feature in-depth typological research and multi-feature broad-scale language comparison

- Collection of structured information for a wide variety of languages
- that allows for different kinds of collaboration
 - ▶ hierarchically organised research projects
 - ▶ more open, wiki-style collaboration
 - ▶ of both specialists and comparativists
 - ▶ of both single-feature in-depth typological research and multi-feature broad-scale language comparison
- that allows reuse, without necessarily having to take over data wholesale

Тавтайморилно уу
Salam alaikum 歡迎
Добро пожаловать
Welcome
ᠨᠭᠣ ᠬᠣᠶ᠋ᠠ
Ngö höya

Max Planck Institute
for Evolutionary Anthropology
Department of Linguistics

Coding Comparative Information

The extent and distribution of variation
is the object of investigation

The extent and distribution of variation
is the object of investigation

so, do not assume that

The extent and distribution of variation
is the object of investigation

so, do not assume that

- you know what is a language

The extent and distribution of variation
is the object of investigation

so, do not assume that

- you know what is a language
- your interpretation of language-particular data is clear to others

The extent and distribution of variation
is the object of investigation

so, do not assume that

- you know what is a language
- your interpretation of language-particular data is clear to others
- a language is a coherent entity

Three proposals

Three proposals

- replace ***language***
by ***source*** (“doculect”)
- replace ***subsidiary notes***
by ***notes as the primary information***
- replace ***typology of languages***
by ***typology of constructions***

Тавтайморилно уу
Salam alaikum 歡迎
Добро пожаловать
Welcome
ᠨᠭᠣ ᠬᠣᠶ᠋ᠠ
Ngö höya

Max Planck Institute
for Evolutionary Anthropology
Department of Linguistics

Coding Comparative Information

- replace **language**
by **source** (“doculect”)

- replace **language** by **source** (“doculect”)
 - ▶ sources are in turn linked to languages

- replace **language** by **source** (“doculect”)
 - ▶ sources are in turn linked to languages
 - ▶ different people can use different source-language linkage

- replace **language** by **source** (“doculect”)
 - ▶ sources are in turn linked to languages
 - ▶ different people can use different source-language linkage
- replace **subsidiary notes** by **notes as the primary information**

- replace **language** by **source** (“doculect”)
 - ▶ sources are in turn linked to languages
 - ▶ different people can use different source-language linkage
- replace **subsidiary notes** by **notes as the primary information**
 - ▶ classifications in a restricted vocabulary are often difficult to reconstruct by others

- replace **language** by **source** (“doculect”)
 - ▶ sources are in turn linked to languages
 - ▶ different people can use different source-language linkage
- replace **subsidiary notes** by **notes as the primary information**
 - ▶ classifications in a restricted vocabulary are often difficult to reconstruct by others
 - ▶ human-readably text (“notes”) is more informative — through text-mining also for computers!

- replace **language** by **source** (“doculect”)
 - ▶ sources are in turn linked to languages
 - ▶ different people can use different source-language linkage
- replace **subsidiary notes** by **notes as the primary information**
 - ▶ classifications in a restricted vocabulary are often difficult to reconstruct by others
 - ▶ human-readably text (“notes”) is more informative — through text-mining also for computers!
 - ▶ the minimal note will reference to a page in a source

Тавтайморилно уу
Salam alaikum 歡迎
Добро пожаловать
Welcome
ᠨᠭᠣ ᠬᠣᠶ᠋ᠠ
Ngö höya

Max Planck Institute
for Evolutionary Anthropology
Department of Linguistics

Coding Comparative Information

- replace ***typology of languages***
by ***typology of constructions***

- replace ***typology of languages***
by ***typology of constructions***
 - ▶ if there is only one relevant construction per language,
these two perspectives coincide

- replace ***typology of languages*** by ***typology of constructions***
 - ▶ if there is only one relevant construction per language, these two perspectives coincide
 - ▶ different people have different opinions about what counts as a construction, but that is a language-particular discussion, not a comparative problem

- replace ***typology of languages*** by ***typology of constructions***
 - ▶ if there is only one relevant construction per language, these two perspectives coincide
 - ▶ different people have different opinions about what counts as a construction, but that is a language-particular discussion, not a comparative problem
 - ▶ Ultimately, each individual example can be seen as a construction (token)

- replace ***typology of languages*** by ***typology of constructions***
 - ▶ if there is only one relevant construction per language, these two perspectives coincide
 - ▶ different people have different opinions about what counts as a construction, but that is a language-particular discussion, not a comparative problem
 - ▶ Ultimately, each individual example can be seen as a construction (token)
 - ▶ different people can have different opinions how the construction tokens are combined to construction types in a particular language