Max Planck Institute
for Evolutionary Anthropology
Department of Linguistics

Structural Challenges for Typological Databases

Michael Cysouw
Martin Haspelmath

Max Planck Institute
for Evolutionary Anthropology
Department of Linguistics

- Collection of structured information for a wide variety of languages

Max Planck Institute
for Evolutionary Anthropology
Department of Linguistics

- 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

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

Max Planck Institute
for Evolutionary Anthropology
Department of Linguistics

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

Max Planck Institute
for Evolutionary Anthropology
Department of Linguistics

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

Max Planck Institute
for Evolutionary Anthropology
Department of Linguistics

Three proposals

- replace *language*
by *source* ("doculect")

- replace *subsidiary notes*
by *notes as the primary information*

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

Max Planck Institute
for Evolutionary Anthropology
Department of Linguistics

- 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

Max Planck Institute
for Evolutionary Anthropology
Department of Linguistics

- 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*

Max Planck Institute
for Evolutionary Anthropology
Department of Linguistics

- 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!

Max Planck Institute
for Evolutionary Anthropology
Department of Linguistics

- 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

Max Planck Institute
for Evolutionary Anthropology
Department of Linguistics

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

Max Planck Institute
for Evolutionary Anthropology
Department of Linguistics

- 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)

Max Planck Institute
for Evolutionary Anthropology
Department of Linguistics

- 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