

**Workshop on the Feasibility of a Web-based Database
of the Syntactic Structures of the World's Languages**

WALS

The World Atlas of Language Structures



Hans-Jörg Bibiko

**Max Planck Institute for evolutionary Anthropology – Department of Linguistics
Leipzig, Germany**

WALS – the book

- Editors:
Martin Haspelmath
Matthew S. Dryer
David Gil
Bernard Comrie
- Interactive Electronic Version:
Hans-Jörg Bibiko



WALS – the book

- an atlas with 142 world maps, showing languages as dots
- each world map shows a single structural feature
- each map was contributed by an author (or team of authors)
- each map is accompanied by a text explaining the feature values and discussing the results

WALS – the book

- 44 authors or author teams
- altogether 2560 different languages
- a core sample of 200 language
- altogether about 58.000 data points
- 6.700 bibliographical references

WALS – the book

Phonology	19	Maddieson, van der Hulst
Morphology	10	Nichols
Nominal Categories	28	Corbett, Cysouw, Dryer
Nominal Syntax	7	Nichols, Gil
Verbal Categories	16	Dahl, van der Auwera
Word Order	17	Dryer
Simple Clauses	24	Comrie, Siewierska, Polinsky
Complex Clauses	7	Comrie, Cristofaro
Lexicon	10	Brown, Kay, Nichols
Others	4	Zeshan, Comrie, Gil

WALS – the program

- Interactive Reference Tool
- All maps are scalable vector graphics
- Interactive maps can be zoomed, dragged, symbols can be customized
- Different display options (rivers, boundaries)
- Mouse-over effect for showing the corresponding language names



WALS – the program

- Language profile with geographical, genealogical information, alternative names and overview of typological features
- 141 predefined customisable feature maps
- Each language-feature pair with data sources
- Generation of compound features of typological, genealogical, geographical data manually
- Import/Export functions

Follow-up – preface

- What is “The Goal” of the project?
 - a clear, concrete, (narrowed) vision
- No database is able to mirror the entire linguistic reality

Follow-up – preface

- What is a “language”?
“We can more easily agree on the existence of a description than on the identification of a language.”
- Language-feature pair
versus
Construction-feature pair

Follow-up – preface

- Usage of Open Source software
- Don't trust 'standards' in respect of primary keys in the database
- Naming of values/properties
 - logically independent properties and values (open-endedness)

Follow-up – preface

- How to cite the database?
 - reconstruction of dynamic data
(unique URI)
 - long-term accessibility/archiving
 - reputation of authors

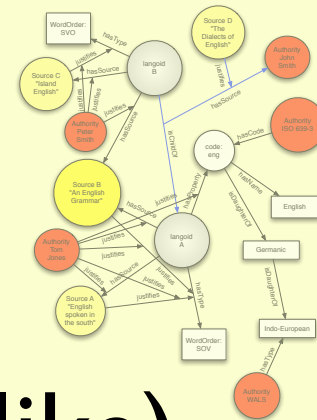
Follow-up – preface

Databases in general not only store large amounts of data, but also **impose an organization in data**, which facilitates access for researchers and applications developers.

A 'linguistic database' organizes collections of linguistic descriptions of data, and it is often the case that such descriptions are reflected in contradictory information embodied by contested hypotheses.

- How to mirror the linguistic reality?
- Nobody can predict the future.

Follow-up – database



- Flexible relational database (graph-like)
field A ; field B ; field C := linkage of A and B
- Each data point also has a time dimension
(correctness and evolution)

Follow-up – database

- Unicode

	Unicode Variants	UTF8 bytes
ô	ô +.	C3B4CCA3
	ọ +^	E1BB8DCC82
	o +. + ^	6FCCA3CC82
	o +^ +.	6FCC82CCA3
	ô	E1BB99

- Normalization

(www.unicode.org/reports/tr15/)

Follow-up – database

- Example sentences – several glosses are possible
- Multi-lingual glossing/translations
‘A’ → Latin, Greek, Cyrillic, ...

Follow-up – database

- **Interoperability**
 - how to link languages
(e.g. *Serbo-Croatian*)
 - problematics of broken links
 - queries using language families etc.

“ad hoc thoughts”

- OpenID (www.openid.org)
- Usage of RDF
(Resource Description Framework)
- Identifying of languages – ISO 639-3⁴⁵⁶
IETF BCP 47 RFC 4647 (www.langtag.net)
Internet Engineering Task Force Best Current Practice 47 Request For Comments 4647
- Software developer with linguistic awareness

Just do it

— step by step —

but be flexible!

Just do it

– step by step –

but be flexible!

Thank you very much
for your attention!