

Language Science Press & Cross-Linguistic Linked Data

Sebastian Nordhoff & Mathias Schenner 2015-04-25

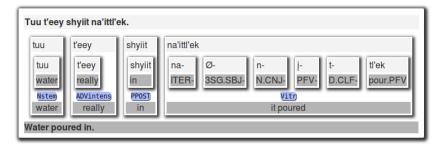
Language Science Press



Langs



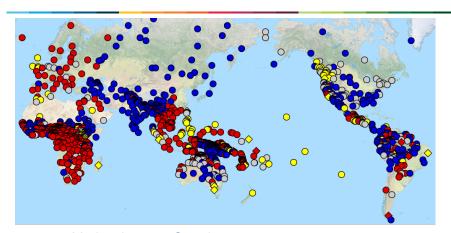
Linguistische Beispiele



-) Übersetzung:
 -) Satz-für-Satz

 - Morphem-für-Morphem

Cross-Linguistic Linked Data | CLLD | The World Atlas of Language Structures | Typologie



-) 180 Merkmale; 2500 Sprachen
- hier: Wortstellung (SOV, SVO, VSO)
-) geographische und genealogische Information

LangSci



- Apics The Atlas of Pidgin and Creole Language Structures
- **eWAVE** The Electronic World Atlas of Varieties of English
- > SAILS The South American Indigenous Language Structures
- PHOIBLE The world's largest database of phonological inventories

language science IDS: The Intercontinental Dictionary Series

CLLD |Lexikon



-) Das Wort für 'Wasser' im IDS
-) 1429 Bedeutungen; 288 Sprachen

- Numerals Numerals in the World's Languages
-) WOLD The World Loanword Database





-) ca. 50k Beispiele
- ca. 500k typologische Datenpunkte
-) ca. 25k Sprachen/Dialekte mit Klassifikation und Literaturangaben
-) ca. 250k Literaturangaben

Formate und Lizenzen

- > csv
-) xml
-) rdf
-) bibtex
- > CC-BY(-SA)

Denkbare Anwendungen

-) Wörterbuch
 - Extraktion der Übersetzungen aus den Beispielsätzen
- \ Lustiges Sprachenraten
 - Zufälliger Beispielsatz
 - Punktzahl nach genealogischem oder geographischem Abstand
- Sprachenbrowser
 - Aggregation aller verfügbarer Datenpunkte
-) Browsebare Visualisierung des typologischen Vektorraums
 - http://citeseerx.ist.psu.edu/viewdoc/download?doi= 10.1.1.50.2428&rep=rep1&type=pdf