

Computergestützte historische Linguistik

Frederic Blum

MPI-EVA, Leipzig



24.02.2024
Workshopsreihe 'Digitale Methoden'

Intro

Punkte

- Fachgebiet
- Konkrete Erwartungen an den Kurs
- 'Bitte-nicht'-Erwartungen
- Interesse an bestimmter Sprachfamilie

Überblick

- Erstellung eines Lexibank-datasets (Digitalisierung von Daten)
- Annotation der Daten in EDICTOR
- Manuelle Auswertung der Annotation
- Automatisierte Auswertung der Annotation

Historische Linguistik

Große Sprachfamilien der Welt

- ~7000 Sprachen weltweit
- Indo-Europäisch: <https://glottolog.org/resource/languoid/id/indo1319>
- Austronesisch: <https://glottolog.org/resource/languoid/id/aust1307>
- Arawak: <https://glottolog.org/resource/languoid/id/araw1281>
- Bantu: <https://glottolog.org/resource/languoid/id/narr1281>

Ähnlichkeiten zwischen Wörtern: Woher und wieso?

Ähnlichkeiten zwischen Wörtern: Woher und wieso?

- Zufall, Kontakt, Verwandtschaft

Ähnlichkeiten zwischen Wörtern: Woher und wieso?

- Zufall, Kontakt, Verwandtschaft
- Beispiele aus etymologischen Wörterbuch für Sprachwandel?

Ähnlichkeiten zwischen Wörtern: Woher und wieso?

- Zufall, Kontakt, Verwandtschaft
- Beispiele aus etymologischen Wörterbuch für Sprachwandel?

(2)	English	Latin	German	Kannada
	<i>mouse</i>	<i>mūs</i>	<i>Maus</i>	<i>ili</i>
	<i>father</i>	<i>pater</i>	<i>Vater</i>	<i>appa</i>
	<i>three</i>	<i>trēs</i>	<i>drei</i>	<i>muru</i>
	<i>fish</i>	<i>piscis</i>	<i>Fisch</i>	<i>minu</i>

Figure 1: Verwandte Wörter in Indo-Europäischen Sprachen (McMahon 1994)

Ähnlichkeiten zwischen Wörtern: Zufall

- Manche Formen sind nur aus Zufall ähnlich

(5)	English <i>man</i> , German <i>nass</i> , Italian [dɔnna],	Korean <i>man</i> Zuni <i>nas</i> Japanese [onzna]	'man' 'wet' 'lady'
-----	--	--	--------------------------

Figure 2: Wörter mit ähnlichen Formen und Bedeutungen in nicht-verwandten Sprachen (McMahon 1994)

Ähnlichkeiten zwischen Wörtern: Kontakt

- *vinum (Latein) > Wein/vine
- *strata (Latein) > Straße/street

Ähnlichkeiten zwischen Wörtern: Verwandtschaft

- Verbindung zwischen Form und Bedeutung ist (meistens) arbiträr

Ähnlichkeiten zwischen Wörtern: Verwandtschaft

- Verbindung zwischen Form und Bedeutung ist (meistens) arbiträr
- Evaluation regelmäßiger Lautkorrespondenzen zwischen Sprachen als Evidenz für gemeinsame Abstammung

- Eine Gruppe junger Forscher in Leipzig
- Systematische und wissenschaftliche Untersuchung des Sprachwandels nach strengen methodologischen Kriterien, die bis heute halten
- Vorher: Oft Referenz auf ‘Urform’ der Sprache, welche verunstaltet wird durch Wandel

- Eine Gruppe junger Forscher in Leipzig
- Systematische und wissenschaftliche Untersuchung des Sprachwandels nach strengen methodologischen Kriterien, die bis heute halten
- Vorher: Oft Referenz auf ‘Urform’ der Sprache, welche verunstaltet wird durch Wandel

Regularitätsprinzip

- Regularitätsprinzip: Aller Lautwandel ist regelmäßig
- Operationalisiert: Es ist dann Lautwandel, wenn es regelmäßig ist (Hoenigswald 1978)

Kognate

- Wörter die regelmäßige und wiederholte Lautkorrespondenzen zwischen mehreren Sprachen sowie eine ähnliche Bedeutung zeigen

Kognate

- Wörter die regelmäßige und wiederholte Lautkorrespondenzen zwischen mehreren Sprachen sowie eine ähnliche Bedeutung zeigen
- Annahme: Die Wörter haben einen gemeinsamen 'Vorfahren'

Sprachfamilien

- Verwandte Sprachen waren einmal die *gleiche Sprache* (Proto-Sprache)
- Beispiel: Proto-Indo-Europäisch

Rekonstruktionen

- Rekonstruktion vorhergehender Sprachstufen (markiert mit Asterisk '*')
- Ableitung von Lautwandel-'Gesetzen'

Rekonstruktionen

- Rekonstruktion vorhergehender Sprachstufen (markiert mit Asterisk '*')
- Ableitung von Lautwandel-'Gesetzen'
- Beispiele:
 - *k > h zwischen Vokalen in Tacana und Warisa
 - *u > o in Ese Ejja

Proto-Takana	d	u	k	u
Kaviñena	d	u	k	u
Tacana	d	u	h	u
Warisa	d	o	h	o
Ese Ejja	d	o	x	o

Table 1: Wörter für INNEN in verschiedenen Takana-Sprachen (Girard 1971)

Lautinventar der Sprache

- Merge: Zwei Laute verbinden sich zu einem (*m und *n werden zu /n/)

Lautinventar der Sprache

- Merge: Zwei Laute verbinden sich zu einem (*m und *n werden zu /n/)
- Split: Ein Laut wird zu zwei Lauten
 - Oft eine kontextabhängige Variation
 - Beispiel: /n/ vor /p/ wird /m/, aber bleibt /n/ in allen anderen Fällen

Lautinventar der Sprache

- Merge: Zwei Laute verbinden sich zu einem (*m und *n werden zu /n/)
- Split: Ein Laut wird zu zwei Lauten
 - Oft eine kontextabhängige Variation
 - Beispiel: /n/ vor /p/ wird /m/, aber bleibt /n/ in allen anderen Fällen
 - Secondary Split: Neuer Laut im Inventar der Sprache
 - Primary split: 'Merge' mit anderem Laut
- Shift: Isolierter Lautwandel, ohne Merge/Split

CLDF

Referenzkataloge

- Glottolog
- Concepticon
- CLTS

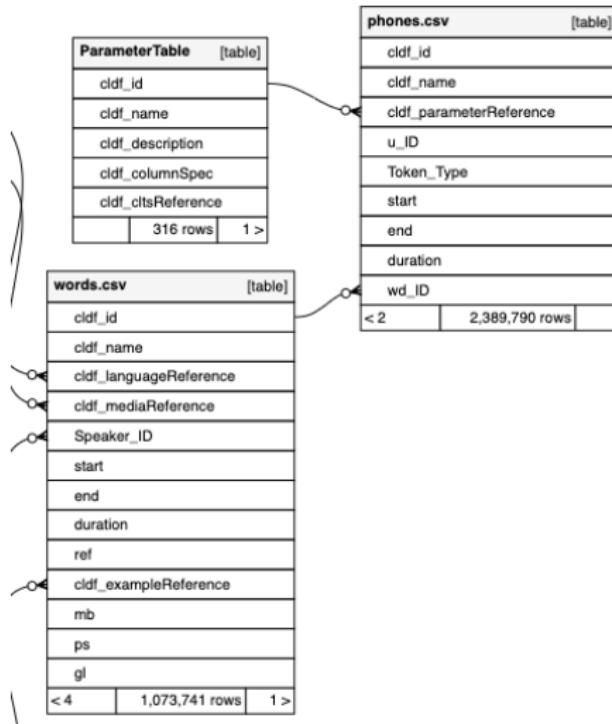


Figure 3: CLDF als relationale Datenbanken (Forkel et al. 2018)

Beispiele für Studien

Expansion der Bantu-Sprachen

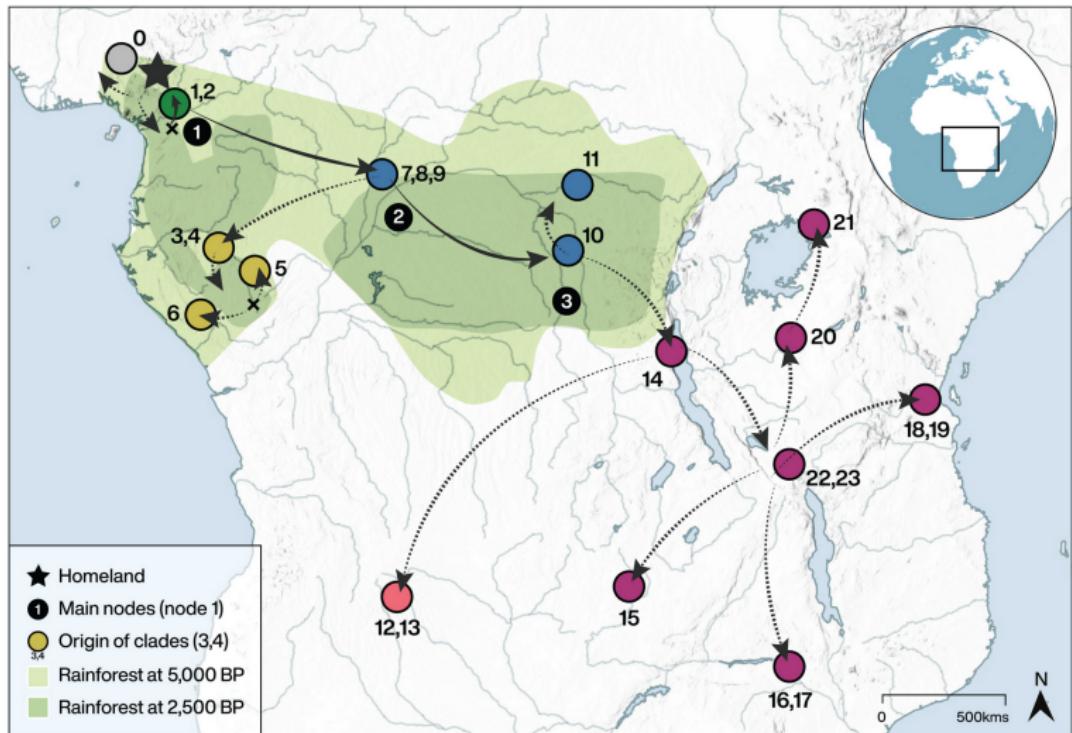


Figure 4: Ausbreitung der Bantu-Sprachen in Afrika (Koile et al. 2022)

Genetische und linguistische Evolution

PNAS

RESEARCH ARTICLE

EVOLUTION

OPEN ACCESS



A global analysis of matches and mismatches between human genetic and linguistic histories

Chiara Barbieri^{a,b,c,1,2}, Damián E. Blasíć^{d,e,2}, Epifanía Arango-Isaza^{a,b}, Alexandros G. Sotiropoulos^f, Harald Hammarström^g, Søren Wichmann^h, Simon J. Greenhill^{c,i}, Russell D. Gray^c, Robert Forkel^{c,3}, Balthasar Bickel^{b,j,3}, and Kentaro K. Shimizu^{a,b,k,3}

Edited by Marcus Feldman, Stanford University, Stanford, CA; received December 10, 2021; accepted October 10, 2022

Figure 5: Matches und mismatches in genetischer und linguistischer Geschichte (Barbieri et al. 2022)

Variation of Emotion Semantics

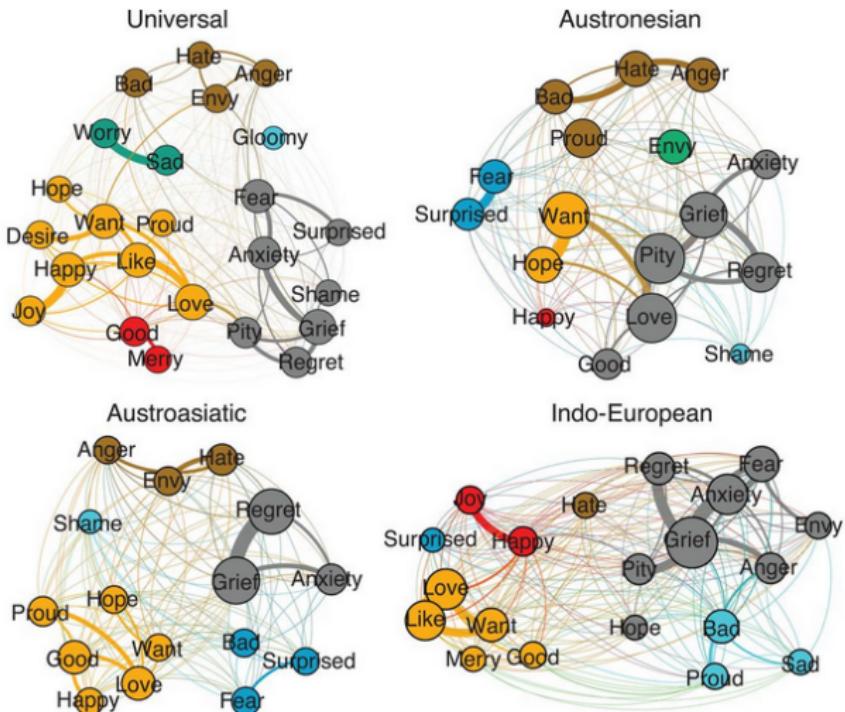


Figure 6: Colexification of emotion concepts (Jackson et al. 2019)

References I

- Barbieri, Chiara, Damián E. Blasi, Epifanía Arango-Isaza, Alexandros G. Sotiropoulos, Harald Hammarström, Søren Wichmann, Simon J. Greenhill, Russell D. Gray, Robert Forkel, Balthasar Bickel & Kentaro K. Shimizu. 2022. A global analysis of matches and mismatches between human genetic and linguistic histories. *Proceedings of the National Academy of Sciences* 119(47). <https://doi.org/10.1073/pnas.2122084119>.
- Forkel, Robert, Johann-Mattis List, Simon J. Greenhill, Christoph Rzymski, Sebastian Bank, Michael Cysouw, Harald Hammarström, Martin Haspelmath, Gereon A. Kaiping & Russell D. Gray. 2018. Cross-Linguistic Data Formats, advancing data sharing and re-use in comparative linguistics. *Scientific Data* 5(1). 1–10. <https://doi.org/10.1038/sdata.2018.205>.
- Girard, Victor. 1971. *Proto-Takanan Phonology*. Berkeley, Los Angeles, London: University of California Press.

References II

- Hoenigswald, Henry M. 1978. The annus mirabilis 1876 and posterity. *Transactions of the Philological Society* 76(1). 17–35.
<https://doi.org/10.1111/j.1467-968x.1978.tb01024.x>.
- Jackson, Joshua Conrad, Joseph Watts, Teague R. Henry, Johann-Mattis List, Robert Forkel, Peter J. Mucha, Simon J. Greenhill, Russell D. Gray & Kristen A. Lindquist. 2019. Emotion semantics show both cultural variation and universal structure. *Science* 366(6472). 1517–1522.
<https://doi.org/10.1126/science.aaw8160>.
- Koile, Ezequiel, Simon J. Greenhill, Damián E. Blasi, Remco Bouckaert & Russell D. Gray. 2022. Phylogeographic analysis of the bantu language expansion supports a rainforest route. *Proceedings of the National Academy of Sciences* 119(32).
<https://doi.org/10.1073/pnas.2112853119>.

References III

McMahon, April M. S. 1994. *Understanding Language Change*.
Cambridge University Press.
<https://doi.org/10.1017/cbo9781139166591>.