



WALS Datenbank

Zihang Sun
Profilierungsmodul Computerlinguistik I
CIS, LMU
Dr. Robert Zangenfeind
24.01.2024



Übersicht

1. WALS Datenbank
 - a. Entwicklung der WALS Datenbank
 - b. Die Vorstellung von WALS online
2. Fallstudie: Der Vergleich von Deutsch und Englisch auf der phonologischen Ebene in WALS
 - a. Phonetik vs. Phonologie
 - b. Beispiel: Deutsch vs. Englisch
3. Ein Beispiel der Nutzung der WALS Datenbank in der Sprachmodelle-Forschungen
4. Fazit



Die Entwicklung der WALS Datenbank (Dryer et al., 2013)

- *WALS: The World Atlas of Language Structures*
- Eine Datenbank von Strukturellen (phonologischen, grammatischen, lexikalischen ...) Merkmalen von Sprachen, die in veröffentlichten Büchern gesammelt wurden.
- Zunächst wurde die WALS Datenbank im Jahr 2005 als ein Buch bei Oxford University Press veröffentlicht. Die erste online-Version von WALS erschien im Jahr 2008 und seitdem gibt es mehrere Updates und neue Versionen von der WALS Datenbank.

Die Vorstellung von **WALS** online:

Chapter/Kapitel (Dryer et al., 2013)

- Es wird in dem Bereich “Chapter/Kapitel” so strukturiert genau wie im Buch. Jedes Kapitel beschreibt detailliert eine linguistische Eigenschaft.

THE WORLD ATLAS
OF LANGUAGE STRUCTURES
ONLINE



Home Features Chapters Languages References Authors

Chapters

Showing 1 to 100 of 152 entries

← Previous 1 2 Next →



Id	Name	Author	Area	Cite
	<input type="text" value="Search"/>		--any--	
	Introduction	Bernard Comrie and Matthew S. Dryer and David Gil and Martin Haspelmath		cite
1	Consonant Inventories	Ian Maddieson	Phonology	cite
2	Vowel Quality Inventories	Ian Maddieson	Phonology	cite
3	Consonant-Vowel Ratio	Ian Maddieson	Phonology	cite
4	Voicing in Plosives and Fricatives	Ian Maddieson	Phonology	cite
5	Voicing and Gaps in Plosive Systems	Ian Maddieson	Phonology	cite
6	Uvular Consonants	Ian Maddieson	Phonology	cite
7	Glottalized Consonants	Ian Maddieson	Phonology	cite
8	Lateral Consonants	Ian Maddieson	Phonology	cite
9	The Velar Nasal	Gregory D.S. Anderson	Phonology	cite

Chapter Consonant Inventories

by Ian Maddieson [cite](#)

1. Introduction

This chapter and the next few chapters will look at various aspects of the complexity of the sound resources used in the world's languages and examine how this complexity is distributed geographically. The first aspect to be examined is the size of the set of consonants used in the language, usually referred to as the **consonant inventory**. This is one element of what is called the **phonology** of the language.

Values of Map 1A. Consonant Inventories

[Go to map](#)

Value	Representation
<input checked="" type="radio"/> Small	89
<input type="radio"/> Moderately small	122
<input type="radio"/> Average	201
<input type="radio"/> Moderately large	94
<input type="radio"/> Large	57
Total:	563

It is usually possible to agree for any given language on a set of elements which are considered to be the **speech sounds** used in that language. The most important consideration in deciding on this set is to find groups of words which sound different from each other by the smallest degree sufficient to make them distinct words of the language. For example, the English one-syllable words *pin, tin, kin, fin, thin, sin, shin* are part of a set which differ by beginning in different ways, *dim, din, ding, did, dig, dish* are part of a set which differ by ending in different ways and *pin, pen, pan, pun, pain, pine, pawn* are part of a set which differ in the middle of the syllable. From a series of such comparisons a list of candidate speech sounds for the language will emerge. Generally the set of those which can appear at the beginnings and ends of syllables will be unlike those which can occur in the middle, hence a distinction is made between **consonants** (sounds typically occurring at the syllable margins) and **vowels** (sounds typically occurring in the syllable centers). In this chapter only consonants will be discussed.

Die Vorstellung von **WALS** online:

Features/Merkmale (Dryer et al., 2013)

- Alle Features sind in insgesamt 11 Area/Bereiche (Phonologie, Morphologie, Wortstellung ...) eingeteilt
- Unter jedem Bereich sind eine Menge von Features/Merkmalen
- Ein Feature/Merkmal beschreibt eine Eigenschaft unter dem jeweiligen Bereich.
- In den meisten Fällen entspricht einem Feature/Merkmal einen Chapter/Kapitel.

Features

A feature is a structural property of language that describes one aspect of cross-linguistic diversity. A WALS feature has between 2 and 28 different values, shown by different colours on the maps. Most features correspond straightforwardly to chapters, but some chapters are about multiple features.

Showing 1 to 100 of 192 entries

← Previous 1 2 Next → ⓘ

Id	Name	Authors	Area	Languages	Details
<input type="text" value="Search"/>	<input type="text" value="Search"/>		--any--	<input type="text" value="Search"/>	
1A	Consonant Inventories	Ian Maddieson	Phonology	563	Values
2A	Vowel Quality Inventories	Ian Maddieson	Phonology	564	Values
3A	Consonant-Vowel Ratio	Ian Maddieson	Phonology	564	Values

Die Vorstellung von WALS online:

Features/Merkmale (Dryer et al., 2013)

Feature 1A: Consonant Inventories



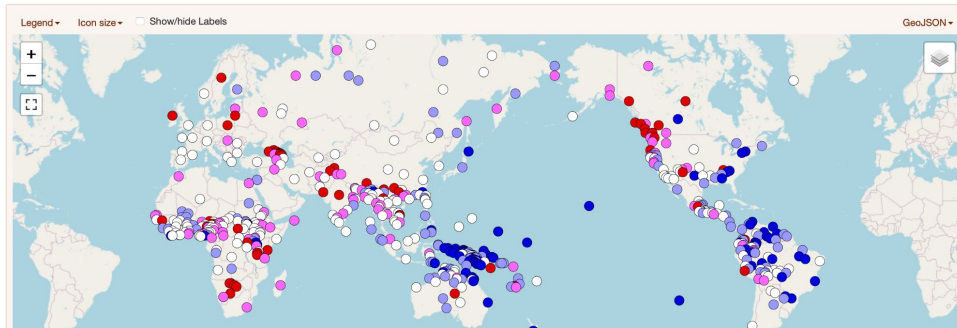
This feature is described in the text of chapter 1 [Consonant Inventories](#) by Ian Maddieson [cite](#)

You may combine this feature with another one. Start typing the feature name or number in the field below.

× 1A: Consonant Inventories

Submit

Values		
 • ▾	Small	89
 • ▾	Moderately small	122
 • ▾	Average	201
 • ▾	Moderately large	94
 • ▾	Large	57
<input type="button" value="reload"/>		



- Zu jedem Feature/Merkmal gibt es zwischen 2 und 28 Werten, dessen Verteilung man auch auf der Weltkarte sehen kann.

Die Vorstellung von **WALS** online: Sprachen (Dryer et al., 2013)

Languages

Search a languoid by name. Matching names are formatted in bold font for **languages**, in italics for *genera* and underlined for families. This search does also take alternative names into account.

Showing 1 to 100 of 2,662 entries

← Previous 1 2 3 4 5 Next →

Name	WALS code	ISO 639-3	Genus	Family	Macroarea	Latitude	Longitude	Countries
<input type="text" value="Search"/>	<input type="text" value="Search"/>	<input type="text" value="Search"/>	<input type="text" value="Search"/>	<input type="text" value="Search"/>	--any--	<input type="text" value="Search"/>	<input type="text" value="Search"/>	<input type="text" value="Search"/>
Aari	aar	a iw	South Omotic	Afro-Asiatic	Africa	6.00	36.58	Ethiopia
Abau	aba	a au	Abau	Sepik	Papunesia	-4.00	141.25	Papua New Guinea
Abaza	abz	ab q	Northwest Caucasian	Northwest Caucasian	Eurasia	44.00	42.00	Russia
Abenakí (Western)	abw	abe	Algonquian	Algic	North America	44.00	-72.25	Canada United States
Abidji	abd	abi	Agneby	Niger-Congo	Africa	5.67	-4.58	Côte d'Ivoire
Abipón	abi	axb	Abipon	Guaicuruan	South America	-29.00	-61.00	Argentina
Abkhaz	abk	abk	Northwest Caucasian	Northwest Caucasian	Eurasia	43.08	41.00	Georgia
Abui	abv	abz	Alor-Pantar	Greater West Bomberai	Papunesia	-8.25	124.67	Indonesia
Abun	abu	kgr	Abun	Abun	Papunesia	-0.50	132.50	Indonesia
Acehnese	ace	ace	Malayo-Sumbawan	Austronesian	Eurasia	5.50	95.50	Indonesia






- Insgesamt 2662 Einträge von Sprachen
- Zu jeder Sprache sind deren Language ID, Sprachfamilie, Breiten- und Längengrad und Länder, wo die Sprache gesprochen wird, verfügbar.

Die Vorstellung von WALS online: Sprachen (Dryer et al., 2013)

Language German

WALS code: ger

Showing 1 to 157 of 157 entries

Fid	Value	Feature	Reference	Area	
	<input type="text" value="Search"/>	<input type="text" value="Search"/>			
1A	Average	Consonant Inventories	Kohler 1999; Moulton 1962; Wangler 1972	Phonology	
2A	Large (7-14)	Vowel Quality Inventories	Kohler 1999; Moulton 1962; Wangler 1972	Phonology	
3A	Low	Consonant-Vowel Ratio	Kohler 1999; Moulton 1962; Wangler 1972	Phonology	
4A	In both plosives and fricatives	Voicing in Plosives and Fricatives	Kohler 1999; Moulton 1962; Wangler 1972	Phonology	
5A	None missing in /p t k b d g/	Voicing and Gaps in Plosive Systems	Kohler 1999; Moulton 1962; Wangler 1972	Phonology	

Nicht jede eingetragene Sprache haben die gleiche Anzahl von Features. Und die Zuverlässigkeit variiert auch wegen fehlenden Referenzen.

Language German (Bavarian)

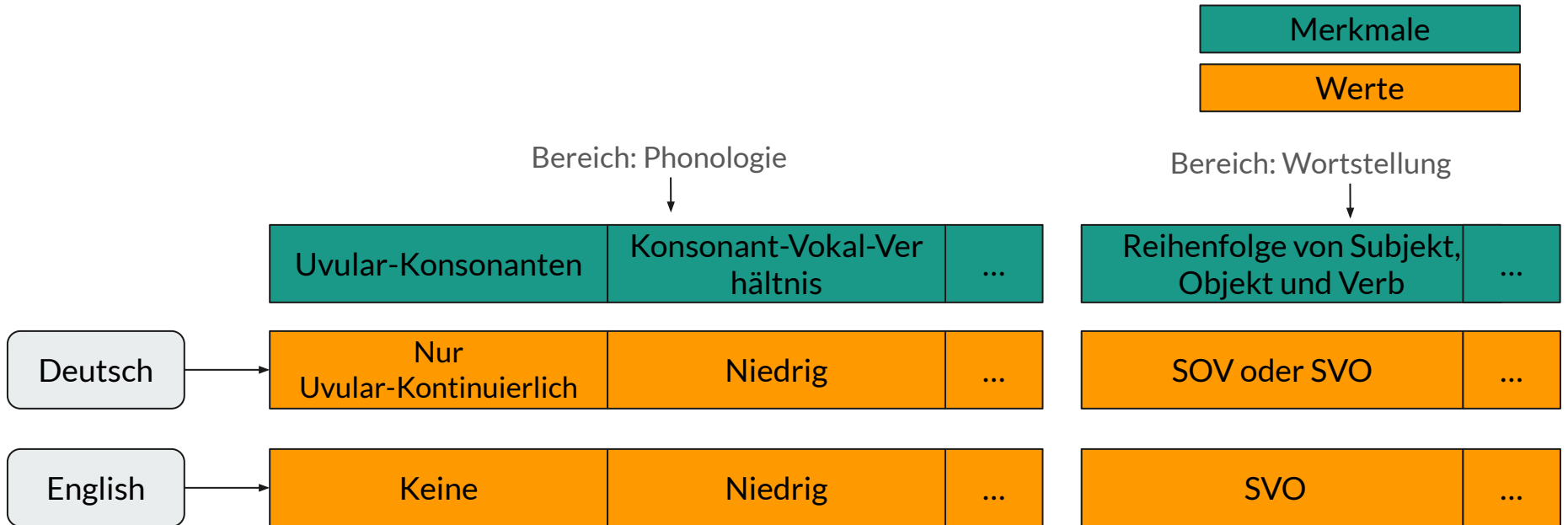
WALS code: gba

Showing 1 to 1 of 1 entries

Fid	Value	Feature	Reference	Area	
	<input type="text" value="Search"/>	<input type="text" value="Search"/>			
54A	No distributive numerals	Distributive Numerals		Nominal Categories	

Showing 1 to 1 of 1 entries

Eine Zwischenzusammenfassung der Struktur der WALS Datenbank





**Fallstudie:
Ein phonologischer Vergleich von Deutsch und
Englisch in WALS**

Exkursion:

Phonetik vs. Phonologie

Phonologie

- Beschäftigung mit dem Lautsystem
- Fokus auf **Phoneme** (bedeutungsunterscheidende Laute)
- Beispiel: /r/ in “Grad” ist für Phonologen nicht bedeutungsunterscheidend, egal ob sie ein **Zungenspitzen-r** ist oder ein **Zäpfchen-r** ist, weil die Bedeutung des Wortes nicht geändert wird. Aber für Phonetiker sind die zwei /r/ zwei Phone
- /r/ und /l/ sind dann zwei Phoneme in dem Wortpaar “Grass” und “Glass”, weil sich die Bedeutungen von den zwei Wörtern wegen einer Laute unterscheiden

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal		m ɱ		n ɳ			ɲ ɳ̺	ŋ	ɴ		
Trill		ʙ		ʀ					ʁ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f ɸ	v θ	ð ʝ	ʒ ʃ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ	ʕ
Lateral fricative				ɬ ɮ							
Approximant			ʋ	ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

Fallstudie: Deutsch vs. Englisch

Phonologische Features/Merkmale in WALIS

- Insgesamt 19 phonologische Features
- Phonetisch?

Id	Name
<input type="text" value="Search"/>	<input type="text" value="Search"/>
1A	Consonant Inventories
2A	Vowel Quality Inventories
3A	Consonant-Vowel Ratio
4A	Voicing in Plosives and Fricatives
5A	Voicing and Gaps in Plosive Systems
6A	Uvular Consonants
7A	Glottalized Consonants
8A	Lateral Consonants
9A	The Velar Nasal
10A	Vowel Nasalization
10B	Nasal Vowels in West Africa
11A	Front Rounded Vowels
12A	Syllable Structure
13A	Tone
14A	Fixed Stress Locations
15A	Weight-Sensitive Stress
16A	Weight Factors in Weight-Sensitive Stress Systems
17A	Rhythm Types
18A	Absence of Common Consonants
19A	Presence of Uncommon Consonants

Fallstudie: Deutsch vs. Englisch

Die Unterscheidungen auf phonologischer Ebene

6A. Uvular Konsonanten

11A. Vorne gerundete Vokale

16A. Weight Factors in Weight-Sensitive Stress Systems/ 16A. Gewichtsfaktoren in gewicht-sensitiven Betonung-Systemen

19A. Presence of Uncommon Consonants/ 19A. Präsenz von unüblichen Konsonanten

Fallstudie: Deutsch vs. Englisch

6A. Uvular Konsonanten (I. Maddieson, 2013)

Mit Zäpfchen artikulierte Konsonanten.

Werten:

1. Keine Uvularen
2. Nur Uvular-Stops
 - Stop: Der Luftstrom wird durch einen Verschluss im Mund kurzzeitig zurückgehalten
3. Nur Uvular-Kontinua
 - Der Luftstrom ist kontinuierlich
4. Uvular-Stops und Uvular-Kontinua

Deutsch: Nur Uvular-Kontinua

Englisch: Keine Uvularen

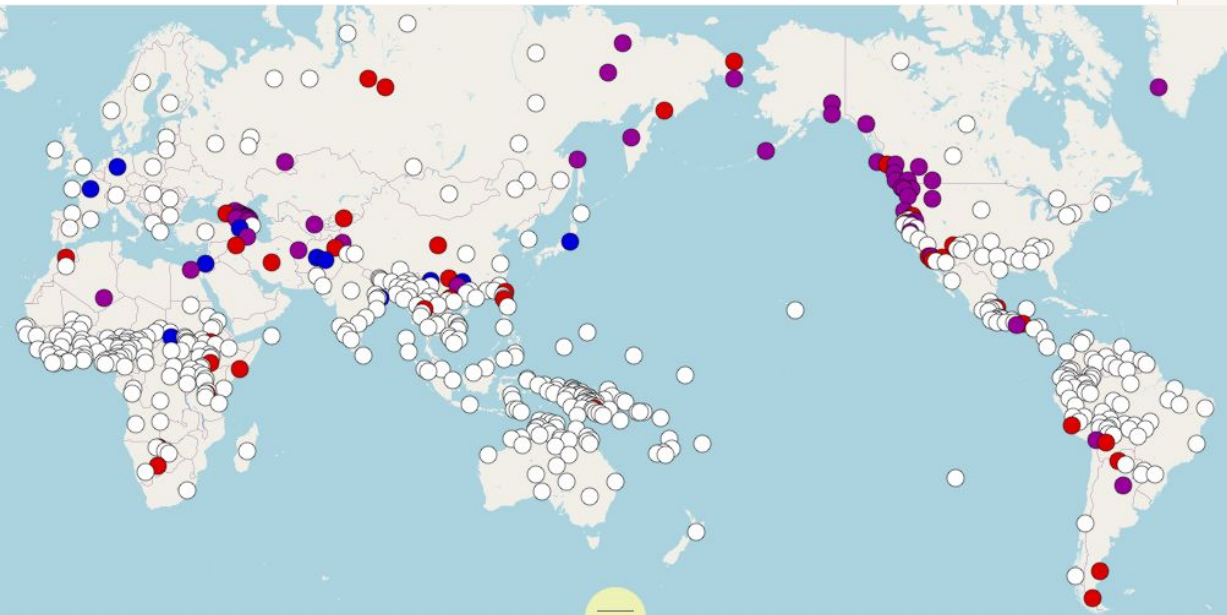
Fallstudie: Deutsch vs. Englisch

6A. Uvular Konsonanten

Die Verteilung auf Weltkarten

Values

<input type="checkbox"/>	<input type="checkbox"/>	None	470
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Uvular stops only	38
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Uvular continuants only	11
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Uvular stops and continuants	48



Fallstudie: Deutsch vs. Englisch

11A. Vorne gerundete Vokale (I. Maddieson, 2013)

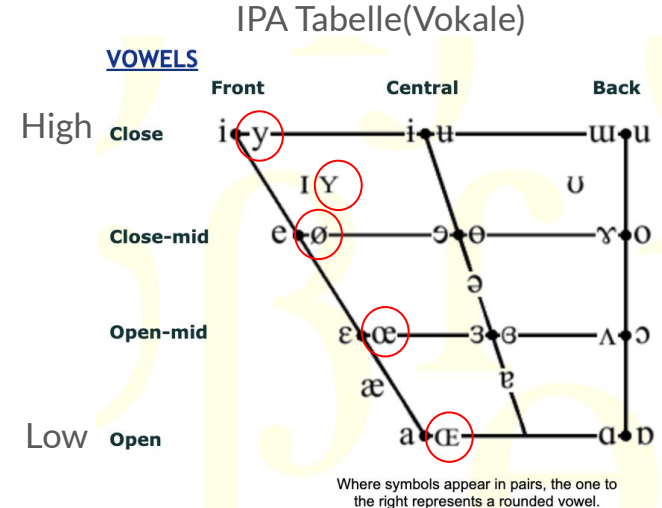
Die Lippenposition von Vokalen ist im Allgemeinen mit der **Höhe** und dem "**Backness**" des Vokals korreliert. Vokale, die hinten im Mund und nicht niedrig, sind mit gerundeter Lippenposition artikuliert.

Werten:

1. Keine
2. Hoch und Mitte
3. Nur hoch
4. Nur mitte

Deutsch: Hoch und Mitte

Englisch: Keine



Fallstudie: Deutsch vs. Englisch

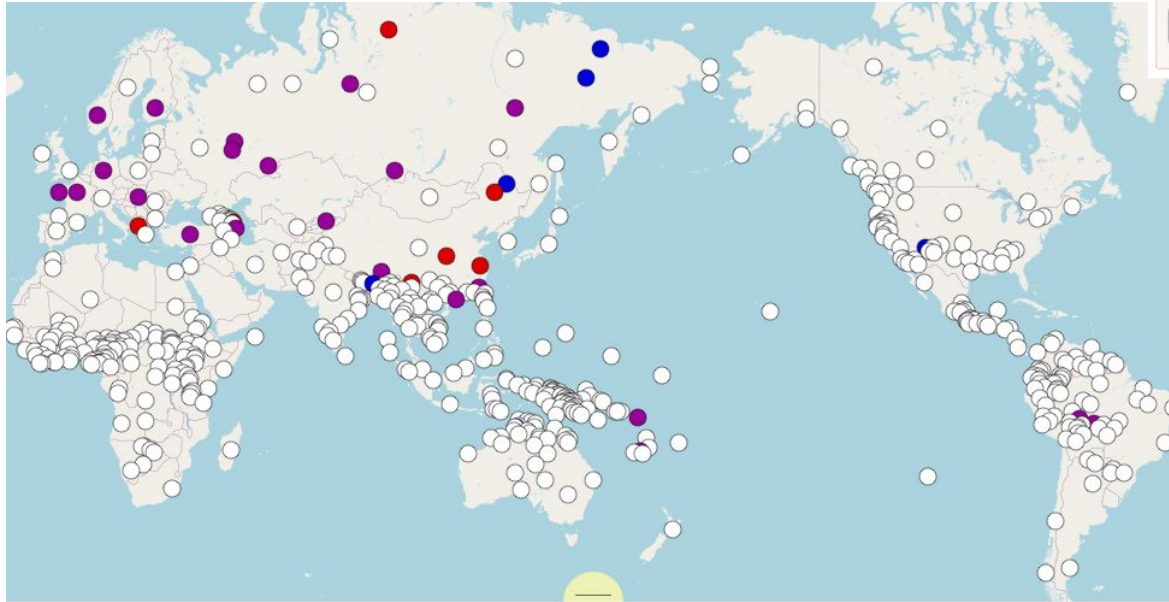
11A. Vorne gerundete Vokale (I. Maddieson, 2013)

Die Verteilung auf Weltkarten

Values

<input type="checkbox"/>	• ▾ ▹	None	525
<input checked="" type="checkbox"/>	• ▾ ▹	High and mid	23
<input checked="" type="checkbox"/>	• ▾ ▹	High only	8
<input checked="" type="checkbox"/>	• ▾ ▹	Mid only	6

reload



Fallstudie: Deutsch vs. Englisch

16A. Gewichtsfaktoren in gewicht-sensitiven Betonung-Systemen

(R. Goedemans, H. van der Hulst, 2013)

- Eine **Silbe** besteht aus **Onset**(dem Teil vor dem Vokal) und **Rhyme** (dem Rest). Das Rhyme besteht aus **Nucleus** (dem Vokal) und **Coda** (dem Teil nach dem Vokal)
- **Silbengewicht:**
 - Wenn die Betonung der Wörter in einer Sprache abhängig von den Eigenschaften der Silben ist, das hat die Sprache ein gewicht-sensitives Betonung-System.
 - Die Silben sind dann in zwei Gruppen eingeteilt. “**heavy/schwer**” und “**light/leicht**” Silben. Eine schwere Silbe wird Betonung “anziehen”, während eine leichte Silbe die Betonung nur bekommen, wenn sie rechts in der Silben-String steht.
- **Gewichtsfaktor:**
 - Faktoren in Silben, die die Silben “heavy/schwer” oder “light/leicht” macht.
 - Wichtige Faktoren sind die Länge des Vokals und der Silbenschluss(syllable closure)

Fallstudie: Deutsch vs. Englisch

16A. Gewichtsfaktoren in gewicht-sensitiven Betonung-Systemen

(R. Goedemans, H. van der Hulst, 2013)

Werte:

- Kein Gewicht, oder Gewichtsfaktor unbekannt
- Langer Vokal: Lange Vokale sind schwer für Betonung
- Coda Konsonant: Silbenschluss ist schwer für Betonung
- Longer Vokal oder Coda: Lange Vokale oder Silbenschluss
- Prominence: andere Faktoren, die schwer für Betonung sind
- Lexikalisch: lexikalische Betonung, (lexical marking) zb. “PER-mit” vs. “per-MIT” auf Englisch
- Kombiniert: zwei der oben erwähnten Faktoren bestimmen Gewicht

Deutsch: Coda Konsonant

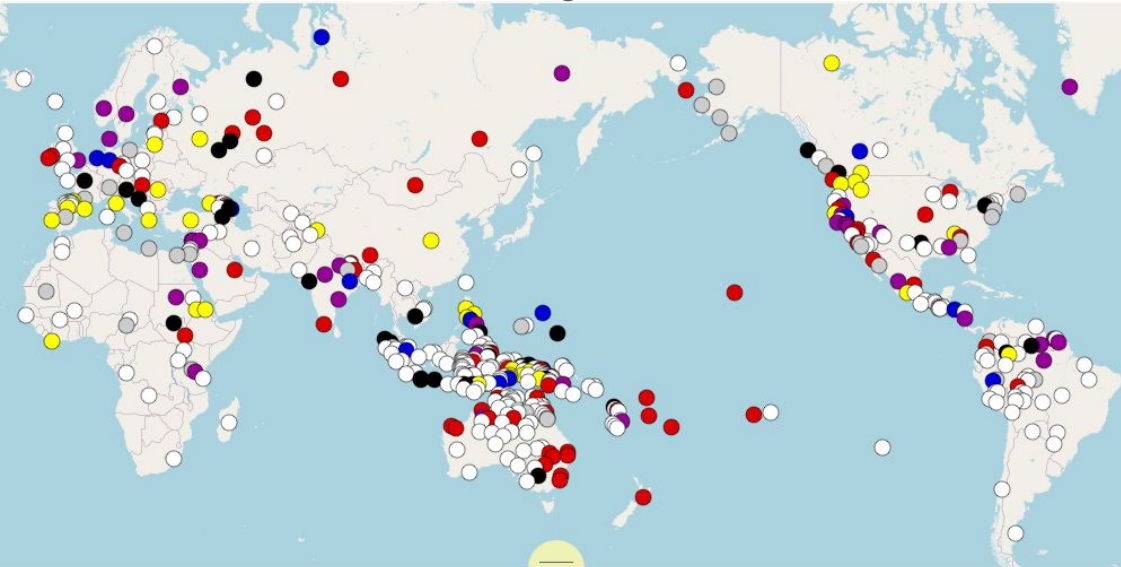
Englisch: Langer Vokal oder Coda

Fallstudie: Deutsch vs. Englisch

16A. Gewichtsfaktoren in gewicht-sensitiven Betonung-Systemen

(R. Goedemans, H. van der Hulst, 2013)

Die Verteilung auf Weltkarten



Values

<input type="checkbox"/>	<input type="text"/>	No weight	261
<input checked="" type="checkbox"/>	<input type="text"/>	Long vowel	65
<input checked="" type="checkbox"/>	<input type="text"/>	Coda consonant	18
<input checked="" type="checkbox"/>	<input type="text"/>	Long vowel or coda consonant	35
<input checked="" type="checkbox"/>	<input type="text"/>	Prominence	41
<input checked="" type="checkbox"/>	<input type="text"/>	Lexical stress	38
<input checked="" type="checkbox"/>	<input type="text"/>	Combined	42

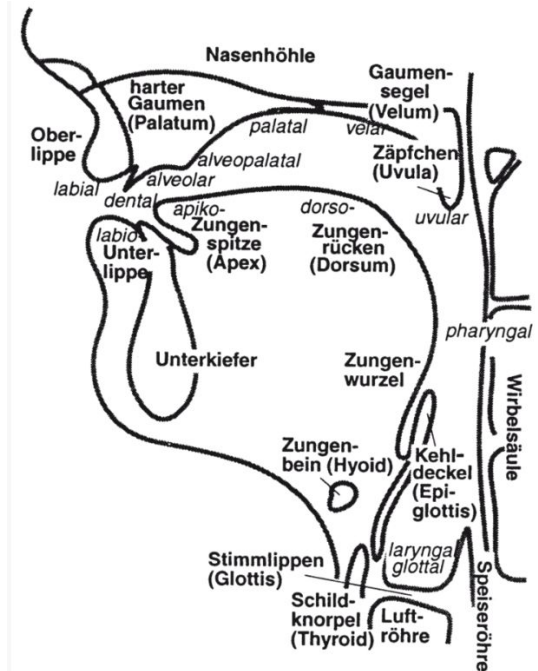
reload

Fallstudie: Deutsch vs. Englisch

19A. Präsenz von unüblichen Konsonanten (I. Maddieson, 2013)

4 Klassen von unüblichen Konsonanten werden genannt:

1. Click
2. Labial-velar Plosive: Verschlüsse bei Lippen und Velar
3. Pharyngeal
4. "Th"



Fallstudie: Deutsch vs. Englisch

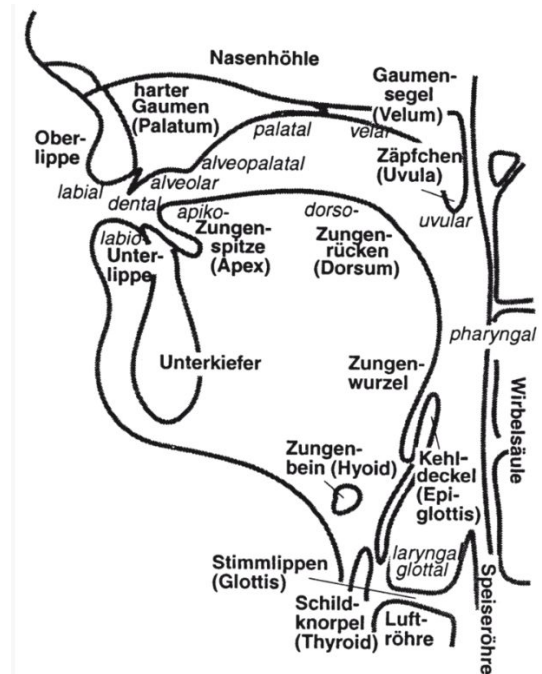
19A. Präsenz von unüblichen Konsonanten (I. Maddieson, 2013)

Werte:

- Keine
- Click
- Labial-Velar
- Pharyngeal
- “Th” Laute
- Click, Pharyngeal und “th” Laute
- Pharyngeal und “th” Laute

Deutsch: Keine

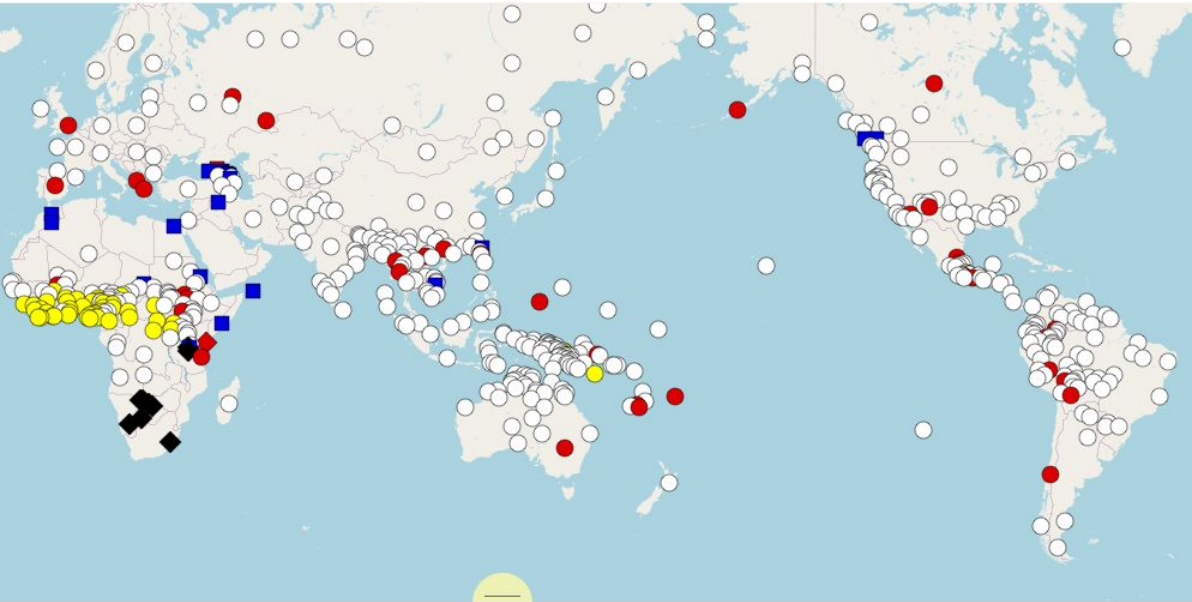
Englisch: “Th” Laute



Fallstudie: Deutsch vs. Englisch

19A. Präsenz von unüblichen Konsonanten (I. Maddieson, 2013)

Die Verteilung auf Weltkarten



Values

<input type="checkbox"/>	• ↕	None	449
<input type="checkbox"/>	↕ ↕	Clicks	9
<input type="checkbox"/>	• ↕	Labial-velars	45
<input type="checkbox"/>	▪ ↕	Pharyngeals	21
<input type="checkbox"/>	• ↕	'Th' sounds	40
<input type="checkbox"/>	↕ ↕	Clicks, pharyngeals, and 'th'	1
<input type="checkbox"/>	▪ ↕	Pharyngeals and "th"	2

reload

Nutzung der WALS Datenbank in Studien

Exploring Linguistic Properties of Monolingual BERTs with Typological Classification among Languages (Ruzzetti et al., 2023)

Motivation:

- Wenn Sprachen in Bezug auf syntaktische oder morphologische Eigenschaften typologisch ähnlich sind, sollten ihre BERT-Modelle in einigen Layers ähnliche Gewichtsmatrizen aufweisen. Damit kann man wissen welcher Layer zb. syntaktische Informationen speichert.

Hypothese:

- Gegeben eine Mengen von Sprachenpaaren (L1,L2)
- Ein Matrix W von einem BERT layer kodiert dann die syntaktische oder morphologische Informationen, wenn die Ähnlichkeit(W_{I1}, W_{I2}) von MonoBert_I1 und MonoBert_I2 mit der **syntaktischen Ähnlichkeit (L1,L2)** oder der **morphologischen Ähnlichkeit (L1,L2)** korreliert.

Schlussfolgerung:

- Mit einer neuen Methode zeigt man, dass die Layers 4,5 und 6 syntaktische linguistischen Informationen kodieren.

Nutzung der WALS Datenbank in Studien

Exploring Linguistic Properties of Monolingual BERTs with Typological Classification among Languages (Ruzzetti et al., 2023)

Rolle der WALS Datenbank:

- Erhalten von Vektoren für Sprachen auf gewünschten linguistischen Ebene

Bereich: Phonologie



Uvular-Konsonanten	Konsonant-Vokal-Verhältnis	...
--------------------	----------------------------	-----

Deutsch	Nur Uvular-Kontinuierlich	Niedrig	...
---------	---------------------------	---------	-----

English	Keine	Niedrig	...
---------	-------	---------	-----

Merkmale

Werte

Bereich: Wortstellung



Reihenfolge von Subjekt, Objekt und Verb	...
--	-----

SOV oder SVO	...
--------------	-----

SVO	...
-----	-----

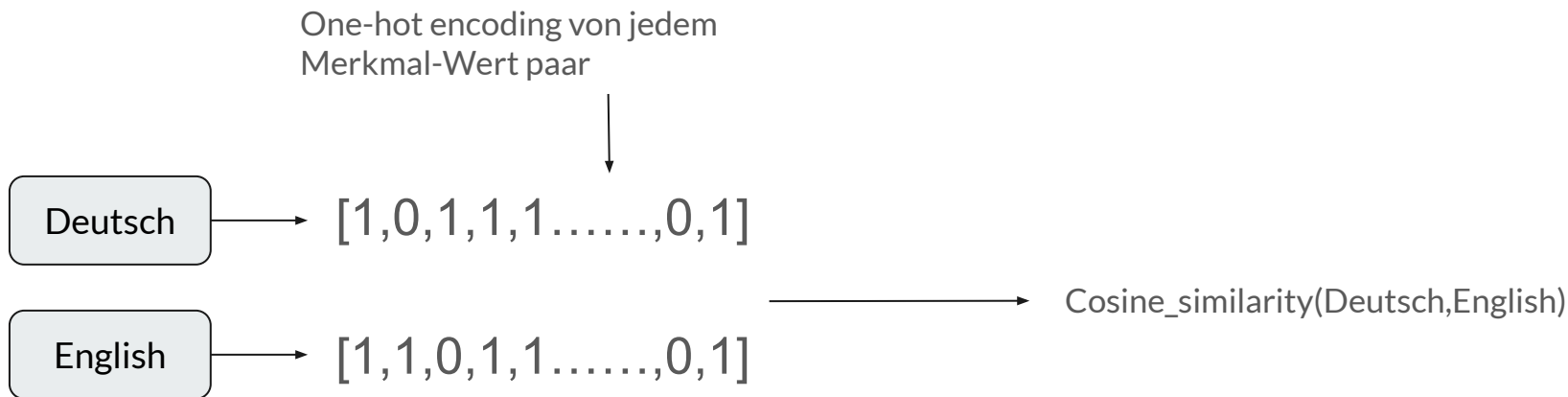
(Dryer et al., 2013)

Nutzung der WALS Datenbank in Studien

Exploring Linguistic Properties of Monolingual BERTs with Typological Classification among Languages (Ruzzetti et al., 2023)

Rolle der WALS Datenbank:

- Erhalten von Vektoren für Sprachen auf gewünschten linguistischen Ebene



Fazit

Vorteile:

1. Ermöglicht cross-linguale Vergleiche von Sprachen
2. Bietet Möglichkeiten zur Vektorisierung von Sprachen auf gewünschten linguistischen Ebene
3. Die meisten Einträge sind durch veröffentlichten Literaturen verifiziert (zitierbar) und daher zuverlässig
4. bietet Breiten- und Längengrad zu jeder Sprache an, was interessante Visualisierungen zulässt.

Nachteile:

1. Low-resource oder international wenig geforschte Sprachen haben deutlich weniger Einträge als die gut dokumentiert gefortschte Sprachen.
2. Kritik?: Phonologische Merkmale sind eher phonetische?



Literatur

- Dryer, Matthew S. & Haspelmath, Martin (eds.) 2013. *WALS Online (v2020.3)* [Data set].
- <https://neon.niederlandistik.fu-berlin.de/de/nedling/phonology/phoneticsphonology> (Accessed on 2024-01-23.)
- <https://web.uvic.ca/ling/resources/ipa/charts/IPAlab/IPAlab.htm> (Accessed on 2024-01-23.)
- <https://www.phonetik.uni-muenchen.de/studium/skripten/SGL/SGLKap2.html>(Accessed on 2024-01-23.)
- <https://dlf.uzh.ch/sites/dialektressource/linguistik/phonologie/#1665427795980-5ad58246-a1e5>(Accessed on 2024-01-23.)
- Elena Ruzzetti, Federico Ranaldi, Felicia Logozzo, Michele Mastromattei, Leonardo Ranaldi, and Fabio Zanzotto. 2023. Exploring Linguistic Properties of Monolingual BERTs with Typological Classification among Languages. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 14447–14461, Singapore. Association for Computational Linguistics.

Fragen?

