

Eine neue Methode zur automatischen Identifikation etymologisch verwandter Wörter

Johann-Mattis List*

* Institut für Romanistik II
Heinrich Heine Universität Düsseldorf

2011/07/01

Gliederung des Vortrags

Kognatenidentifikation in der historischen Linguistik

- Die Gegenwart als Schlüssel zur Vergangenheit
- Beziehungen zwischen sprachlichen Zeichen
- Rekonstruktion von Zeichenbeziehungen
- Probleme

Vorüberlegungen für eine automatische Implementierung

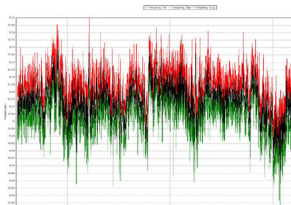
- Alinierung
- Ähnlichkeit
- Lautklassen

Die neue Methode zur automatischen Kognatenidentifikation

- Arbeitsweise
- Arbeitsschritte
- Implementierung

Evaluierung der Methode

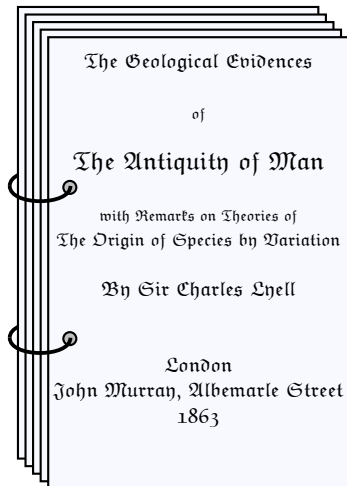
- Kontrolldatensätze
- Ergebnisse



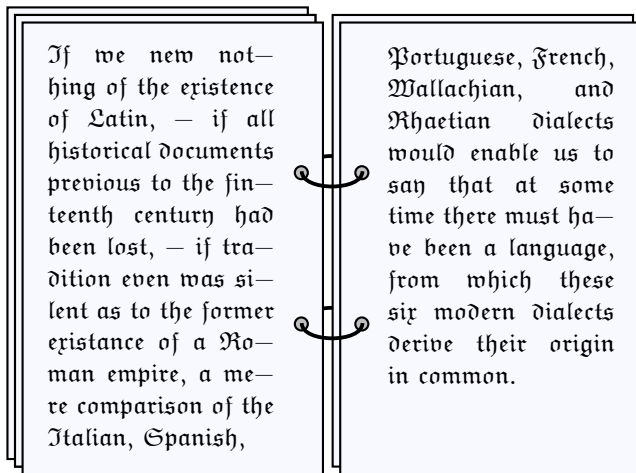
Kognatenidentifikation



Von Schichten zu Geschichten



Von Schichten zu Geschichten



Uniformitarianismus

Uniformitarianismus

Uniformität des Wandels: Es wird davon ausgegangen, dass die Gesetze des Wandels *uniform* sind, d. h. dass sie genauso in der Vergangenheit galten, wie sie in der Gegenwart gelten und in der Zukunft gelten werden.

Uniformitarianismus

- ▶ **Uniformität des Wandels:** Es wird davon ausgegangen, dass die Gesetze des Wandels *uniform* sind, d. h. dass sie genauso in der Vergangenheit galten, wie sie in der Gegenwart gelten und in der Zukunft gelten werden.
- ▶ **Gradualität des Wandels:** Es wird davon ausgegangen, dass Wandel graduell vor sich geht.

Uniformitarianismus

- ▶ **Uniformität des Wandels:** Es wird davon ausgegangen, dass die Gesetze des Wandels *uniform* sind, d. h. dass sie genauso in der Vergangenheit galten, wie sie in der Gegenwart gelten und in der Zukunft gelten werden.
- ▶ **Gradualität des Wandels:** Es wird davon ausgegangen, dass Wandel graduell vor sich geht.
- ▶ **Indiziengestützte Beweisführung:** Ausgehend von in der Gegenwart gegebenen Tatsachen wird, vor dem Hintergrund der Annahme graduellen Wandels, auf Tatsachen in der Vergangenheit geschlossen.

Das sprachliche Zeichen

Das sprachliche Zeichen

- ▶ **Form:** Die *materielle* Erscheinungsform des sprachlichen Zeichens (eine Lautkette).

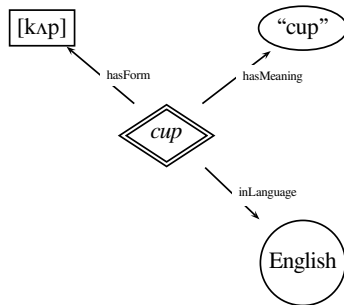
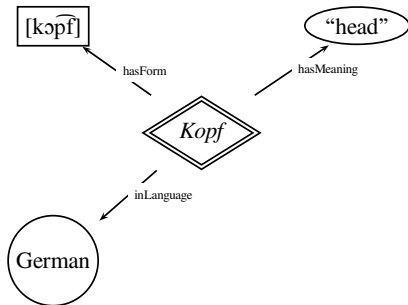
Das sprachliche Zeichen

- ▶ **Form:** Die *materielle* Erscheinungsform des sprachlichen Zeichens (eine Lautkette).
- ▶ **Funktion:** Die Bedeutung des sprachlichen Zeichens.

Das sprachliche Zeichen

- ▶ **Form:** Die *materielle* Erscheinungsform des sprachlichen Zeichens (eine Lautkette).
- ▶ **Funktion:** Die Bedeutung des sprachlichen Zeichens.
- ▶ **System:** Die Sprache, in der das sprachliche Zeichen durch seine Form mit einer Bedeutung verknüpft wird.

Das sprachliche Zeichen



Beziehungen zwischen sprachlichen Zeichen

Beziehungen zwischen sprachlichen Zeichen

- ▶ **Etymologische Beziehung:** Sprachliche Zeichen teilen eine gemeinsame Geschichte.

Beziehungen zwischen sprachlichen Zeichen

- ▶ **Etymologische Beziehung:** Sprachliche Zeichen teilen eine gemeinsame Geschichte.
- ▶ **Vorgänger-Nachfolger-Beziehung:** Ein Zeichen ist aus einem anderen Zeichen durch einen *graduellen* Wandelprozess hervorgegangen.

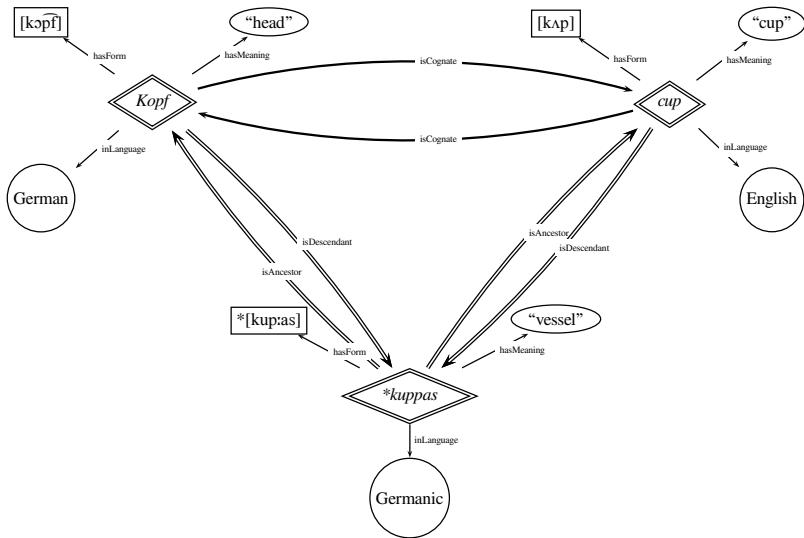
Beziehungen zwischen sprachlichen Zeichen

- ▶ **Etymologische Beziehung:** Sprachliche Zeichen teilen eine gemeinsame Geschichte.
- ▶ **Vorgänger-Nachfolger-Beziehung:** Ein Zeichen ist aus einem anderen Zeichen durch einen *graduellen* Wandelprozess hervorgegangen.
- ▶ **Donor-Rezipient-Beziehung:** Ein Zeichen ist aus einem anderen Zeichen durch einen *diskreten* Prozess der Übertragung (aus einer anderen Sprache) hervorgegangen.

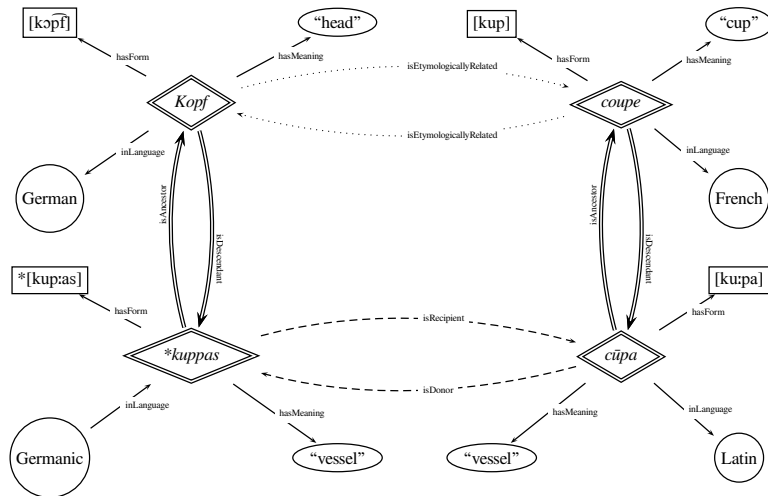
Beziehungen zwischen sprachlichen Zeichen

- ▶ **Etymologische Beziehung:** Sprachliche Zeichen teilen eine gemeinsame Geschichte.
- ▶ **Vorgänger-Nachfolger-Beziehung:** Ein Zeichen ist aus einem anderen Zeichen durch einen *graduellen* Wandelprozess hervorgegangen.
- ▶ **Donor-Rezipient-Beziehung:** Ein Zeichen ist aus einem anderen Zeichen durch einen *diskreten* Prozess der Übertragung (aus einer anderen Sprache) hervorgegangen.
- ▶ **Kognatenbeziehung:** Zwei Zeichen sind Nachfolger desselben Zeichens.

Beziehungen zwischen sprachlichen Zeichen



Beziehungen zwischen sprachlichen Zeichen



Identifizieren historisch bedingter Ähnlichkeiten

Identifizieren historisch bedingter Ähnlichkeiten

- ▶ **zufällige Ähnlichkeiten:** Wörter klingen zufällig gleich oder ähnlich.

Identifizieren historisch bedingter Ähnlichkeiten

- ▶ **zufällige Ähnlichkeiten:** Wörter klingen zufällig gleich oder ähnlich.
- ▶ **natürliche Ähnlichkeiten:** Wörter klingen gleich, weil sie universellen Denotationsmustern folgen.

Identifizieren historisch bedingter Ähnlichkeiten

- ▶ **zufällige Ähnlichkeiten:** Wörter klingen zufällig gleich oder ähnlich.
- ▶ **natürliche Ähnlichkeiten:** Wörter klingen gleich, weil sie universellen Denotationsmustern folgen.
- ▶ **historisch bedingte Ähnlichkeiten:**

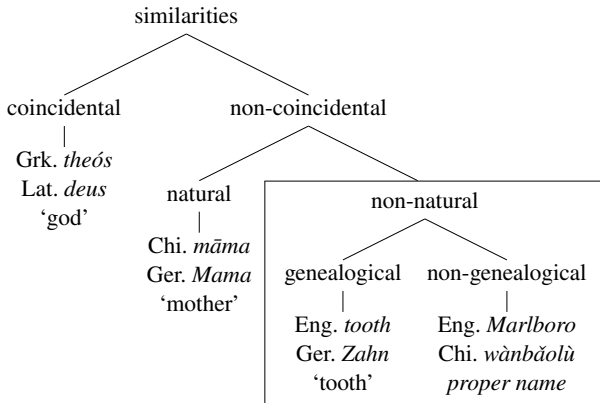
Identifizieren historisch bedingter Ähnlichkeiten

- ▶ **zufällige Ähnlichkeiten:** Wörter klingen zufällig gleich oder ähnlich.
- ▶ **natürliche Ähnlichkeiten:** Wörter klingen gleich, weil sie universellen Denotationsmustern folgen.
- ▶ **historisch bedingte Ähnlichkeiten:**
 - ▶ **genealogische Ähnlichkeiten:** Wörter klingen ähnlich, weil sie aus einer gemeinsamen Vorgängerform entstanden sind.

Identifizieren historisch bedingter Ähnlichkeiten

- ▶ **zufällige Ähnlichkeiten:** Wörter klingen zufällig gleich oder ähnlich.
- ▶ **natürliche Ähnlichkeiten:** Wörter klingen gleich, weil sie universellen Denotationsmustern folgen.
- ▶ **historisch bedingte Ähnlichkeiten:**
 - ▶ **genealogische Ähnlichkeiten:** Wörter klingen ähnlich, weil sie aus einer gemeinsamen Vorgängerform entstanden sind.
 - ▶ **nicht-genealogische Ähnlichkeiten:** Wörter klingen ähnlich, weil das eine Wort durch einen Entlehnungsprozess aus dem anderen hervorgegangen ist.

Identifizieren historisch bedingter Ähnlichkeiten



Kumulative Evidenz

The force of our evidence is cumulative; while it might be possible to doubt the validity of each item taken separately, the inference from all the items combined is in many cases practically certain.

Sturtevant (1940: 29)

Linguistische Rekonstruktion

German

ts^h

a:

n

English

t

ʊ:

θ

Italian

d

ε

n

t

e

French

d

ã

Linguistische Rekonstruktion

*German*ts^h

a:

n

English

t

ʊ:

θ

Italian

d

ɛ

n

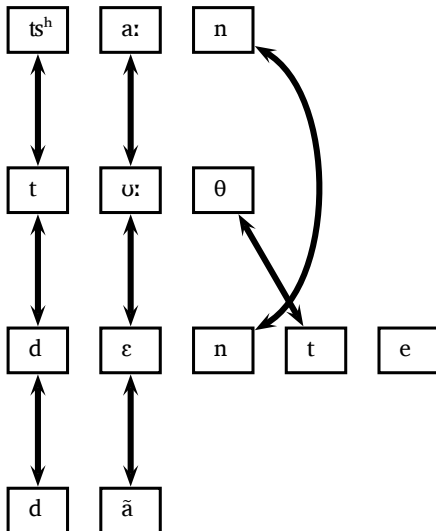
t

e

French

d

ã



Linguistische Rekonstruktion

German

ts ^h	a:	n	-	-
-----------------	----	---	---	---

English

t	ʊ:	-	θ	-
---	----	---	---	---

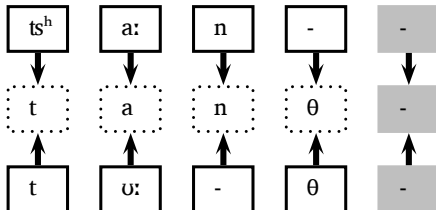
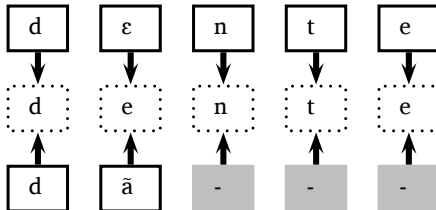
Italian

d	ε	n	t	e
---	---	---	---	---

French

d	ã	-	-	-
---	---	---	---	---

Linguistische Rekonstruktion

German*Proto-Germanic**English**Italian**Proto-Romance**French*

Linguistische Rekonstruktion

Proto-Germanic

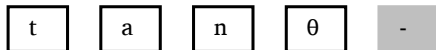
t	a	n	θ	-
---	---	---	---	---

Proto-Romance

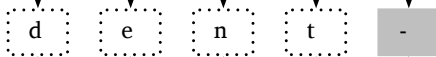
d	e	n	t	e
---	---	---	---	---

Linguistische Rekonstruktion

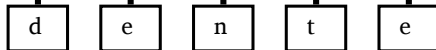
Proto-Germanic



Proto-Indo-European



Proto-Romance



Linguistische Rekonstruktion

Proto-Indo-European

d

e

n

t

Linguistische Rekonstruktion

German

Proto-Germanic

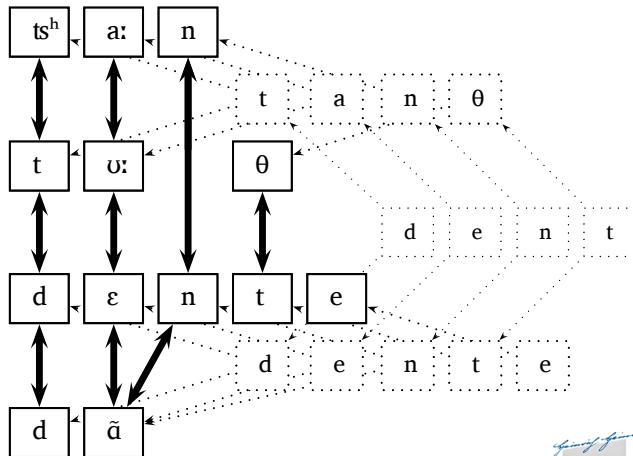
English

Proto-Indo-European

Italian

Proto-Romance

French



Intuition und Objektivität

Part of the process of “becoming” a competent Indo-Europeanist has always been recognized as coming to grasp “intuitively” concepts and types of changes in language so as to be able to pick and choose between alternative explanations for the history and development of specific features of the reconstructed language and its offspring.

Schwink (1994: 29)

Zirkularität und iteratives Vorgehen

Zirkularität und iteratives Vorgehen

- ▶ **Kognatenbeziehungen** können nur auf der Grundlage nachgewiesener **systematischer Korrespondenzen** nachgewiesen werden.

Zirkularität und iteratives Vorgehen

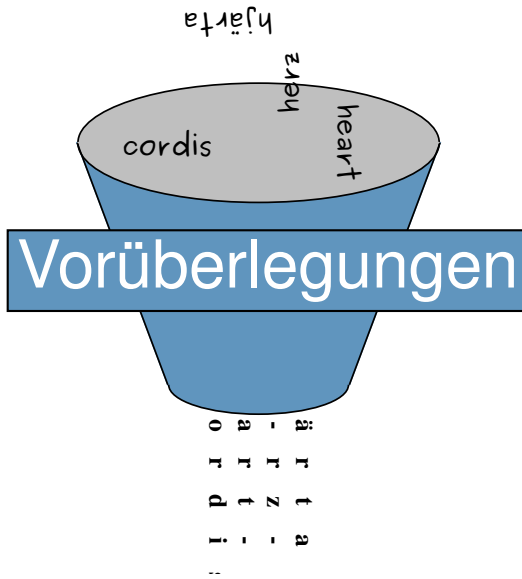
- ▶ **Kognatenbeziehungen** können nur auf der Grundlage nachgewiesener **systematischer Korrespondenzen** nachgewiesen werden.
- ▶ **Systematische Korrespondenzen** können nur auf der Grundlage nachgewiesener **Kognatenbeziehungen** ermittelt werden.

Zirkularität und iteratives Vorgehen

- ▶ **Kognatenbeziehungen** können nur auf der Grundlage nachgewiesener **systematischer Korrespondenzen** nachgewiesen werden.

- ▶ **Systematische Korrespondenzen** können nur auf der Grundlage nachgewiesener **Kognatenbeziehungen** ermittelt werden.

→ Um das Problem der Zirkularität zu umgehen, wird allgemein eine iterative Heuristik verwendet, mit deren Hilfe erste Hypothesen aufgestellt und sukzessive bestärkt oder verworfen werden.



Alinierung

Alinierung

- ▶ Die Alinierung stellt die am weitesten verbreitete Methode für Sequenzvergleiche dar.

Alinierung

- ▶ Die Alinierung stellt die am weitesten verbreitete Methode für Sequenzvergleiche dar.
In Alinierungsanalysen werden zwei oder mehrere Sequenzen dergestalt in einer Matrix angeordnet, dass
- ▶ miteinander korrespondierende Segmente in derselben Spalte erscheinen, wobei Nullkorrespondenzen mit Hilfe von Lückensymbolen dargestellt werden.

Alinierung

- ▶ Die Alinierung stellt die am weitesten verbreitete Methode für Sequenzvergleiche dar.

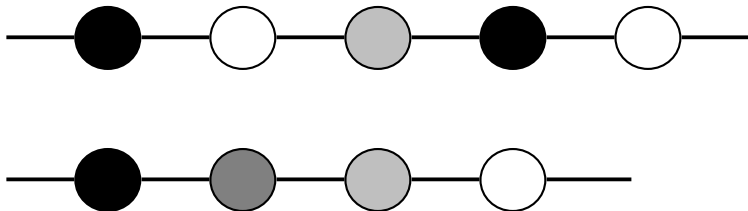
In Alinierungsanalysen werden zwei oder mehrere Sequenzen dergestalt in einer Matrix angeordnet, dass

- ▶ miteinander korrespondierende Segmente in derselben Spalte erscheinen, wobei Nullkorrespondenzen mit Hilfe von Lückensymbolen dargestellt werden.

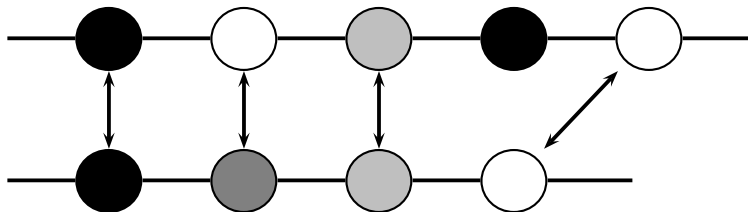
Alinierungsanalysen sind grundsätzlich indifferent in Bezug auf die Natur der Korrespondenzen, die mit

- ▶ ihrer Hilfe dargestellt oder ermittelt werden. Eine explizite historische Interpretation der Fakten muss immer in einem zweiten Schritt vorgenommen werden.

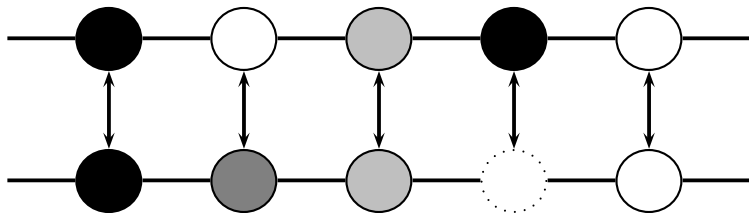
Alinierung



Alinierung



Alinierung



Phänotypische und genotypische Ähnlichkeit

Phänotypische und genotypische Ähnlichkeit

- ▶ **Phänotypische Ähnlichkeit:** Wörter werden als ähnlich angesehen, wenn sie ähnlich klingen (*phenotypic resemblance*, Lass 1997).

Phänotypische und genotypische Ähnlichkeit

- ▶ **Phänotypische Ähnlichkeit:** Wörter werden als ähnlich angesehen, wenn sie ähnlich klingen (*phenotypic resemblance*, Lass 1997).
- ▶ **Genotypische Ähnlichkeit:** Wörter werden als ähnlich angesehen, wenn gezeigt werden kann, dass systematische Korrespondenzbeziehungen zwischen ihren Segmenten bestehen (*genotypic resemblance*, Lass 1997).

Phänotypische und genotypische Ähnlichkeit

► Phänotypische Ähnlichkeit

Greek	mati	‘eye’	≈	Malay	mata	‘eye’
Greek	θεος	‘god’	≈	Spanish	dios	‘god’

Phänotypische und genotypische Ähnlichkeit

► Phänotypische Ähnlichkeit

Greek	mati	‘eye’	≈	Malay	mata	‘eye’
Greek	θεος	‘god’	≈	Spanish	dios	‘god’

► Genotypische Ähnlichkeit

German	ts ^h a:n	‘tooth’	≈	English	tu:θ	‘tooth’
Spanish	etfo	‘fact’	≈	French	fɛ	‘fact’

Das Problem mit den Ähnlichkeiten

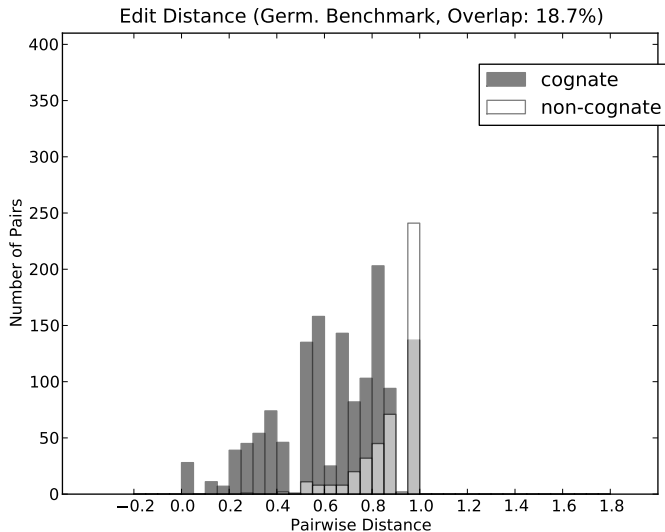
Das Problem mit den Ähnlichkeiten

- ▶ Phänotypische Ähnlichkeit allein ist unbrauchbar als Ähnlichkeits- oder Distanzmaß für automatische Applikationen.

Das Problem mit den Ähnlichkeiten

- ▶ Phänotypische Ähnlichkeit allein ist unbrauchbar als Ähnlichkeits- oder Distanzmaß für automatische Applikationen.
- Genotypische Ähnlichkeit allein beruht auf einem bereits vorgenommenen, im Laufe jahrelanger manueller
- ▶ Forschung betriebenen Vergleich der Sprachen einer Sprachfamilie. Sie taugt nicht als Heuristik für automatische Ansätze.

Das Problem mit den Ähnlichkeiten



Lautklassen

Lautklassen

- ▶ Phänotypische und genotypische Ähnlichkeit stellen zwei gegensätzliche Positionen dar, von denen keine für automatische Applikationen geeignet ist.

Lautklassen

- ▶ Phänotypische und genotypische Ähnlichkeit stellen zwei gegensätzliche Positionen dar, von denen keine für automatische Applikationen geeignet ist.

Um bisher unerforschte Sprachen automatisch zu analysieren, bedarf es eines Ähnlichkeitsmaßes, welches einen Kompromiss zwischen den beiden Ähnlichkeitsmodellen darstellt, also

- ▶ basierend auf der phänotypischen Ähnlichkeit von Sequenzen auf die Wahrscheinlichkeit genotypischer Ähnlichkeiten schließen lässt.

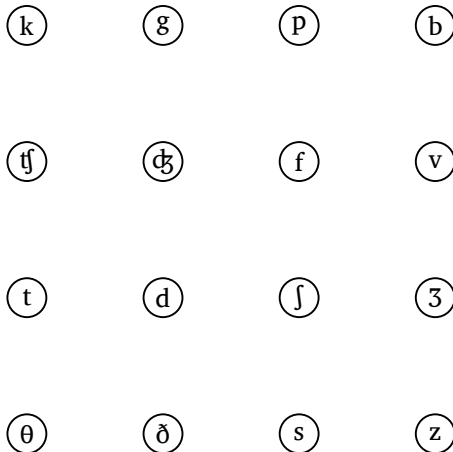
Lautklassen

- ▶ Phänotypische und genotypische Ähnlichkeit stellen zwei gegensätzliche Positionen dar, von denen keine für automatische Applikationen geeignet ist.

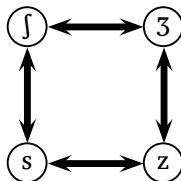
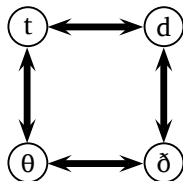
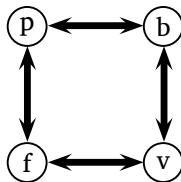
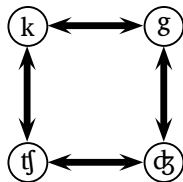
Um bisher unerforschte Sprachen automatisch zu analysieren, bedarf es eines Ähnlichkeitsmaßes, welches einen Kompromiss zwischen den beiden Ähnlichkeitsmodellen darstellt, also basierend auf der phänotypischen Ähnlichkeit von Sequenzen auf die Wahrscheinlichkeit genotypischer Ähnlichkeiten schließen lässt.

- ▶ Lautklassenbasierte Ansätze (vgl. bes. Dolgopolsky 1986) gruppieren sprachliche Laute basierend auf empirischen Studien in Korrespondenzklassen, wobei angenommen wird, dass Laute, die derselben Klasse angehören, signifikant häufiger in Korrespondenzbeziehungen in genetisch verwandten Sprachen auftauchen.

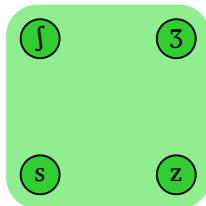
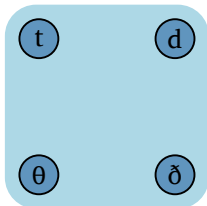
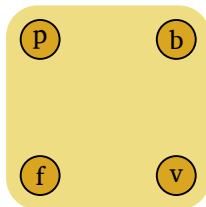
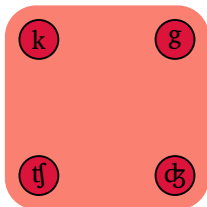
Lautklassen



Lautklassen



Lautklassen



Lautklassen

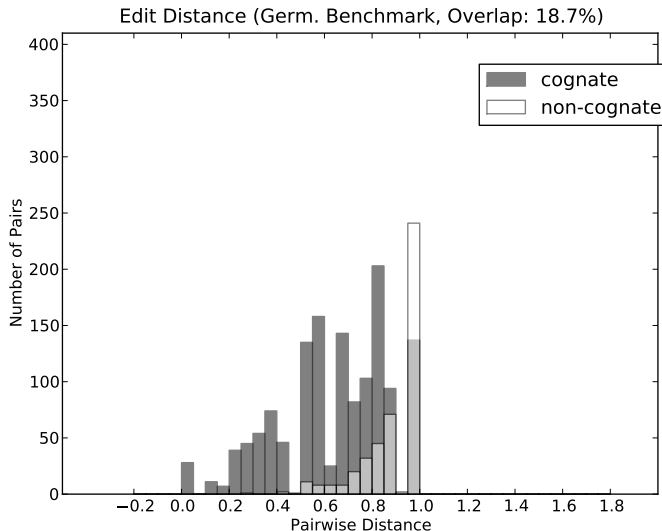
K

P

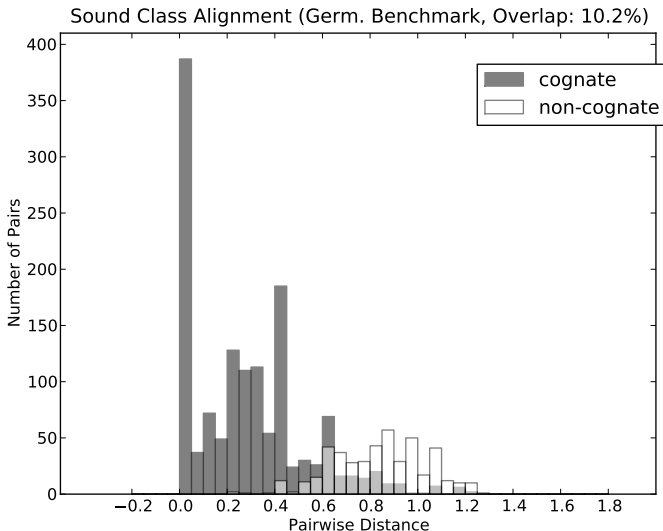
T

S

Lautklassenalinierung vs. phänotypische Alinierung



Lautklassenalinierung vs. phänotypische Alinierung





Die neue Methode

Arbeitsweise

Arbeitsweise

- ▶ **Eingabe:** Bedeutungslisten (Swadeshlisten) für eine beliebige Anzahl von Sprachen und von beliebiger Länge.

Arbeitsweise

- ▶ **Eingabe:** Bedeutungslisten (Swadeshlisten) für eine beliebige Anzahl von Sprachen und von beliebiger Länge.
- ▶ **Eingabeformat:** Generell IPA, für Tonsprachen wird das in der chinesischen Linguistik übliche Format aus ein bis drei Ziffern verwendet.

Arbeitsweise

- ▶ **Eingabe:** Bedeutungslisten (Swadeshlisten) für eine beliebige Anzahl von Sprachen und von beliebiger Länge.
- ▶ **Eingabeformat:** Generell IPA, für Tonsprachen wird das in der chinesischen Linguistik übliche Format aus ein bis drei Ziffern verwendet.
- ▶ **Analyse:** Der Algorithmus analysiert alle Wortlisten, ermittelt mögliche Lautkorrespondenzen und ermittelt, basierend auf Clusteranalysen mögliche Kognatensets.

Arbeitsweise

- ▶ **Eingabe:** Bedeutungslisten (Swadeshlisten) für eine beliebige Anzahl von Sprachen und von beliebiger Länge.
- ▶ **Eingabeformat:** Generell IPA, für Tonsprachen wird das in der chinesischen Linguistik übliche Format aus ein bis drei Ziffern verwendet.
- ▶ **Analyse:** Der Algorithmus analysiert alle Wortlisten, ermittelt mögliche Lautkorrespondenzen und ermittelt, basierend auf Clusteranalysen mögliche Kognatensets.
- ▶ **Ausgabe:** Eine Liste der Eingabedaten mitsamt Kognazitätsurteilen, wobei die Kognatensets in alinierter Form ausgegeben werden, um spätere manuelle Analysen zu erleichtern.

Arbeitsweise

Number	Words	ger	eng	dan	...
1	all	al	ɔ:l	æ [?] l	...
2	ashes	aʃə	æʃ	asg	...
3	bark	rində	bɑ:rk	ba:g	...
4	belly	baux	bɛli	ɔnəbliw [?]	...
4	belly	–	–	mæ:və	...
5	big	gro:s	big	sdo [?] ʁ	...
5	big	–	greit	–	...

Arbeitsweise

Basic Concept: belly						
CogID	Language	Gloss	GlossID	Entry	Aligned Entry	
5	ger	belly	4	baux	b	au x
5	swe	belly	4	buk	b	u k
5	nld	belly	4	bœyk	b	œy k
5	nor	belly	4	bæ:k	b	æ: k
6	eng	belly	4	bɛli	b	ɛ l i
7	dan	belly	4	ɔnæbliw [?]	ɔ	n ə ɾ l i w [?]
8	dan	belly	4	mæ:və	m	æ: v ə
8	swe	belly	4	ma:ge	m	a: g e
8	nor	belly	4	mɑ:gə	m	ɑ: g ə

Arbeitsschritte

Arbeitsschritte

1. Paarweise, lautklassenbasierte Alinierung zur Ermittlung möglicher Korrespondenzpaare (List 2010).

Arbeitsschritte

1. Paarweise, lautklassenbasierte Alinierung zur Ermittlung möglicher Korrespondenzpaare (List 2010).
2. Erstellung einer korrespondenzbasierten Bewertungsmatrix auf Grundlage des Vergleichs randomisierter mit nicht-randomisierten Daten (folgt in Grundzügen Henikoff & Henikoff 1992).

Arbeitsschritte

1. Paarweise, lautklassenbasierte Alinierung zur Ermittlung möglicher Korrespondenzpaare (List 2010).
2. Erstellung einer korrespondenzbasierten Bewertungsmatrix auf Grundlage des Vergleichs randomisierter mit nicht-randomisierten Daten (folgt in Grundzügen Henikoff & Henikoff 1992).
3. Berechnung korrespondenzbasierter paarweiser Distanzen.

Arbeitsschritte

1. Paarweise, lautklassenbasierte Alinierung zur Ermittlung möglicher Korrespondenzpaare (List 2010).
2. Erstellung einer korrespondenzbasierten Bewertungsmatrix auf Grundlage des Vergleichs randomisierter mit nicht-randomisierten Daten (folgt in Grundzügen Henikoff & Henikoff 1992).
3. Berechnung korrespondenzbasierter paarweiser Distanzen.
4. Gruppierung der Wörter zu Kognatensets basierend auf einer Clusteranalyse.

Arbeitsschritte

1. Paarweise, lautklassenbasierte Alinierung zur Ermittlung möglicher Korrespondenzpaare (List 2010).
2. Erstellung einer korrespondenzbasierten Bewertungsmatrix auf Grundlage des Vergleichs randomisierter mit nicht-randomisierten Daten (folgt in Grundzügen Henikoff & Henikoff 1992).
3. Berechnung korrespondenzbasierter paarweiser Distanzen.
4. Gruppierung der Wörter zu Kognatensets basierend auf einer Clusteranalyse.
5. Multiple Alinierung der ermittelten Kognatensets (List 2011), um die manuelle Kontrolle der Ergebnisse zu erleichtern.

LingPy

- ▶ LingPy (List 2011) ist eine frei verfügbare Pythonbibliothek, welche Klassen und Funktionen für verschiedenste Aufgaben in quantitativen Ansätzen in der historischen Linguistik zur Verfügung stellt.

* *

v o l - d e m o r t

v - l a d i m i r -

v a l - d e m a r -

*

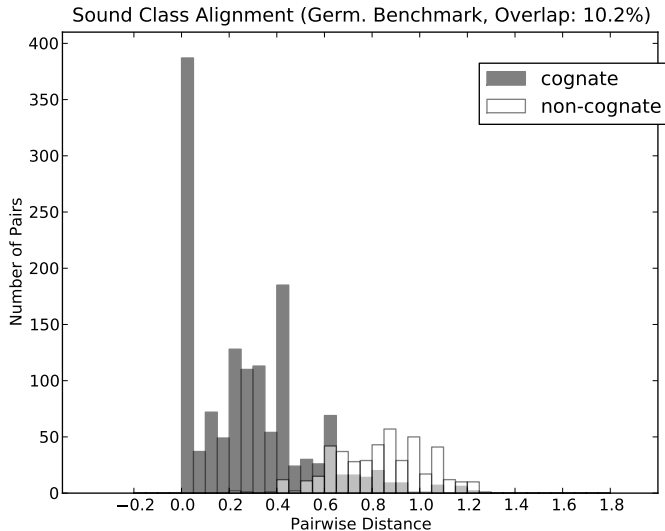
Evaluierung



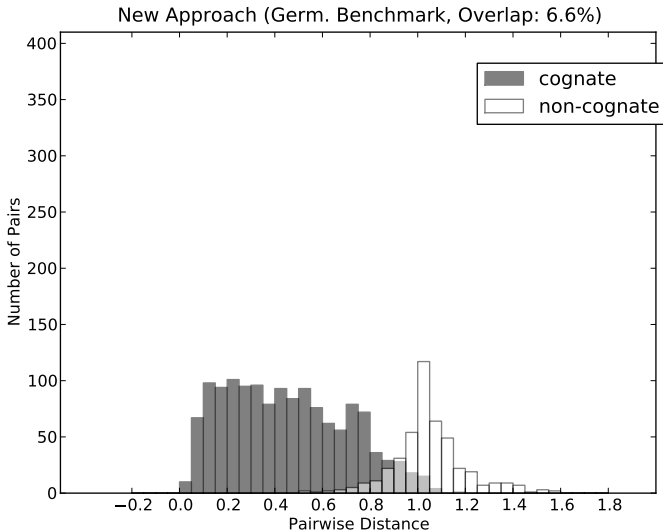
Benchmark-Datensätze für automatische Analysen

Sprachfamilie	Sprachen	Umfang	Urheber
Germanisch	6 germanische Sprachen	110 Basiswörter	Tower of Babel (orthographische Listen), Mattis List (phonetische Kodierung)
Romanisch	5 romanische Sprachen	110 Basiswörter	Tower of Babel (orthographische Listen), Mattis List (phonetische Kodierung)
Bai	9 Varietäten der Bai-Sprache	110 Basiswörter	Wang Feng (phonetische Listen)
Slavisch	4 slavische Sprachen	110 Basiswörter	Tower of Babel (orthographische Listen), Mattis List (phonetische Kodierung)
Varia	8 Sprachen (5 indogermanische, 3 unverwandte)	100 Basiswörter	Brat Kessler (phonetische Listen)
Japanisch	10 Dialekte	200 Basiswörter	Shiro (1973, orthographische Listen), Mathias Dickmanns (Digitalisierung)

Paarweiser Vergleich



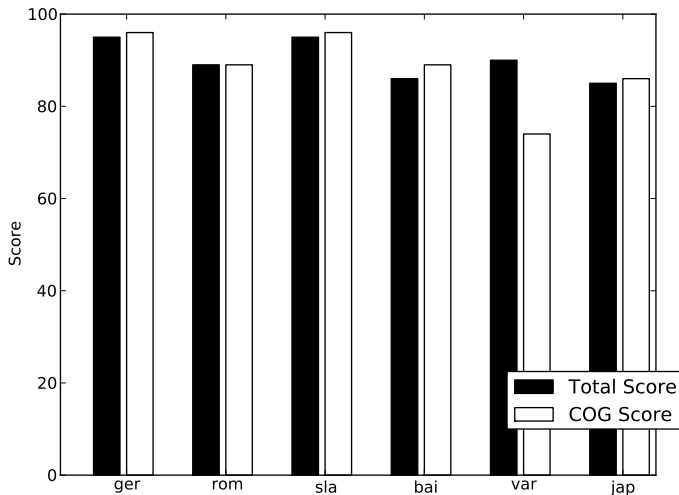
Paarweiser Vergleich



Erste Ergebnisse

Datensatz	Total	False Pos.	False Neg.	Id. Kog.	Kog. / Nicht-K.
Germanisch	95	2	3	96	3:1
Romanisch	89	2	9	89	2:1
Slavisch	95	2	3	96	3:1
Bai	86	6	8	89	3:1
Varia	90	3	7	74	1:7
Japanisch	85	6	9	86	2:1

Ergebnisse



Comparison of the Performance on the Testsets

Schlussbetrachtung

Schlussbetrachtung

- ▶ Die neue Methode funktioniert definitiv besser als reine auf Ediertistanzen basierende Methoden.

Schlussbetrachtung

- ▶ Die neue Methode funktioniert definitiv besser als reine auf Editierdistanzen basierende Methoden.
- ▶ Die Methode kommt in ihrer Arbeitsweise dem traditionellen Verfahren der historischen Linguistik näher als bisherige Methoden.

Schlussbetrachtung

- ▶ Die neue Methode funktioniert definitiv besser als reine auf Editierdistanzen basierende Methoden.
- ▶ Die Methode kommt in ihrer Arbeitsweise dem traditionellen Verfahren der historischen Linguistik näher als bisherige Methoden.
Im Gegensatz zum Blackbox-Charakter vieler automatischer
- ▶ Analysen ermöglicht es die neue Methode, explizit die Ergebnisse der Computeranalyse mit den Ergebnissen der traditionellen komparativen Methode zu vergleichen.

Schlussbetrachtung

- ▶ Die neue Methode funktioniert definitiv besser als reine auf Ediertistanzen basierende Methoden.
- ▶ Die Methode kommt in ihrer Arbeitsweise dem traditionellen Verfahren der historischen Linguistik näher als bisherige Methoden.
Im Gegensatz zum Blackbox-Charakter vieler automatischer Analysen ermöglicht es die neue Methode, explizit die Ergebnisse der Computeranalyse mit den Ergebnissen der traditionellen komparativen Methode zu vergleichen.
- ▶ Abgesehen von den Grundannahmen des Verfahrens, das noch weiter modifiziert werden muss, liegen dessen Grenzen definitiv in den Daten. Es gibt Fälle von Sprachverwandtschaft, die im Rahmen von Swadeshlisten schlicht und einfach nicht zu klären sind.

Schlussbetrachtung

- ▶ Die neue Methode funktioniert definitiv besser als reine auf Ediertistanzen basierende Methoden.
- ▶ Die Methode kommt in ihrer Arbeitsweise dem traditionellen Verfahren der historischen Linguistik näher als bisherige Methoden.
Im Gegensatz zum Blackbox-Charakter vieler automatischer Analysen ermöglicht es die neue Methode, explizit die Ergebnisse der Computeranalyse mit den Ergebnissen der traditionellen komparativen Methode zu vergleichen.
- ▶ Abgesehen von den Grundannahmen des Verfahrens, das noch weiter modifiziert werden muss, liegen dessen Grenzen definitiv in den Daten. Es gibt Fälle von Sprachverwandtschaft, die im Rahmen von Swadeshlisten schlicht und einfach nicht zu klären sind.
Automatische Ansätze, die ihre Ergebnisse explizit machen, können uns helfen, unsere traditionellen Methoden neu zu überdenken, auf deren Grenzen hinweisen, und zeigen, wo Spekulation über objektive Analyse triumphiert.

