

Of Words, Waves, and Webs

Using bioinformatics to study the lateral component of language evolution

Johann-Mattis List

Forschungszentrum Deutscher Sprachatlas
Philipps-Universität Marburg

31.01.2014



语言



язык

Languages



language



språk

Languages and Dialects

Norwegian, Swedish, and Danish are different languages

Běijīng-Chinese, Shànghǎi-Chinese und Hakka-Chinese
are dialects of the same language

Languages and Dialects

Beijing Chinese	1	iou ²¹	i ⁵⁵	xuei ³⁵	pei ²¹ fəŋ ⁵⁵	kən ⁵⁵	tʰai ⁵¹ ian ¹¹	tʂəŋ ⁵⁵	tsai ⁵³	naɛ ⁵¹	tʂəŋ ⁵⁵ luən ⁵¹
Hakka Chinese	1	iu ³³	it ⁵⁵	pai ³³ a ¹¹	pet ³³ fuiŋ ³³	tʰuŋ ¹¹	nit ¹¹ tʰeu ¹¹	hɔk ³³	e ⁵³	au ⁵⁵	
Shanghai Chinese	1	fi ²²		tʰɑ̃ ⁵⁵ tsɿ ²¹	po? ³ fɔŋ ⁴⁴	ta? ⁵	tʰa ³³ fia ⁴⁴	tsəŋ ³³ hɔ ⁴⁴		lə? ¹ lə ²³ tsa ⁵³	
Beijing Chinese	2	ʂei ³⁵		də ⁵⁵		pən ³⁵	lɪŋ ²¹	ta ⁵¹			
Hakka Chinese	2	man ³³	ɲin ¹¹			kʷɔ ⁵⁵	vɔi ⁵³				
Shanghai Chinese	2	sa ³³	ɲin ⁵⁵	fia? ²¹		pəŋ ³³	zɿ ⁴⁴	du ¹³			
Norwegian	1	nu:ravin'ɳ	ɔ	su:lɳ					krɑŋlət	ɔm	
Swedish	1	nu:qanvindən	ɔ	su:lən		tyistadə	ən gɔj			ɔm	
Danish	1	noðʌnven?ɳ	ʌ	so:l?n	kʰʌm		en ɣɑŋ	i sðøið?		ʌm?	
Norwegian	2	vem	a	dem	sŋ	va:	ɖɳ		stæfkəstə		
Swedish	2	vem	av	dɔm	sɔm	va			staikast		
Danish	2	vem?	a	bm	d	va	ɖn		sdæʌgøsðə		

Languages and Dialects

From the perspective of the lexicon and the sound system, the Chinese **dialects** are at least as diverse as the Scandinavian **languages**

Language as a diasystem

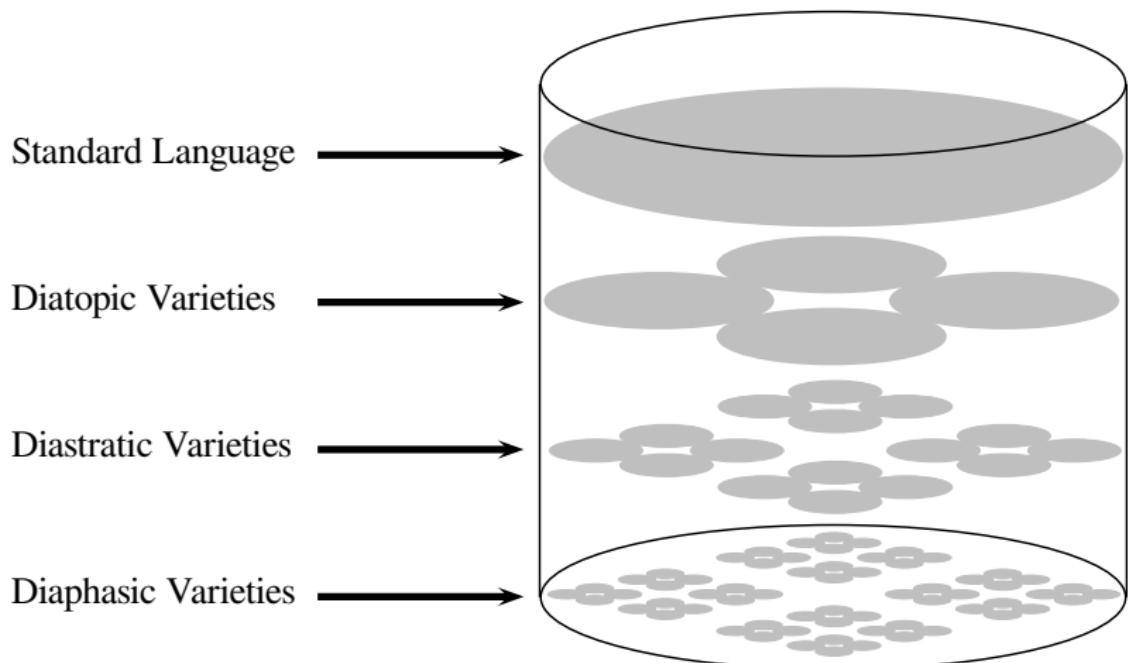
Languages are complex aggregates of different linguistic systems which “miteinander koexistieren und einander beeinflussen” (Coseriu 1973: 40).

Language as a diasystem

Languages are complex aggregates of different linguistic systems which “miteinander koexistieren und einander beeinflussen” (Coseriu 1973: 40).

A linguistic diasystem needs a “roof language” (Goossens 1973: 11), a linguistic variety that serves as a standard for interdialectal communication.

Language as a diasystem



Change



Change



expected

Mandarin

[ma₅₅po₂₁lou]

Change



expected

Mandarin

[ma₅₅po₂₁lou]

attested

Mandarin

[wan₅₁paw₂₁lu₅₁]

Change



expected

Mandarin

[ma₅₅po₂₁lou]

attested

Mandarin

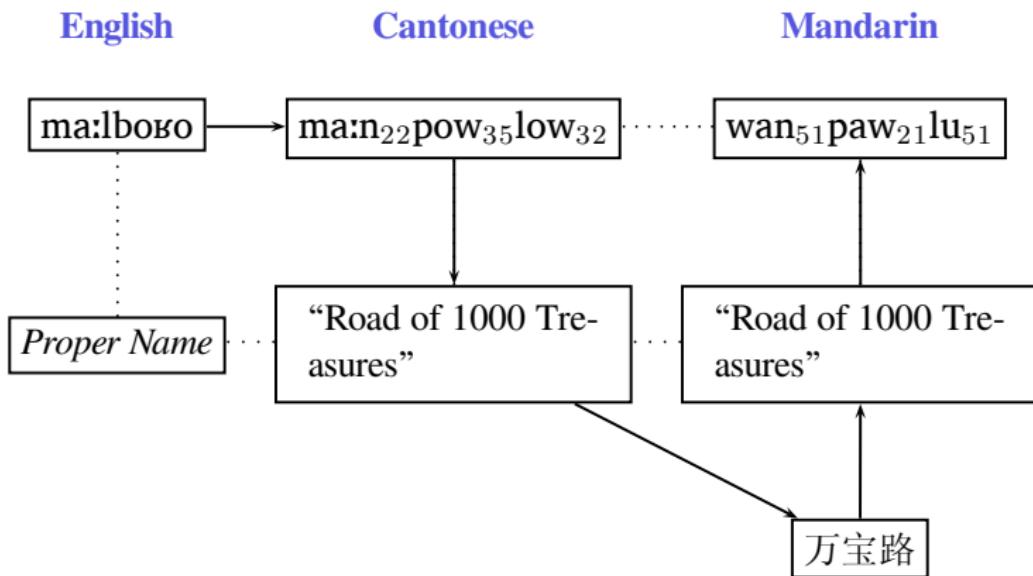
[wan₅₁paw₂₁lu₅₁]

explanation

Cantonese

[ma:n₂₂pow₃₅low₃₂]

Change



Wind of Sound Change in China

燕	燕	于	飛，	下	上	其	音。	The swallows go flying, falling and rising are their voices;
zhī	zǐ	yú	guī,	yuǎn	sòng	yú	nán	This young lady goes to her new home, far I accompany her to the south.
zhān	wàng	fú	jí,	shí	láo	wǒ	xīn	I gaze after her, can no longer see her, truly it grieves my heart.

Wind of Sound Change in China

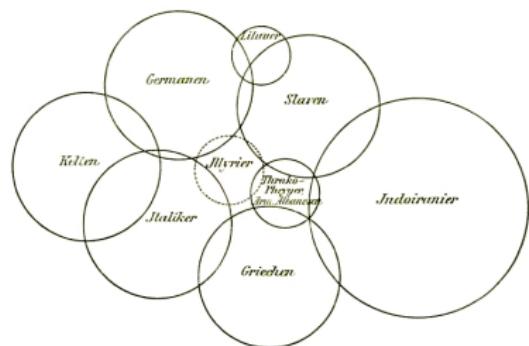
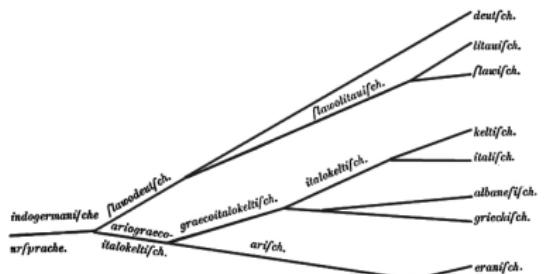
燕	燕	于	飛，	下	上	其	音。	The swallows go flying, falling and rising are their voices;
yān	yān	yú	fēi	xià	shàng	qí	yīn	
之	子	于	歸，	遠	送	于	南。	This young lady goes to her new home, far I accompany her to the south.
zhī	zǐ	yú	guī,	yuǎn	sòng	yú	nán	
瞻	望	弗	及，	實	勞	我	心。	I gaze after her, can no longer see her, truly it grieves my heart.
zhān	wàng	fú	jí,	shí	láo	wǒ	xīn	

Wind of Sound Change in China

燕	燕	于	飛，	下	上	其	音。	The swallows go flying, falling and rising are their voices;
<i>yān</i>	<i>yān</i>	<i>yú</i>	pjij	<i>xià</i>	<i>shàng</i>	<i>qí</i>	2jím	
之	子	于	歸，	遠	送	于	南。	This young lady goes to her new home, far I accompany her to the south.
<i>zhī</i>	<i>zǐ</i>	<i>yú</i>	kʷjij,	<i>yuǎn</i>	<i>sòng</i>	<i>yú</i>	ním	
瞻	望	弗	及，	實	勞	我	心。	I gaze after her, can no longer see her, truly it grieves my heart.
<i>zhān</i>	<i>wàng</i>	<i>fú</i>	<i>jí,</i>	<i>shí</i>	<i>láo</i>	<i>wǒ</i>	sjím	



Modelling Language History



Dendrophilia

August Schleicher
(1821-1868)



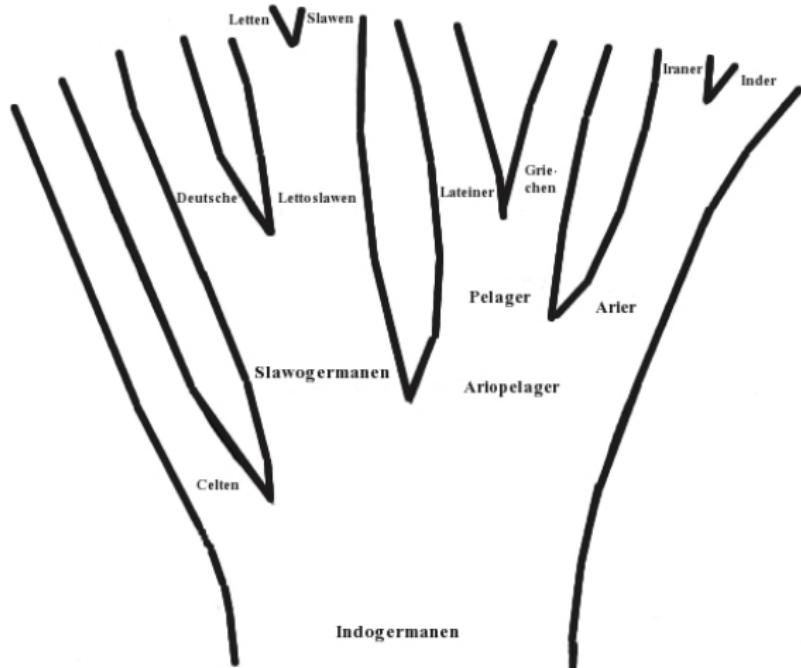
Dendrophilia

“Diese Annahmen, logisch folgend aus den Ergebnissen der bisherigen Forschung, lassen sich am besten unter dem Bilde eines sich verästelnden Baumes anschaulich machen.”(Schleicher 1853: 787)



August Schleicher
(1821-1868)

Dendrophilia



Schleicher (1853)

Dendrophobia



Johannes Schmidt
(1843-1901)

Dendrophobia



„Man mag sich also drehen und wenden wie man will, so lange man an der anschauung fest hält, dass die in historischer zeit erscheinenden sprachen durch merfache gabelungen aus der ursprache hervorgegangen seien, d.h. so lange man einen stammbaum der indo-germanischen sprachen annimmt, wird man nie dazu gelangen alle die hier in frage stehenden tatsachen wissenschaftlich zu erklären.“ (Schmidt 1872: 17, my translation)

Johannes Schmidt
(1843-1901)

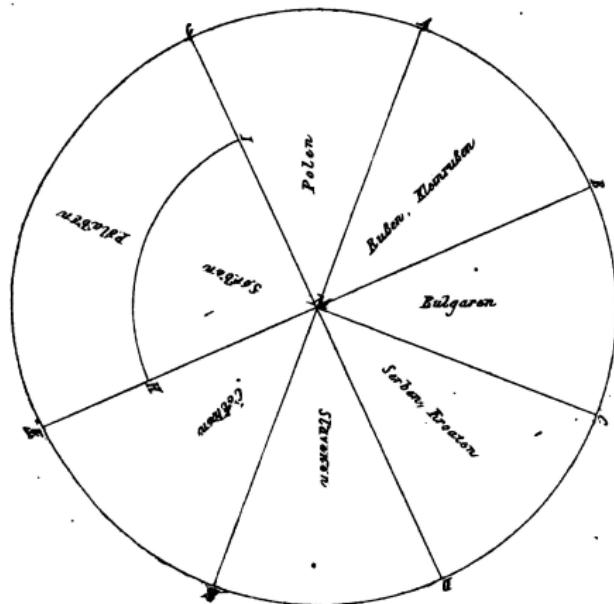
Dendrophobia



„Ich möchte an seine [des Baumes] stelle das bild der welle setzen, welche sich in concentrischen mit der entfernung vom mittelpunkte immer schwächer werdenden ringen ausbreitet.“ (Schmidt 1872: 27)

Johannes Schmidt
(1843-1901)

Dendrophobia

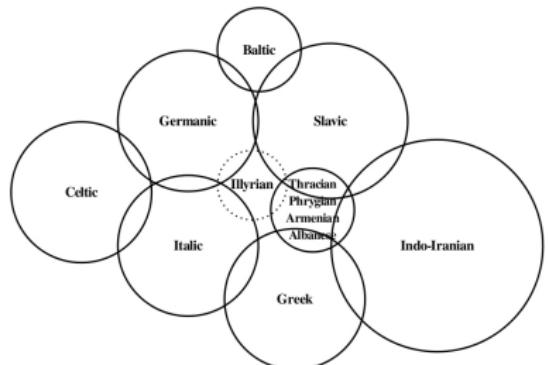


Schmidt (1875)

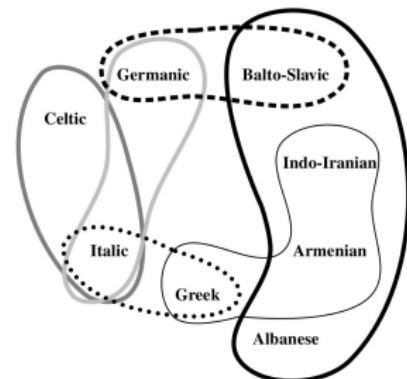
Dendrophobia



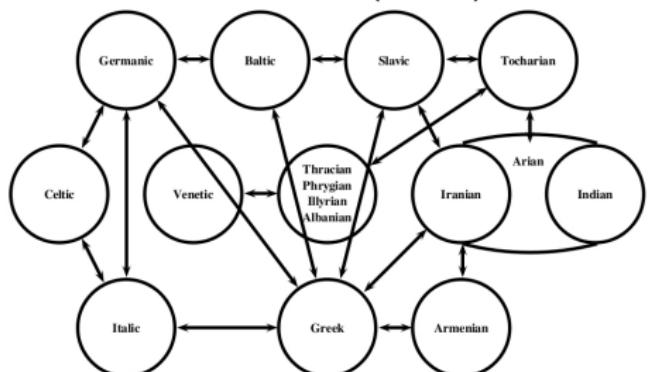
Meillet (1908)



Hirt (1905)



Bloomfield (1933)



Bonfante (1931)

Phylogenetic Networks

Trees are bad, because...

Phylogenetic Networks

Trees are bad, because...

- they are difficult to reconstruct

Phylogenetic Networks

Trees are bad, because...

- they are difficult to reconstruct
- languages do not always split

Phylogenetic Networks

Trees are bad, because...

- they are difficult to reconstruct
- languages do not always split
- they are boring, since they only model the vertical aspects of language history

Phylogenetic Networks

Trees are bad, because...

- they are difficult to reconstruct
- languages do not always split
- they are boring, since they only model the vertical aspects of language history

Waves are bad, because

- nobody knows how to reconstruct them

Phylogenetic Networks

Trees are bad, because...

- they are difficult to reconstruct
- languages do not always split
- they are boring, since they only model the vertical aspects of language history

Waves are bad, because

- nobody knows how to reconstruct them
- languages still diverge, even if not necessarily in split processes

Phylogenetic Networks

Trees are bad, because...

- they are difficult to reconstruct
- languages do not always split
- they are boring, since they only model the vertical aspects of language history

Waves are bad, because

- nobody knows how to reconstruct them
- languages still diverge, even if not necessarily in split processes
- they are boring, since they only model the horizontal aspects of language history

Phylogenetic Networks



Hugo Schuchardt
(1842-1927)

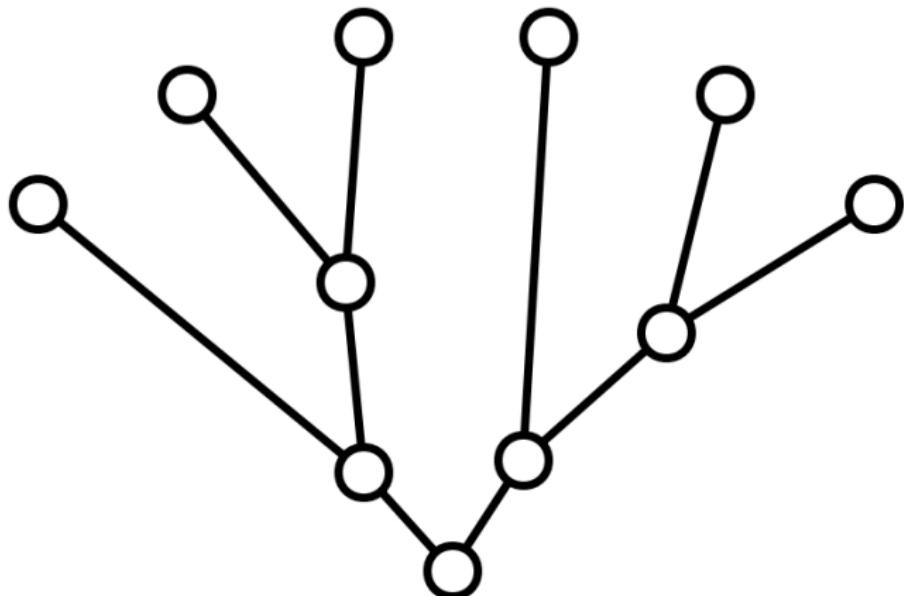
Phylogenetic Networks



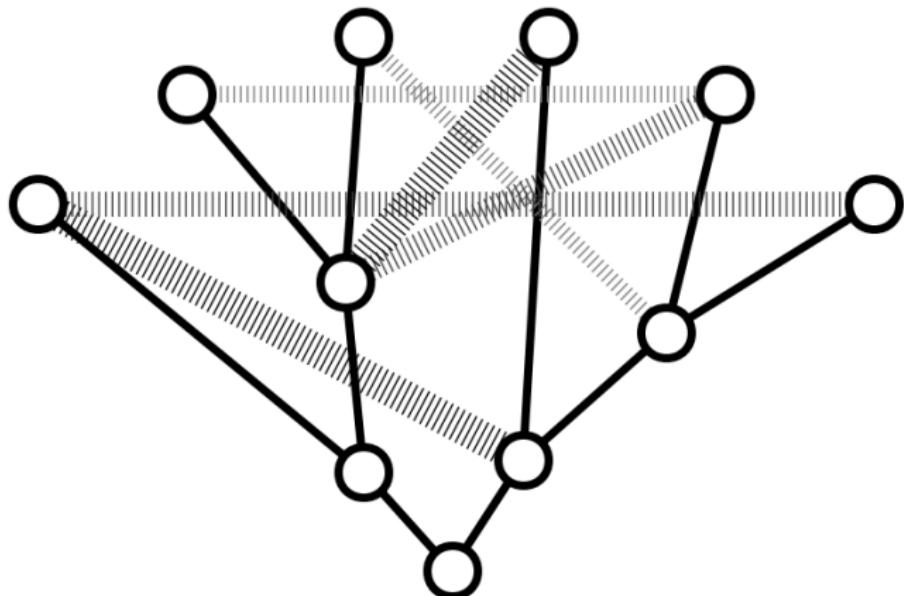
“Wir verbinden die Äste und Zweige des Baumes mit zahllosen horizontalen Linien, und er hört auf ein Baum zu sein.” (Schuchardt 1870 [1900]: 11)

Hugo Schuchardt
(1842-1927)

Phylogenetic Networks



Phylogenetic Networks



Linguistics and Biology



historical linguistics and evolutionary biology



historical linguistics and **evolutionary biology**

historical linguistics and **generative grammar**

historical linguistics and **the comparative study of african languages**

historical linguistics and **philology**

Weitere Informationen

The Quantitative Turn

A screenshot of a Facebook page for "Historical Linguistics". The page header features the word "Linguistics" in a stylized font. Below it, the page title is "Historical Linguistics" and the status message is "is on Facebook." A call-to-action button says "To connect with Linguistics, sign up for Facebook today." Below the status message are "Sign Up" and "Log In" buttons. To the right of the status message is a "Like" button with a count of 248 likes. The "About" section describes the page as "Education everything about linguistics". The status message "Historical Linguistics hat leider noch keine Freunde" is displayed in red. Below the status message are "Photos" and "Likes" sections. A "Post" and "Photo / Video" button is at the bottom left. A "Highlights" dropdown menu is visible above the post area. A "Likes" section shows "Tuane Linguistics Organization" liked the page.

Historical Linguistics is on Facebook.

To connect with Linguistics, sign up for Facebook today.

Sign Up Log In

Linguistics

248 likes · 3 talking about this

Education everything about linguistics

Historical Linguistics hat leider noch keine Freunde

About Photos Likes

Post Photo / Video

Write something...

Highlights ▾

Tuane Linguistics Organization

See All

Like

The Quantitative Turn

A screenshot of a Facebook page for the group "Linguistics". The page has 248 likes and 3 people talking about it. A red annotation on the left side of the page reads: "Historical Linguistics ist jetzt mit Biologie befreundet! :-)" (Historical Linguistics is now friends with Biology! :-)). The page also features a post from "Tuane Linguistics" and a link to "See All Likes". The top right corner of the screenshot shows a red annotation with a crossed-out thumbs-up icon and the number "5000!!!".

Historical Linguistics
is on Facebook.

To connect with Linguistics, sign up for Facebook today.

Sign Up Log In

Linguistics
248 likes · 3 talking about this

Education
everything about linguistics

Historical Linguistics ist jetzt mit Biologie befreundet! :-)

About Photos Likes

Post Photo / Video Write something... Highlights ▾

Tuane Linguistics Organization See All

Like

Like

The Quantitative Turn

- “Indo-European and computational cladistics” (Ringe, Warnow and Taylor 2002)
- “Language-tree divergence times support the Anatolian theory of Indo-European origin” (Gray und Atkinson 2003)
- “Language classification by numbers” (McMahon und McMahon 2005)
- “Curious Parallels and Curious Connections: Phylogenetic Thinking in Biology and Historical Linguistics” (Atkinson und Gray 2005)
- “Automated classification of the world’s languages” (Brown et al. 2008)
- “Indo-European languages tree by Levenshtein distance” (Serva and Petroni 2008)
- “Networks uncover hidden lexical borrowing in Indo-European language evolution” (Nelson-Sathi et al. 2011)

The Quantitative Turn

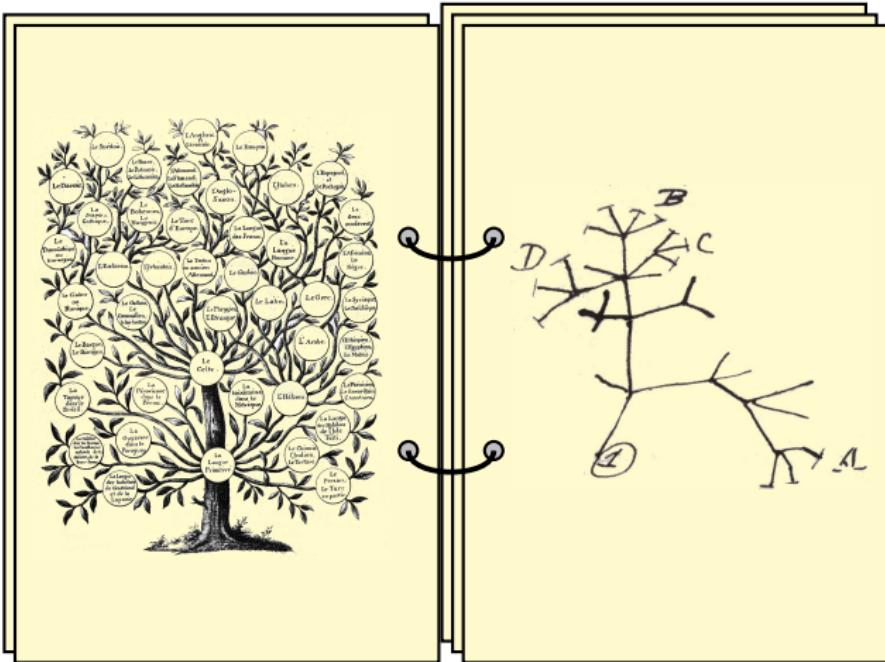
- “Indo-European and computational **cladistics**” (Ringe, Warnow and Taylor 2002)
- “**Language-tree** divergence times support the Anatolian theory of Indo-European origin” (Gray und Atkinson 2003)
- “Language **classification** by numbers” (McMahon und McMahon 2005)
- “Curious Parallels and Curious Connections: Phylogenetic Thinking in Biology and Historical Linguistics” (Atkinson und Gray 2005)
- “Automated **classification** of the world’s languages” (Brown et al. 2008)
- “Indo-European languages **tree** by Levenshtein distance” (Serva and Petroni 2008)
- “Networks uncover hidden lexical borrowing in Indo-European language evolution” (Nelson-Sathi et al. 2011)

Parallels

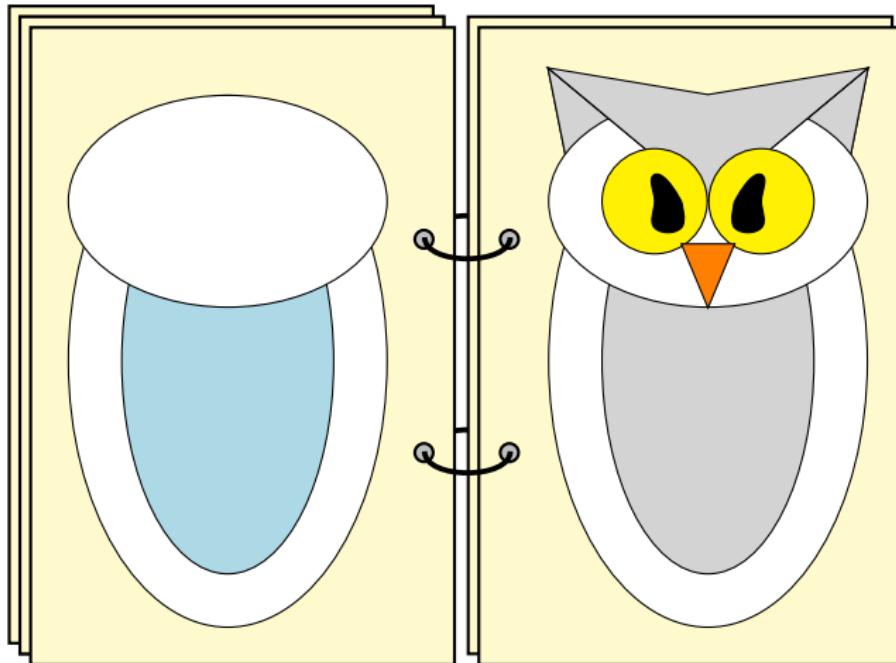
Parallels according to Pagel (2009)

aspect	species	languages
unit of replication	gene	word
replication	asexual und sexual reproduction	learning
speciation	cladogenesis	language split
forces of change	natural selection and genetic drift	social selection and trends
differentiation	tree-like	tree-like

Parallels?



Parallels?



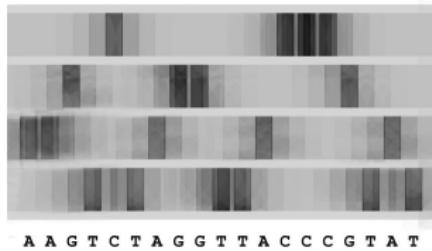
Differences

Differences (Geisler & List 2013)

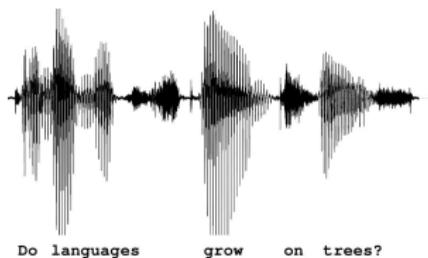
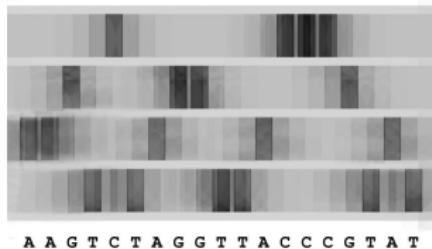
Aspect	Species	Languages
domain	Popper's World I	Popper's World III
relation between form and function	mechanical	arbitrary
origin	monogenesis	unclear
sequence similarity	universal (independent of species)	language-specific
differentiation	tree-like	network-like

These differences are ignored in most of the recent applications of bioinformatic methods in historical linguistics.

Differences: Alphabets

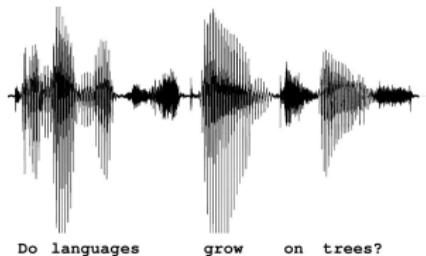
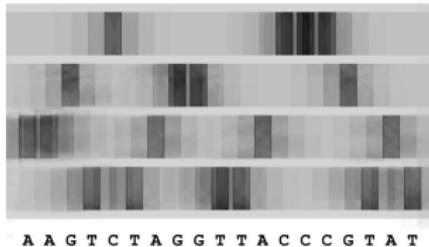


Differences: Alphabets



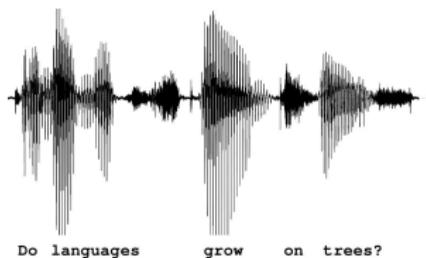
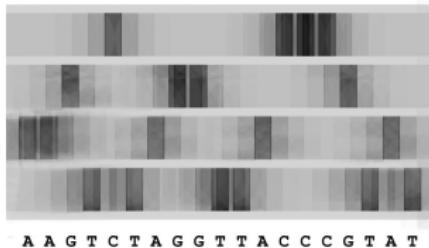
- universal
- language-specific

Differences: Alphabets



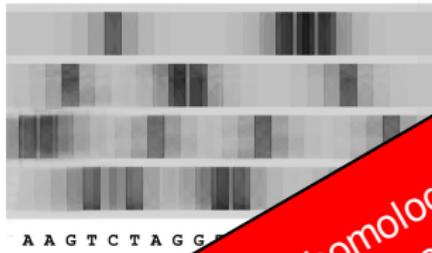
- universal
- limited
- language-specific
- widely varying

Differences: Alphabets



- universal
- limited
- constant
- language-specific
- widely varying
- mutable

Differences: Alphabets



In order to identify homologous words in different languages, not only corresponding segments have to be identified, but also mappings between the alphabets. Phonetic alignment is thus similar to the task of aligning two sequences which have been drawn from **two different alphabets!**

- language-specific
- widely varying
- mutable

Differences: Alphabets

Differences: Alphabets

Cognate List	
German	<i>dünn</i>
English	<i>thin</i>
German	<i>Ding</i>
English	<i>thing</i>
German	<i>dumm</i>
English	<i>dumb</i>

Differences: Alphabets

Cognate List		Alignment		
German	<i>dünn</i>	d	y	n
English	<i>thin</i>	θ	i	n
German	<i>Ding</i>	d	i	ŋ
English	<i>thing</i>	θ	i	ŋ
German	<i>dumm</i>	d	ʊ	m
English	<i>dumb</i>	d	ʌ	m

Differences: Alphabets

Cognate List		Alignment			Correspondence List		
German	English	d	y	n	GER	ENG	Frequ.
German <i>dünn</i>	English <i>thin</i>	d	θ	n	d	θ	2 x
German <i>Ding</i>	English <i>thing</i>	d	i	ŋ	d	d	1 x
German <i>dumm</i>	English <i>dumb</i>	d	ʊ	m	n	n	1 x
		d	ʌ	m	m	m	1 x
					ŋ	ŋ	1 x

Differences: Alphabets

Cognate List		Alignment			Correspondence List		
German	English	d	y	n	GER	ENG	Frequ.
<i>dünn</i>	<i>thin</i>	d	θ	n	d	θ	2 x
<i>Ding</i>	<i>thing</i>	d	i	ŋ	d	d	1 x
<i>dumm</i>	<i>dumb</i>	d	ʊ	m	n	n	1 x
<i>Dorn</i>	<i>thorn</i>	d	ʌ	m	m	m	1 x
		ŋ	ɔ:	n	ŋ	ŋ	1 x

Differences: Alphabets

Cognate List		Alignment			Correspondence List		
German	English	d	y	n	GER	ENG	Frequ.
<i>dünn</i>	<i>thin</i>	d	θ	n	d	θ	3 x
<i>Ding</i>	<i>thing</i>	d	ɪ	ŋ	d	d	1 x ?
<i>dumm</i>	<i>dumb</i>	d	ʊ	m	n	n	2 x
<i>Dorn</i>	<i>thorn</i>	d	ʌ	m	m	m	1 x
		ŋ	ɔ:	n	ŋ	ŋ	1 x

Differences: Alphabets

Cognate List		Alignment			Correspondence List		
		d	y	n	GER	ENG	Frequ.
German	<i>dünn</i>	d	y	n	d	θ	3 x
English	<i>thin</i>	θ	i	n	d	d	1 x
German	<i>Ding</i>	d	i	ŋ	n	n	2 x
English	<i>thing</i>	θ	i	ŋ	m	m	1 x
German	<i>dumm</i>	d	u	m	ŋ	ŋ	1 x
English	<i>dumb</i>	d	ʌ	m			
German	<i>Dorn</i>	d	ac	n			
English	<i>thorn</i>	θ	ɔ:	n			

Differences: Alphabets

Cognate List		Alignment			Correspondence List		
German	English	d	y	n	GER	ENG	Frequ.
German <i>dünn</i>	English <i>thin</i>	d	y	n	d	θ	3 x
German <i>Ding</i>	English <i>thing</i>	d	i	ŋ	n	n	2 x
German <i>Dorn</i>	English <i>thorn</i>	d	əc	n	ŋ	ŋ	1 x

Differences: Borrowing

Of the 1,000 most frequent Latin words (Stefenelli 1992),

Differences: Borrowing

Of the 1,000 most frequent Latin words (Stefenelli 1992),

- 67% were directly inherited in at least one of the descendant languages of Latin,

Differences: Borrowing

Of the 1,000 most frequent Latin words (Stefenelli 1992),

- 67% were directly inherited in at least one of the descendant languages of Latin,
- 14% were directly inherited in all descendant languages,

Differences: Borrowing

Of the 1,000 most frequent Latin words (Stefenelli 1992),

- 67% were directly inherited in at least one of the descendant languages of Latin,
- 14% were directly inherited in all descendant languages,
- only 33% are completely lost,

Differences: Borrowing

Of the 1,000 most frequent Latin words (Stefenelli 1992),

- 67% were directly inherited in at least one of the descendant languages of Latin,
- 14% were directly inherited in all descendant languages,
- only 33% are completely lost,
- about 50% of the words survive as borrowings from Latin in the descendant languages

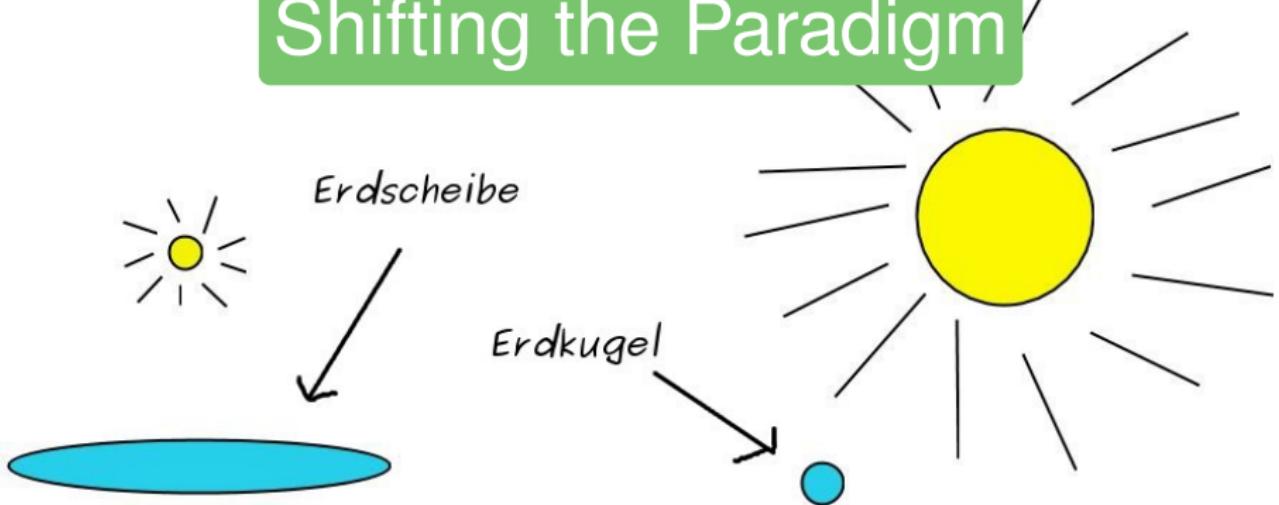
Differences: Borrowing

Of the 1,000 most frequent Latin words:

- 67% were directly inherited from the languages of Latin
- 14% were directly derived from their descendant languages,
- only 32% survive as borrowings from Latin in the descendant languages
- about 10% of the words in English are derived from Latin, but it's only a part of the whole interesting story.

Saying that languages evolve in tree-like processes is similar to saying that penguins walk: It may be true, but it's only a part of the whole interesting story.

Shifting the Paradigm



New Parallels

If we sequence 61 human genomes, we will find more or less the same collection of about 30,000 genes in each individual. But if we sequence 61 genomes of *Escherichia coli* (Lukjancenko et al. 2010)

New Parallels

If we sequence 61 human genomes, we will find more or less the same collection of about 30,000 genes in each individual. But if we sequence 61 genomes of *Escherichia coli* (Lukjancenko et al. 2010)

- we find about 4,500 genes in each individual,

New Parallels

If we sequence 61 human genomes, we will find more or less the same collection of about 30,000 genes in each individual. But if we sequence 61 genomes of *Escherichia coli* (Lukjancenko et al. 2010)

- we find about 4,500 genes in each individual,
- we find 1,000 genes present in all genomes,

New Parallels

If we sequence 61 human genomes, we will find more or less the same collection of about 30,000 genes in each individual. But if we sequence 61 genomes of *Escherichia coli* (Lukjancenko et al. 2010)

- we find about 4,500 genes in each individual,
- we find 1,000 genes present in all genomes,
- we find about 18,000 different genes distributed among all genomes.

New Parallels

Eukaryotic and Prokaryotic Evolution

Eukaryotic populations generate tree-like divergence structures over time, while genome evolution in prokaryotes generates both tree-like and net-like components.

New Parallels

Eukaryotic and Prokaryotic Evolution

Eukaryotic populations generate tree-like divergence structures over time, while genome evolution in prokaryotes generates both tree-like and net-like components.

Evolution and Language History

Recalling the scores on borrowing frequency in the descendant languages of Latin, it seems obvious that language history shows a much closer resemblance to prokaryotic evolution than to eukaryotic evolution. When trying to apply methods from bioinformatics to linguistic problems, it seems therefore more fruitful to use those methods that explicitly deal with prokaryotic evolution.

Minimal Lateral Networks

Biological Workflow (Dagan and Martin 2007, Dagan et al. 2008)

- ① collect phyletic pattern data (shared gene families) of the taxa that shall be investigated
- ② use **gain-loss mapping** techniques with different **weighting models**, allowing for different amounts of gain events to analyze how the gene families evolved along a given **reference tree**
- ③ use **ancestral genome sizes** as an external criterion to determine the best weighting model
- ④ assume that all patterns for which the best model yields more than one gain event result from lateral gene transfer
- ⑤ reconstruct a **minimal lateral network** by connecting multiple gains for the same gene family by lateral edges

Minimal Lateral Networks

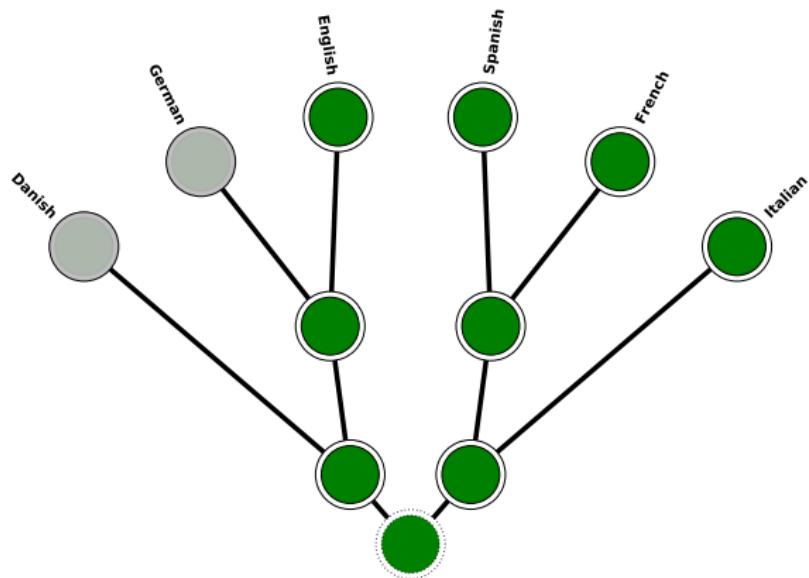
Linguistic Workflow (Nelson-Sathi et al. 2011, List et al. 2014)

- ① collect phyletic pattern data (**shared cognates**) of the languages that shall be investigated
- ② use **gain-loss mapping** techniques with different **weighting models**, allowing for different amounts of to analyze how the **cognates** evolved along a given **reference tree**
- ③ use **ancestral vocabulary size distributions** as an external criterion to determine the best weighting model
- ④ allow for a substantial amount (5%) of **parallel evolution**
- ⑤ assume that all patterns for which the best model yields more than one gain event result from lateral gene transfer
- ⑥ reconstruct a **minimal lateral network** by connecting multiple gains of the same **cognate** by lateral edges

Minimal Lateral Networks: Gain-Loss Mapping

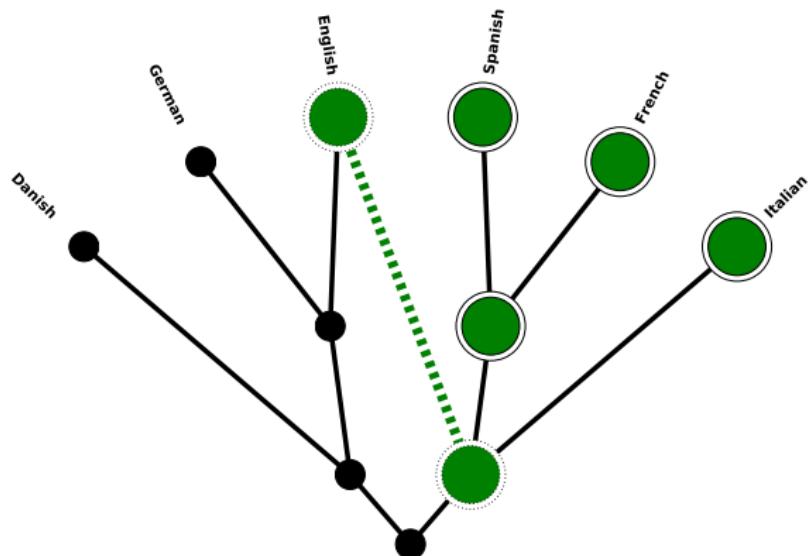
Language Variety	Danish	German	English	Spanish	French	Italian
“to count”	<i>tælle</i>	<i>zählen</i>	<i>count</i>	<i>contar</i>	<i>compter</i>	<i>contare</i>
Latin <i>computare</i>	0	0	1	1	1	1
Proto-Germanic * <i>tal-</i>	1	1	0	0	0	0

Minimal Lateral Networks: Gain-Loss Mapping



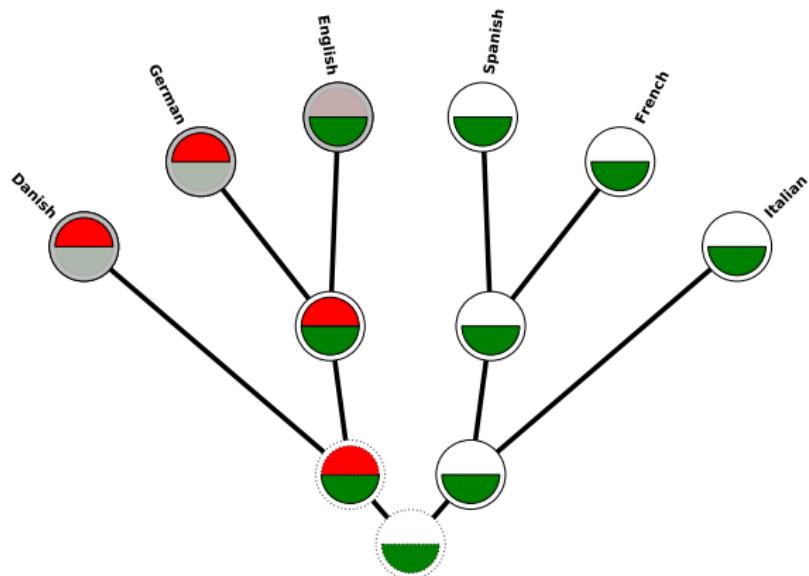
Language Variety	Danish	German	English	Spanish	French	Italian
“to count”	<i>tælle</i>	<i>zählen</i>	<i>count</i>	<i>contar</i>	<i>compter</i>	<i>contare</i>
Latin <i>computare</i>	0	0	1	1	1	1
Proto-Germanic * <i>tal-</i>	1	1	0	0	0	0

Minimal Lateral Networks: Gain-Loss Mapping



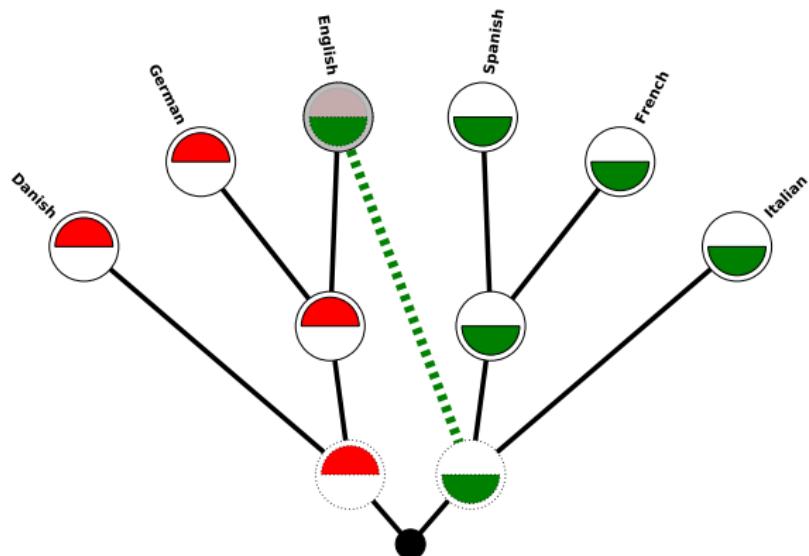
Language Variety	Danish	German	English	Spanish	French	Italian
“to count”	<i>tælle</i>	<i>zählen</i>	<i>count</i>	<i>contar</i>	<i>compter</i>	<i>contare</i>
Latin <i>computare</i>	0	0	1	1	1	1
Proto-Germanic * <i>tal-</i>	1	1	0	0	0	0

Minimal Lateral Networks: Gain-Loss Mapping



Language Variety	Danish	German	English	Spanish	French	Italian
“to count”	<i>tælle</i>	<i>zählen</i>	<i>count</i>	<i>contar</i>	<i>compter</i>	<i>contare</i>
Latin <i>computare</i>	0	0	1	1	1	1
Proto-Germanic * <i>tal-</i>	1	1	0	0	0	0

Minimal Lateral Networks: Gain-Loss Mapping



Language Variety	Danish	German	English	Spanish	French	Italian
“to count”	<i>tælle</i>	<i>zählen</i>	<i>count</i>	<i>contar</i>	<i>compter</i>	<i>contare</i>
Latin <i>computare</i>	0	0	1	1	1	1
Proto-Germanic * <i>tal-</i>	1	1	0	0	0	0

Application: Indo-European Data (List et al. 2014)

Data

- 40 Indo-European languages (taken from the IELEX, Dunn 2012)
- 1190 cognate sets (207 semantic glosses)
- 105 cognate sets contain known borrowings
- traditional reference tree, reflecting a very broad consensus, taken from Ethnologue (Lewis and Fennig 2013)

Application: Indo-European Data (List et al. 2014)

Analysis

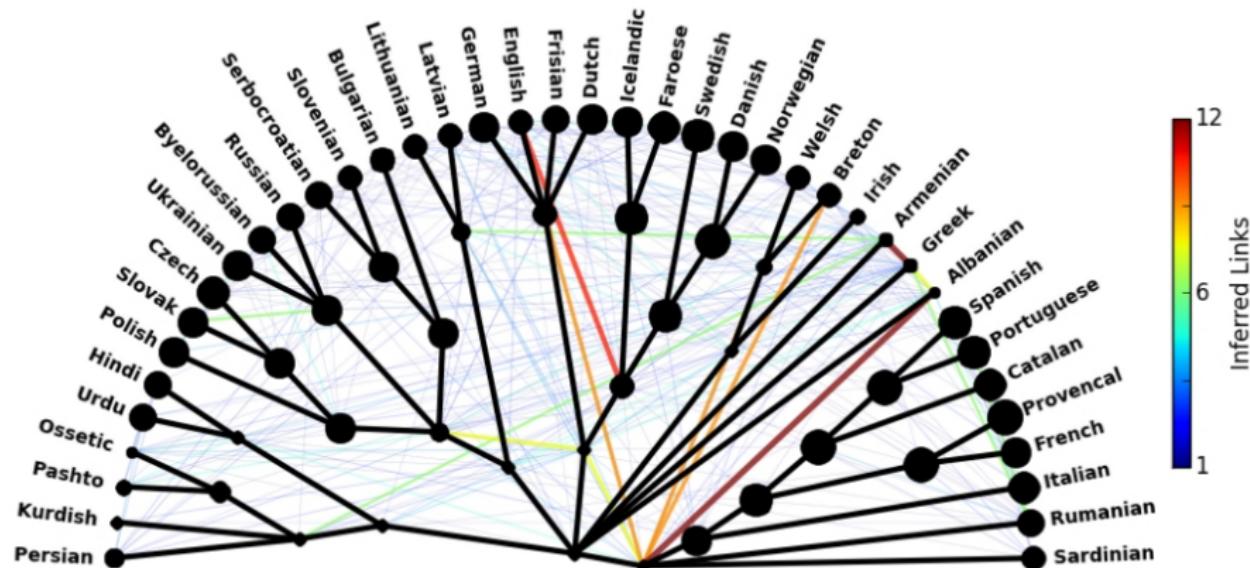
- bottom-up parsimony-based approach for gain-loss mapping using different weight ratios for gain and loss events
- modified analysis allows for multifurcating (polytomic) reference trees
- specific factor for parallel evolution was added to the evaluation procedure
- implementation as part of the LingPy Python library for quantitative tasks in historical linguistics (<http://lingpy.org>, Version 2.2, List et al. 2013)

Application: Indo-European Data (List et al. 2014)

Results

- 76 cognate sets correctly identified as borrowings
- 31% of all cognate sets could not be properly explained by the reference tree
- 17 out of 19 borrowings in English correctly identified
- well-known contact situations among major groups and languages were correctly identified

Application: Indo-European Data (List et al. 2014)



Application: Chinese Dialects (List et. al forthcoming)

Data

- lexical data of 40 Chinese dialects (Hóu 2004)
- 1056 cognate sets (180 semantic glosses)
- two traditional reference trees reflecting competing hypotheses, and two automatically generated reference trees (Neighbor-Joining and UPGMA)

Application: Chinese Dialects (List et. al forthcoming)

Analysis

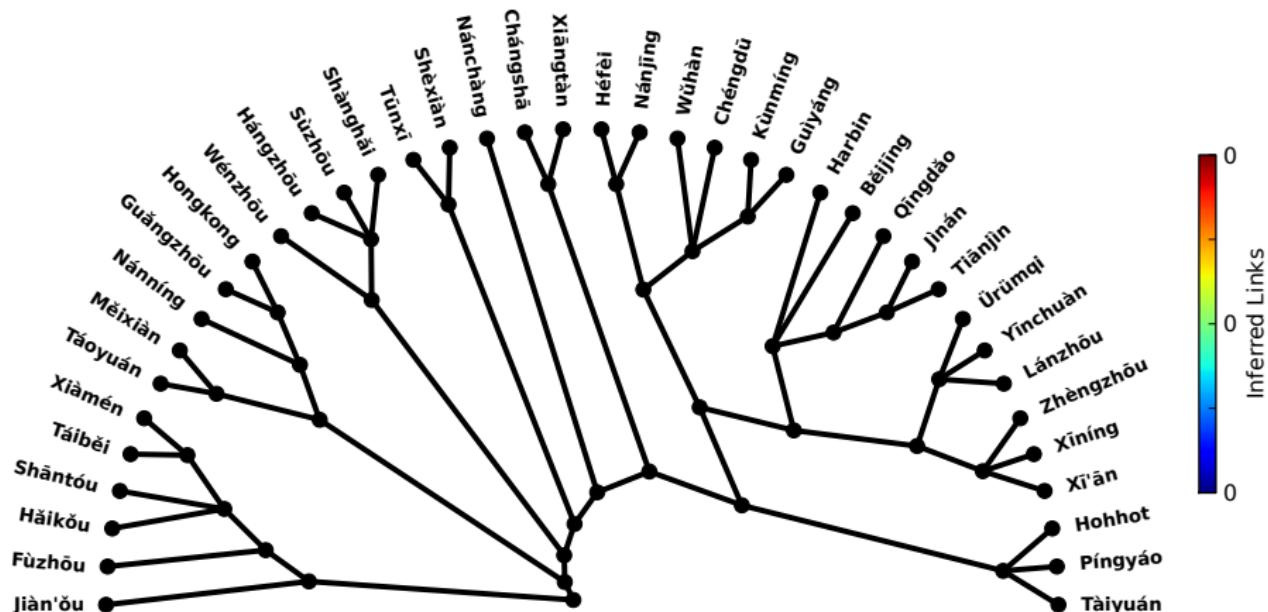
- calculate **minimal spatial networks** by plotting the inferred lateral connections onto geographic maps

Application: Chinese Dialects (List et. al forthcoming)

Results

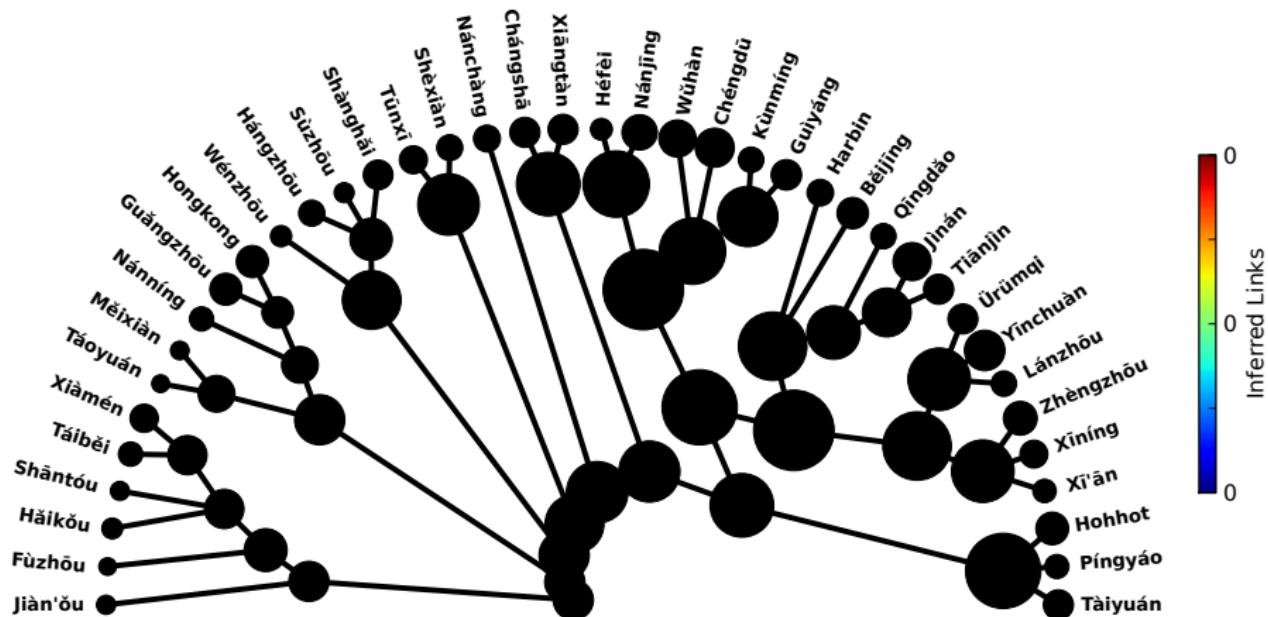
- between 48% (UPGMA) and 55% (Neighbor-Joining) of the characters cannot be explained by the reference trees
- although not showing the highest degree (be it weighted or unweighted) in the minimal lateral network, Běijīng Chinese shows the highest proportion of cognate sets which are suggestive of borrowing (40-42%): this reflects the important role that Běijīng Chinese plays as the current standard language for interdialectal communication and education in China

Application: Chinese Dialects (List et al. forthcoming)



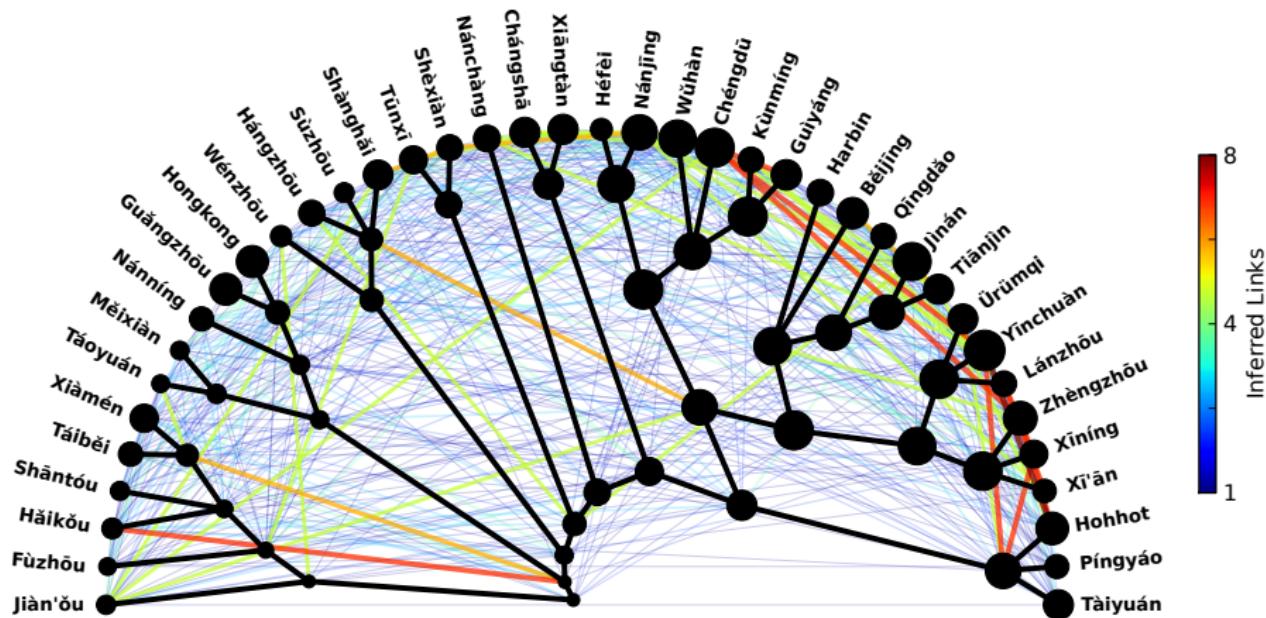
Reference tree of the Chinese dialects

Application: Chinese Dialects (List et al. forthcoming)



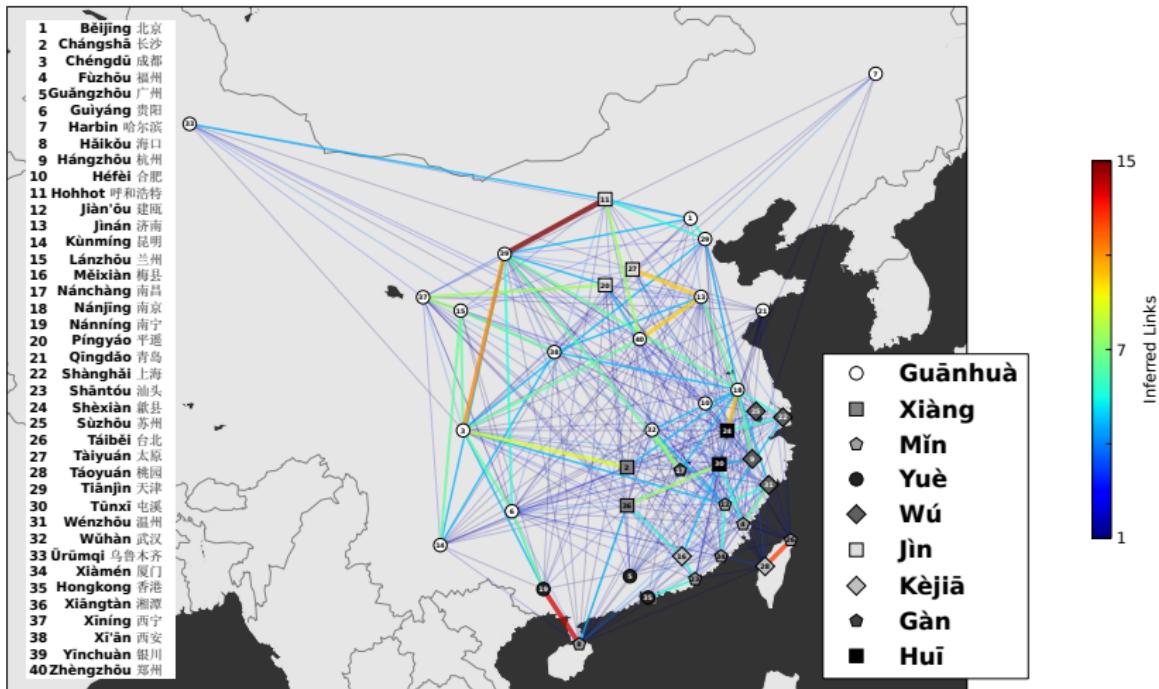
MLN analysis, no borrowing allowed

Application: Chinese Dialects (List et al. forthcoming)

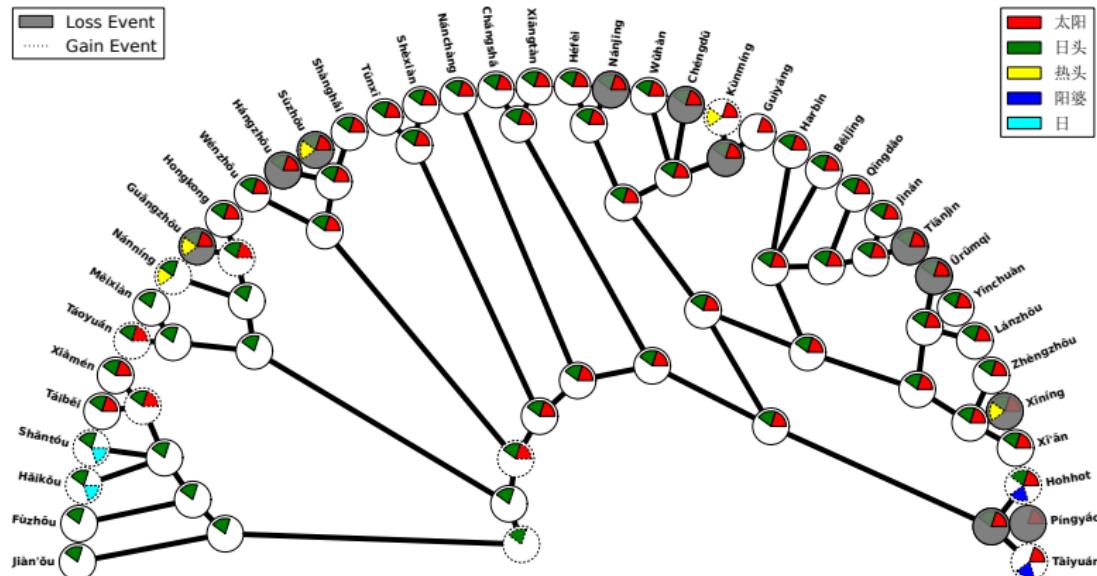


MLN analysis, best fit of borrowing and inheritance

Application: Chinese Dialects (List et al. forthcoming)

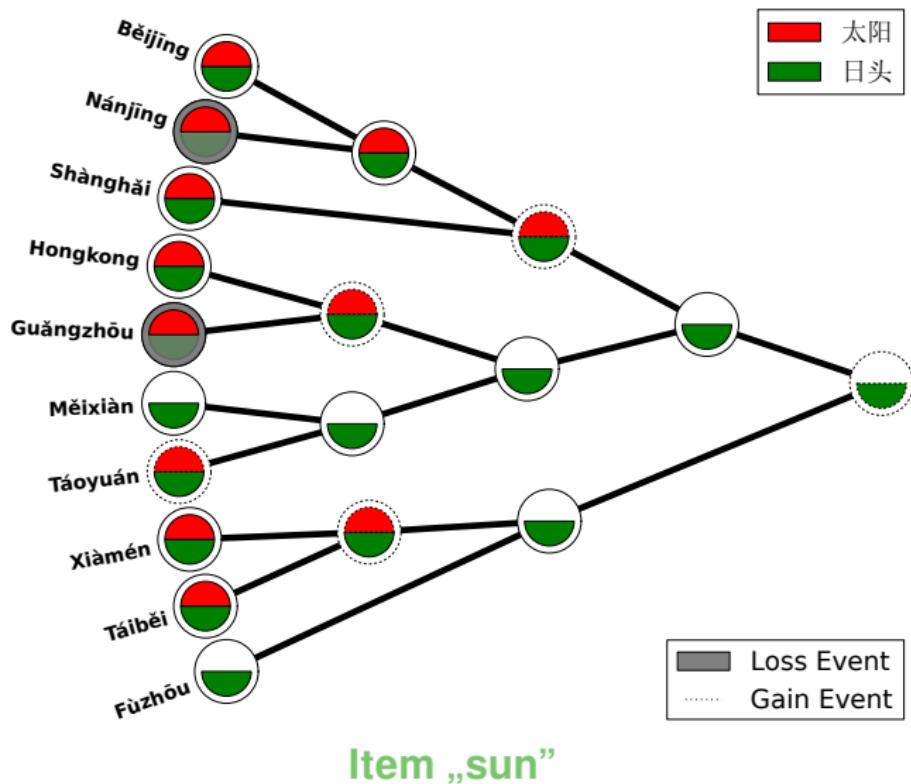


Application: Chinese Dialects (work in progress)

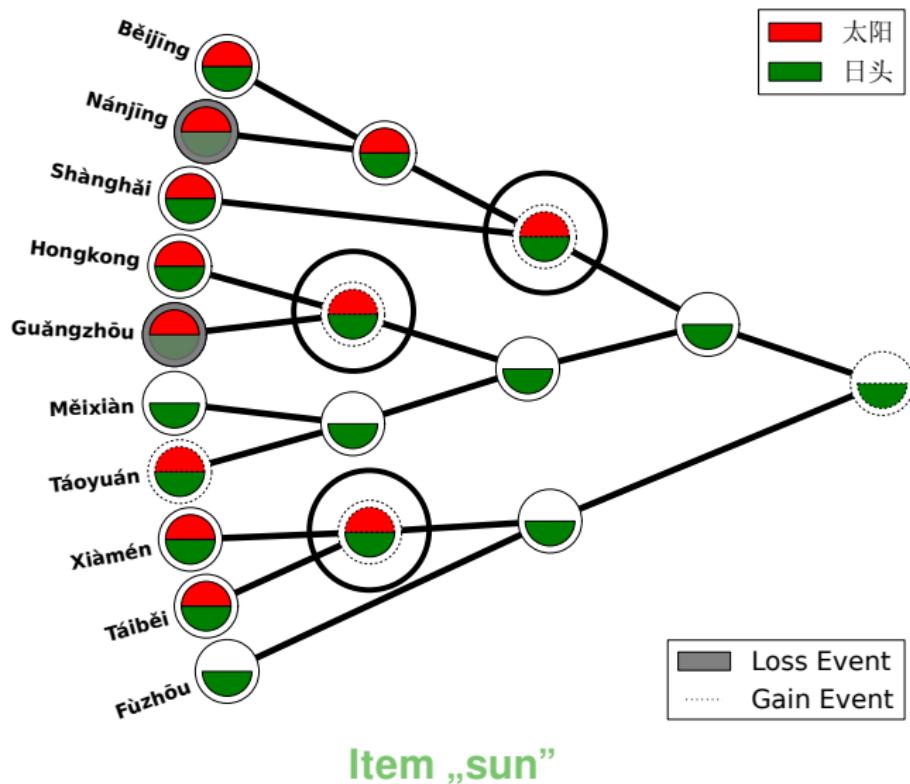


Item „sun”

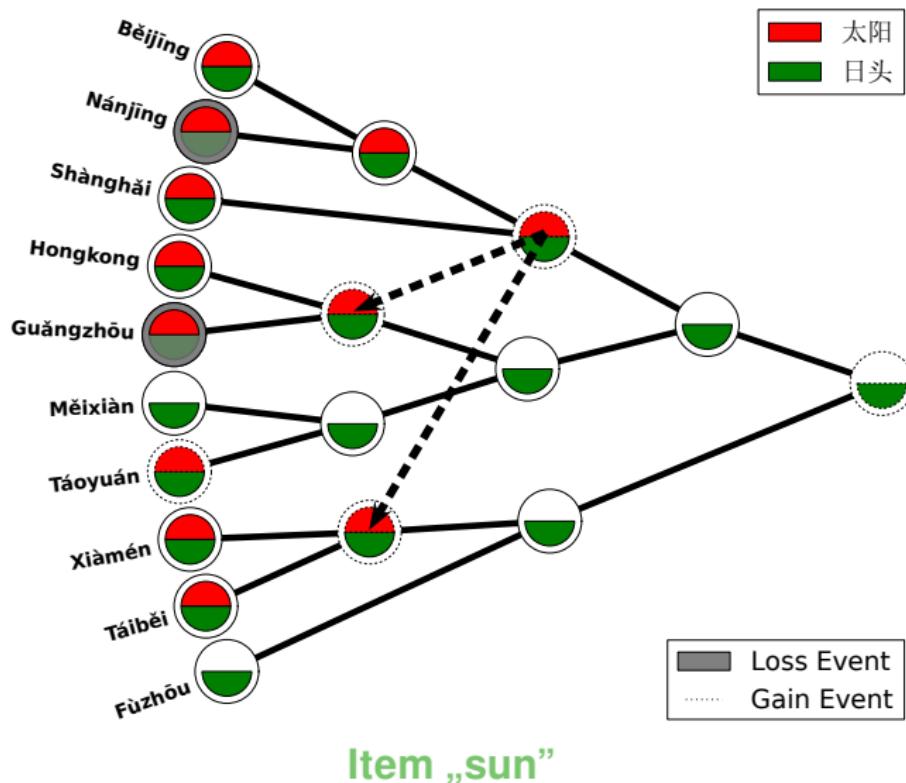
Application: Chinese Dialects (work in progress)



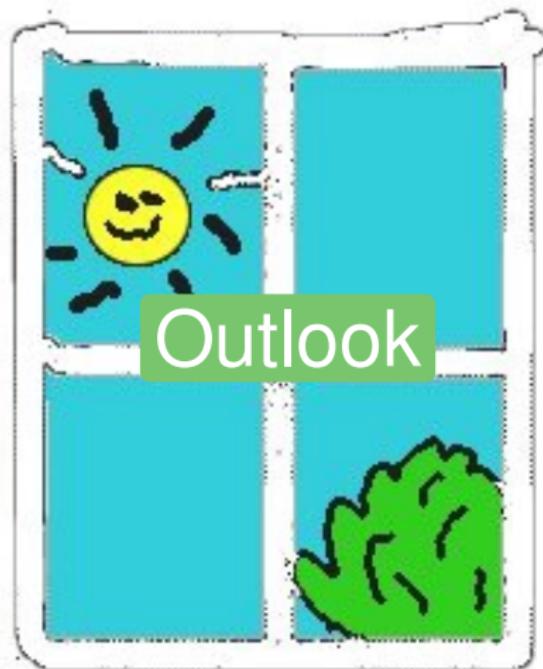
Application: Chinese Dialects (work in progress)



Application: Chinese Dialects (work in progress)



Outlook



- further test the MLN method on linguistic data

- further test the MLN method on linguistic data
- increase the transparency of the results in order to provide linguistic experts with a valid starting point for further not necessarily automatic research

- further test the MLN method on linguistic data
- increase the transparency of the results in order to provide linguistic experts with a valid starting point for further not necessarily automatic research
- improve the capability of the models: Similarly to gene fusion in biology, we have complex processes of compounding, regularly contributing to lexical change. Gain-loss models are not enough to deal with these cases of partial homology.

- further test the MLN method on linguistic data
- increase the transparency of the results in order to provide linguistic experts with a valuable point for further not necessarily automatic refinement
- improve the capability of the system. Similarly to gene fusion in biology, we have to find a way of combining, regularly contributing to the system. These cases are not enough to deal with these cases.

A simple cartoon illustration of a brown rabbit with long ears, sitting upright and facing forward. It has large, expressive eyes and a small smile. A green rectangular box is positioned in front of the rabbit, containing the text.

Thank You for listening!