

LingPy – A Python Library for Quantitative Tasks in Historical Linguistics

Johann-Mattis List*

*Institut für Romanistik II
Heinrich Heine University Düsseldorf

19. Juli 2012

Introductory Remarks

What the *** is historical
linguistics?

Introductory Remarks



Introductory Remarks

*Konservativ kommt
nicht von Konserve!*

Angela Merkel (10. April 2000)

Introductory Remarks

konservativ

Introductory Remarks

konservativ

Part of Speech	adjective
----------------	-----------

frequency	moderate
-----------	----------

Introductory Remarks

konservativ

Part of Speech	adjective
----------------	-----------

frequency	moderate
-----------	----------

meaning	“sticking to the traditional”
---------	-------------------------------

synonyms	“rückschrittlich”, “antiquiert”, “rückständig”, “unzeitgemäß”
----------	---

pronunciation	kɔnzɛrvati:f
---------------	--------------

Introductory Remarks

konservativ

Part of Speech	adjective
----------------	-----------

frequency	moderate
-----------	----------

meaning	“sticking to the traditional”
---------	-------------------------------

synonyms	“rückschrittlich”, “antiquiert”, “rückständig”, “unzeitgemäß”
----------	---

pronunciation	kɔnzɛrvati:f
---------------	--------------

origin	English <i>conservative</i>
--------	-----------------------------

Introductory Remarks

English *conservative*
“antiquated”

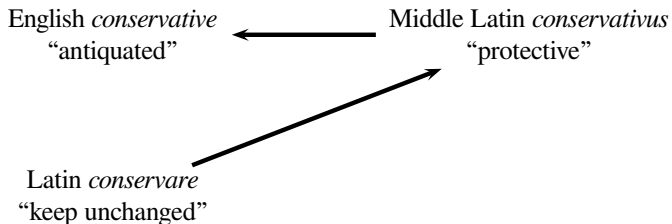
Introductory Remarks

English *conservative*
“antiquated”

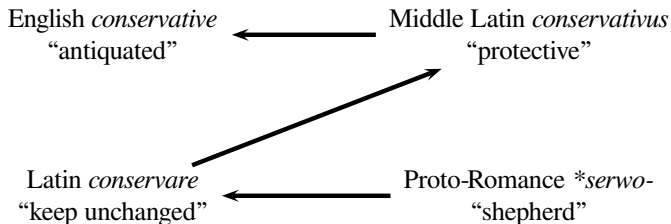


Middle Latin *conservativus*
“protective”

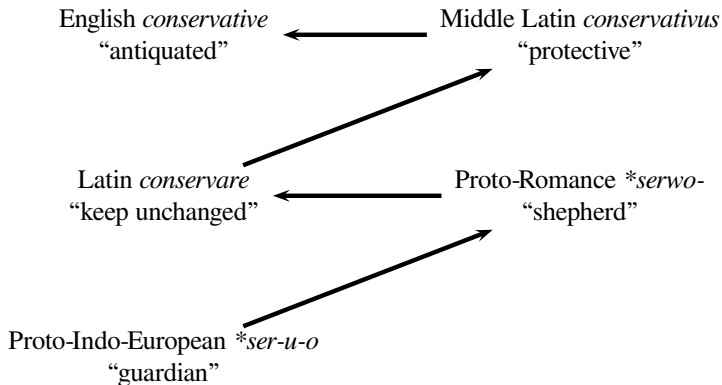
Introductory Remarks



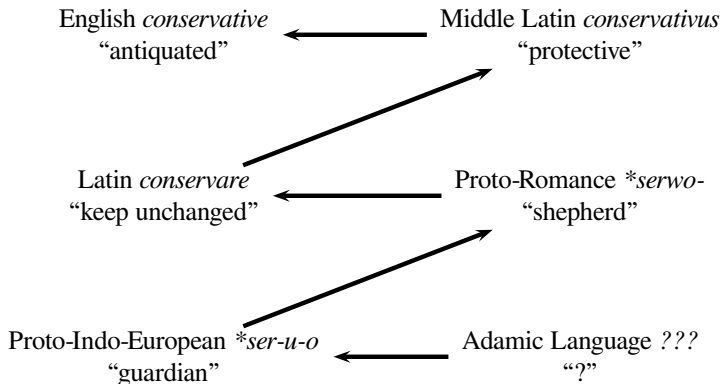
Introductory Remarks



Introductory Remarks



Introductory Remarks



Introductory Remarks

Konserve

Introductory Remarks

Konserve

Part of Speech	noun
----------------	------

frequency	moderate
-----------	----------

Introductory Remarks

Konserve

Part of Speech	noun
----------------	------

frequency	moderate
-----------	----------

meaning	“preserve tin”
---------	----------------

synonyms	“Büchse”, “Konservenbüchse”, “Konservendose”
----------	---

pronunciation	kɔnzɜrvə
---------------	----------

Introductory Remarks

Konserve

Part of Speech	noun
----------------	------

frequency	moderate
-----------	----------

meaning	“preserve tin”
---------	----------------

synonyms	“Büchse”, “Konservenbüchse”, “Konservendose”
----------	---

pronunciation	kɔnzɜrvə
---------------	----------

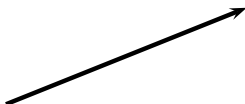
origin	Middle Latin <i>conserva</i>
--------	------------------------------

Introductory Remarks

Middle Latin *conserva*

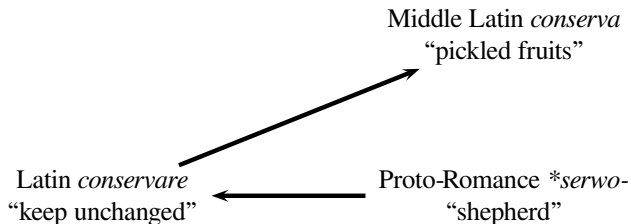
Introductory Remarks

Latin *conservare*
“keep unchanged”

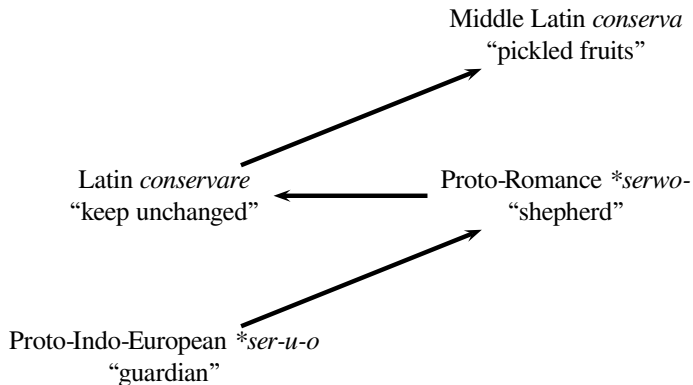


Middle Latin *conserva*
“pickled fruits”

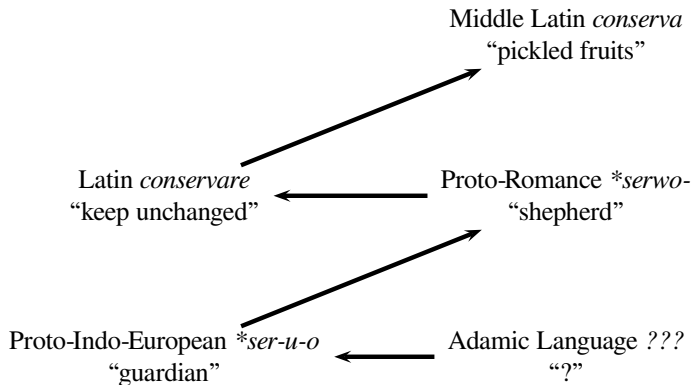
Introductory Remarks



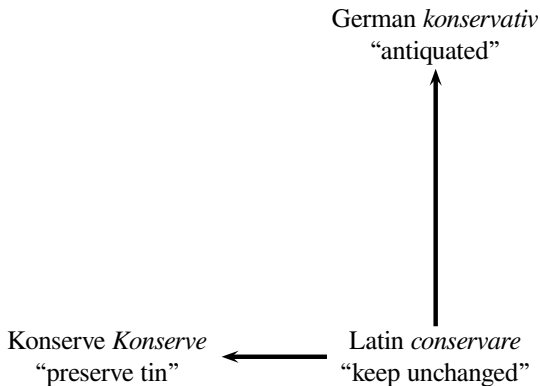
Introductory Remarks



Introductory Remarks



Introductory Remarks



Introductory Remarks

*Konservativ kommt
nicht von Konserve!*

Introductory Remarks

*Konservativ kommt
nicht von Konserve!*

Q. E. D.

Introductory Remarks

What the *** is historical
linguistics?

Introductory Remarks

What the *** is historical linguistics?

Historical linguistics is a (scientific) discipline that does not specifically care about what words mean, but where they come from...

Structure of the Talk

- 1 Basic Ideas
- 2 Sequence Modelling
- 3 Phonetic Alignment
- 4 Automatic Cognate Detection

What is LingPy?

What is LingPy?

- LingPy is a Python library (see <http://linguist.de/lingpy>) for automatic tasks in historical linguistics.

What is LingPy?

- LingPy is a Python library (see <http://linguist.de/lingpy>) for automatic tasks in historical linguistics.
- The current release of LingPy (lingpy-1.0) provides methods for sequence modelling, pairwise and multiple sequence alignment (SCA, List 2012a), automatic cognate detection (LexStat, List 2012b), and plotting routines (see the online documentation for details).

What is LingPy?

- LingPy is a Python library (see <http://linguist.de/lingpy>) for automatic tasks in historical linguistics.
- The current release of LingPy (lingpy-1.0) provides methods for sequence modelling, pairwise and multiple sequence alignment (SCA, List 2012a), automatic cognate detection (LexStat, List 2012b), and plotting routines (see the online documentation for details).
- LingPy can be invoked from the Python shell or inside Python scripts (examples are given in the online documentation).

Current Features

Current Features

- tokenize phonetic sequences

Current Features

- tokenize phonetic sequences
- convert phonetic sequences into abstract sound classes

Current Features

- tokenize phonetic sequences
- convert phonetic sequences into abstract sound classes
- calculate basic prosodic characteristics of phonetic sequences

Current Features

- tokenize phonetic sequences
- convert phonetic sequences into abstract sound classes
- calculate basic prosodic characteristics of phonetic sequences
- carry out pairwise alignment analyses of phonetic sequences

Current Features

- tokenize phonetic sequences
- convert phonetic sequences into abstract sound classes
- calculate basic prosodic characteristics of phonetic sequences
- carry out pairwise alignment analyses of phonetic sequences
- carry out multiple alignment analyses of phonetic sequences

Current Features

- tokenize phonetic sequences
- convert phonetic sequences into abstract sound classes
- calculate basic prosodic characteristics of phonetic sequences
- carry out pairwise alignment analyses of phonetic sequences
- carry out multiple alignment analyses of phonetic sequences
- automatically search for cognates and etymologically related words in multilingual word lists

Basic Ideas

Basic Ideas

- **adapt common techniques for sequence comparison and phylogenetic reconstruction to the specific needs of historical linguistics (general idea)**

Basic Ideas

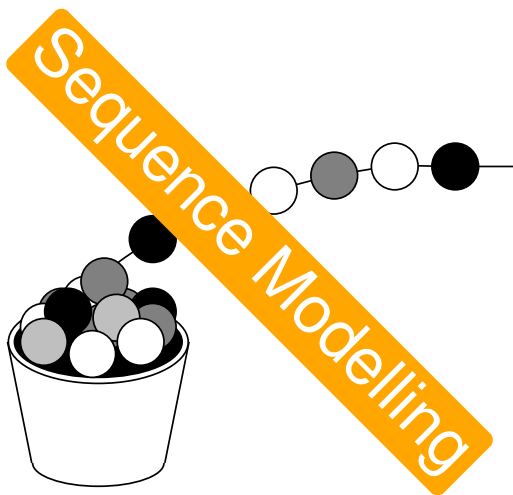
- **adapt common techniques for sequence comparison and phylogenetic reconstruction to the specific needs of historical linguistics (general idea)**
- construct realistic models of phonetic sequences by distinguishing external and internal representations (module `lingpy.sequence`)

Basic Ideas

- **adapt common techniques for sequence comparison and phylogenetic reconstruction to the specific needs of historical linguistics (general idea)**
- construct realistic models of phonetic sequences by distinguishing external and internal representations (module `lingpy.sequence`)
- compare sequences in a way that closely reflects common linguistic theory (module `lingpy.compare`)

Basic Ideas

- **adapt common techniques for sequence comparison and phylogenetic reconstruction to the specific needs of historical linguistics (general idea)**
- construct realistic models of phonetic sequences by distinguishing external and internal representations (module `lingpy.sequence`)
- compare sequences in a way that closely reflects common linguistic theory (module `lingpy.compare`)
- compare languages in a way that closely reflects the basic methods of historical linguistics (module `lingpy.lexstat`)



Excursus: Sequence Similarity

Excursus: Sequence Similarity

Synchronic Sequence Similarity

Sequences are judged to be similar if the segments of the sequences are phonetically similar ('phenotypic resemblance', Lass 1997).

Excursus: Sequence Similarity

Synchronic Sequence Similarity

Sequences are judged to be similar if the segments of the sequences are phonetically similar ('phenotypic resemblance', Lass 1997).

Diachronic Sequence Similarity

Sequences are judged to be similar if the segments of the sequences correspond *systematically* ('genotypic resemblance', Lass 1997).

Excursus: Sequence Similarity

Synchronic Sequence Similarity

Diachronic Sequence Similarity

Excursus: Sequence Similarity

Synchronic Sequence Similarity

Greek	mati	'eye'	\approx	Malay	mata	'eye'
Greek	θεος	'god'	\approx	Spanish	dios	'god'

Diachronic Sequence Similarity

Excursus: Sequence Similarity

Synchronic Sequence Similarity

Greek	mati	'eye'	≈	Malay	mata	'eye'
Greek	θεος	'god'	≈	Spanish	dios	'god'

Diachronic Sequence Similarity

German	ts ^h a:n	'tooth'	≈	English	tʊ:θ	'tooth'
Spanish	etʃo	'fact'	≈	French	fɛ	'fact'

Paradigmatic Aspects

Sound Classes

Paradigmatic Aspects

Sound Classes

Sounds which often occur in correspondence relations in genetically related languages can be clustered into classes (types). It is assumed “that phonetic correspondences inside a ‘type’ are more regular than those between different ‘types’” (Dolgopolsky 1986: 35).

Paradigmatic Aspects

Sound Classes

Sounds which often occur in correspondence relations in genetically related languages can be clustered into classes (types). It is assumed “that phonetic correspondences inside a ‘type’ are more regular than those between different ‘types’” (Dolgopolsky 1986: 35).

k

g

p

b

ŋ

ŋ

f

v

t

d

ʃ

ʒ

θ

ð

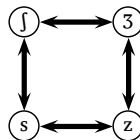
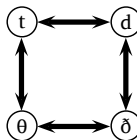
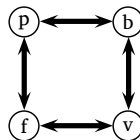
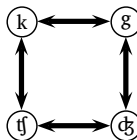
s

z

Paradigmatic Aspects

Sound Classes

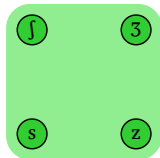
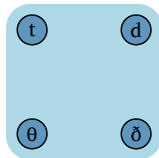
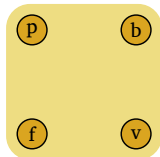
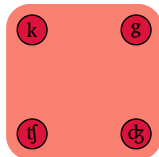
Sounds which often occur in correspondence relations in genetically related languages can be clustered into classes (types). It is assumed “that phonetic correspondences inside a ‘type’ are more regular than those between different ‘types’” (Dolgopolsky 1986: 35).



Paradigmatic Aspects

Sound Classes

Sounds which often occur in correspondence relations in genetically related languages can be clustered into classes (types). It is assumed “that phonetic correspondences inside a ‘type’ are more regular than those between different ‘types’” (Dolgopolsky 1986: 35).



Paradigmatic Aspects

Sound Classes

Sounds which often occur in correspondence relations in genetically related languages can be clustered into classes (types). It is assumed “that phonetic correspondences inside a ‘type’ are more regular than those between different ‘types’” (Dolgopolsky 1986: 35).

**K****P****T****S**

Syntagmatic Aspects

Syntagmatic Aspects

- sound change occurs more frequently in prosodically **weak** positions of phonetic sequences (Geisler 1992)

Syntagmatic Aspects

- sound change occurs more frequently in prosodically **weak** positions of phonetic sequences (Geisler 1992)
- given the sonority structure of a phonetic sequence, one can distinguish positions that differ regarding their **prosodic context**

Syntagmatic Aspects

- sound change occurs more frequently in prosodically **weak** positions of phonetic sequences (Geisler 1992)
- given the sonority structure of a phonetic sequence, one can distinguish positions that differ regarding their **prosodic context**
- prosodic context can be modelled by representing a sequence by a **prosodic string**, indicating the different prosodic contexts of each segment

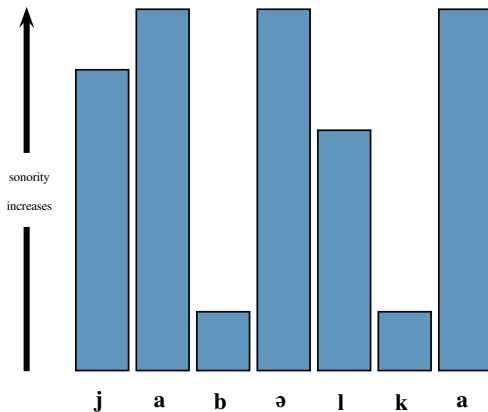
Syntagmatic Aspects

- sound change occurs more frequently in prosodically **weak** positions of phonetic sequences (Geisler 1992)
- given the sonority structure of a phonetic sequence, one can distinguish positions that differ regarding their **prosodic context**
- prosodic context can be modelled by representing a sequence by a **prosodic string**, indicating the different prosodic contexts of each segment
- based on the relative strength of all sites in a phonetic sequence, substitution scores and gap penalties can be modified when carrying out alignment analyses

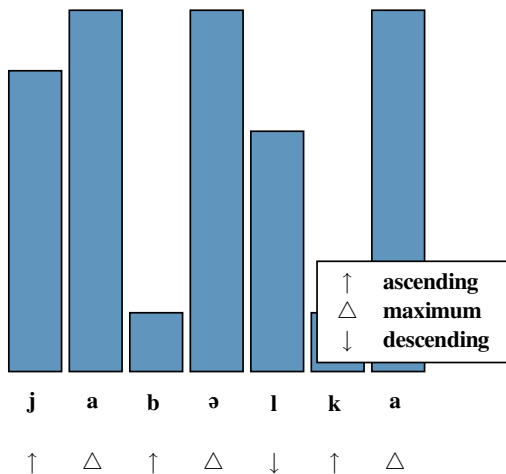
Syntagmatic Aspects

j a b ə l k a

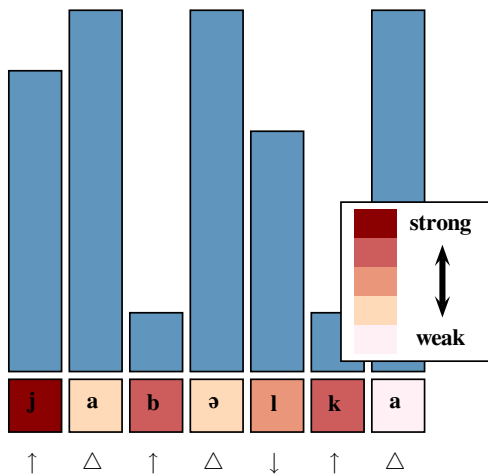
Syntagmatic Aspects



Syntagmatic Aspects



Syntagmatic Aspects



Examples

Examples

```
>>> from lingpy import *
```

Examples

```
>>> from lingpy import *  
>>> konservativ = Sequence("kɔnzɛrvati:f")
```

Examples

```
>>> from lingpy import *  
>>> konservativ = Sequence("kɔnzɛrvati:f")  
>>> print ' '.join(konservativ.tokens)
```

Examples

```
>>> from lingpy import *  
>>> konservativ = Sequence("kɔnzɛrvati:f")  
>>> print ' '.join(konservativ.tokens)  
k ɔ n z ɛ r v a t i: f
```


Examples

```
>>> from lingpy import *
>>> konservativ = Sequence("kɔnzɛrvati:f")
>>> print ' '.join(konservativ.tokens)
k ɔ n z ɛ r v a t i: f
>>> print konservativ.classes
```

Examples

```
>>> from lingpy import *
>>> konservativ = Sequence("kɔnzɛrvati:f")
>>> print ' '.join(konservativ.tokens)
k ɔ n z ɛ r v a t i: f
>>> print konservativ.classes
KUNSERBATIB
```

Examples

```
>>> from lingpy import *
>>> konservativ = Sequence("kɔnzɛrvati:f")
>>> print ' '.join(konservativ.tokens)
k ɔ n z ɛ r v a t i: f
>>> print konservativ.classes
KUNSERBATIB
>>> print ' '.join([str(i) for i in konservativ.sonar])
```

Examples

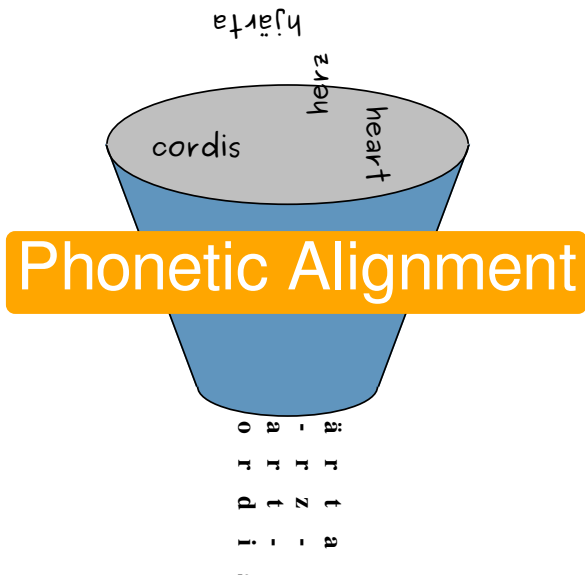
```
>>> from lingpy import *
>>> konservativ = Sequence("kɔnzɛrvati:f")
>>> print ' '.join(konservativ.tokens)
k ɔ n z ɛ r v a t i: f
>>> print konservativ.classes
KUNSERBATIB
>>> print ' '.join([str(i) for i in konservativ.sonar])
1 7 4 3 7 5 3 7 1 7 3
```

Examples

```
>>> from lingpy import *
>>> konservativ = Sequence("kɔnzɛrvati:f")
>>> print ' '.join(konservativ.tokens)
k ɔ n z ɛ r v a t i: f
>>> print konservativ.classes
KUNSERBATIB
>>> print ' '.join([str(i) for i in konservativ.sonar])
1 7 4 3 7 5 3 7 1 7 3
>>> print konservativ.prostring
```

Examples

```
>>> from lingpy import *
>>> konservativ = Sequence("kɔnzɛrvati:f")
>>> print ' '.join(konservativ.tokens)
k ɔ n z ɛ r v a t i: f
>>> print konservativ.classes
KUNSERBATIB
>>> print ' '.join([str(i) for i in konservativ.sonar])
1 7 4 3 7 5 3 7 1 7 3
>>> print konservativ.prostring
#VcCvcCvCv$
```



Excursus: Alignment Analyses

Alignment Analyses

Excursus: Alignment Analyses

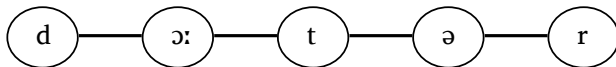
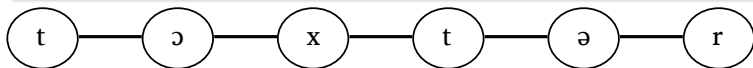
Alignment Analyses

In alignment analyses, sequences are arranged in a matrix in such a way that corresponding elements occur in the same column, while empty cells resulting from non-corresponding elements are filled with gap symbols.

Excursus: Alignment Analyses

Alignment Analyses

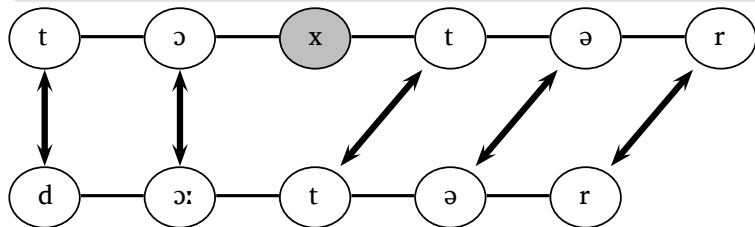
In alignment analyses, sequences are arranged in a matrix in such a way that corresponding elements occur in the same column, while empty cells resulting from non-corresponding elements are filled with gap symbols.



Excursus: Alignment Analyses

Alignment Analyses

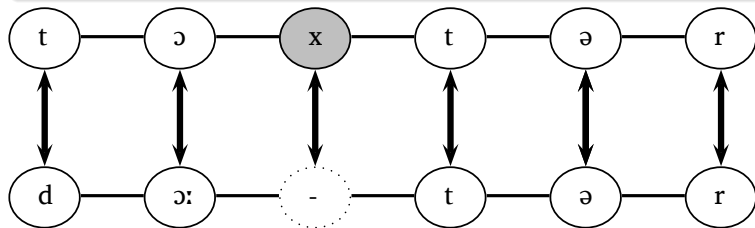
In alignment analyses, sequences are arranged in a matrix in such a way that corresponding elements occur in the same column, while empty cells resulting from non-corresponding elements are filled with gap symbols.



Excursus: Alignment Analyses

Alignment Analyses

In alignment analyses, sequences are arranged in a matrix in such a way that corresponding elements occur in the same column, while empty cells resulting from non-corresponding elements are filled with gap symbols.



Phonetic Alignment

Sound-Class Based Alignment (SCA, List 2012a)

Phonetic Alignment

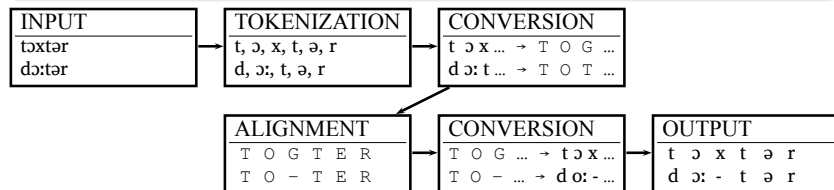
Sound-Class Based Alignment (SCA, List 2012a)

Sound classes and alignment analyses can be easily combined by representing phonetic sequences internally as sound classes and comparing the sound classes with traditional alignment algorithms.

Phonetic Alignment

Sound-Class Based Alignment (SCA, List 2012a)

Sound classes and alignment analyses can be easily combined by representing phonetic sequences internally as sound classes and comparing the sound classes with traditional alignment algorithms.



Example

```
>>> from lingpy import *
```


Example

```
>>> from lingpy import *  
>>> konservativ = "kɔnzɛrvati:f"
```

Example

```
>>> from lingpy import *  
>>> konservativ = "kɔnzɛrvati:f"  
>>> Konserve = "kɔnzɛrvə"
```

Example

```
>>> from lingpy import *  
>>> konservativ = "kɔnzɛrvati:f"  
>>> Konserve = "kɔnzɛrvə"  
>>> pair = Pairwise(konservativ, Konserve)
```

Example

```
>>> from lingpy import *  
>>> konservativ = "kɔnzɛrvati:f"  
>>> Konserve = "kɔnzɛrvə"  
>>> pair = Pairwise(konservativ, Konserve)  
>>> pair.align()
```

Example

```
>>> from lingpy import *  
>>> konservativ = "kɔnzɛrvati:f"  
>>> Konserve = "kɔnzɛrvə"  
>>> pair = Pairwise(konservativ, Konserve)  
>>> pair.align()  
>>> pair.distance()
```

Example

```
>>> from lingpy import *
>>> konservativ = "kɔnzɛrvati:f"
>>> Konserve = "kɔnzɛrvə"
>>> pair = Pairwise(konservativ, Konserve)
>>> pair.align()
>>> pair.distance()
>>> print pair
```

Example

```
>>> from lingpy import *
>>> konservativ = "kɔnzɛrvati:f"
>>> Konserve = "kɔnzɛrvə"
>>> pair = Pairwise(konservativ,Konserve)
>>> pair.align()
>>> pair.distance()
>>> print pair
k ɔ n z ɛ r v a t i: f
k ɔ n z ɛ r v - - ə -
0.3
```

Example

Taxon	Alignment											
Xiangtan	-	i	-	24	d	əu	12	-	-	-	-	-
Qingdao	-	i	-	42	t ^h	ou	-	-	-	-	-	-
Xi'an	-	ə	r	21	t ^h	ou	-	-	-	-	-	-
Wuhan	-	u	-	213	t ^h	eu	-	-	-	-	-	-
Guangzhou	j	i	t	2	t ^h	eu	21	-	-	-	-	-
Nanning	j	i	t	22	t ^h	eu	21	-	-	-	-	-
Xianggang	j	e	t	2	t ^h	eu	21	-	-	-	-	-
Nanning	j	e	t	22	t ^h	eu	21	-	-	-	-	-
Taibei	l	i	t	44	t ^h	au	24	-	-	-	-	-
Xiamen	l	i	t	5	t ^h	au	-	-	-	-	-	-
Jian'ou	m	i	-	33	t ^h	e	33	-	-	-	-	-
Shexian	n	i	-	22	t ^h	iu	44	-	-	-	-	-
Fuzhou	n	i	?	5	t ^h	au	53	-	-	-	-	-
Shantou	z	i	k	5	-	-	-	-	-	-	-	-
Shantou	z	i	k	5	t ^h	au	55	-	-	-	-	-
Haikou	z	i	t	3	-	-	-	-	-	-	-	-
Haikou	z	i	t	3	h	au	31	-	-	-	-	-
Taoyuan	ŋ	i	t	22	t ^h	eu	11	-	-	-	-	-
Meixian	ŋ	i	t	1	t ^h	eu	11	-	-	-	-	-
Wenzhou	n	i	-	213	d	su	-	-	-	-	-	-
Wenzhou	n	i	-	213	d	xu	31	v	ai	213	-	-
Nanchang	n	i	?	5	t ^h	u	62	-	-	-	-	-
Tunxi	n	ie	-	11	t ^h	iu	44	-	-	-	-	-
Suzhou	n	iə	?	3	d	ɤ	13	-	-	-	-	-
Shanghai	n	i	?	1	d	ɤ	13	-	-	-	-	-
Kunming	ʒ	ə	-	31	t ^h	əu	31	-	-	-	-	-
Hefei	ʒ	ə	?	5	t ^h	u	-	-	-	-	-	-
Xining	ʒ	ɛ	-	44	t ^h	u	24	-	ɛ	24	-	-
Jinan	ʒ	ɿ	-	21	t ^h	ou	-	-	-	-	-	-
Changsha	ʒ	ɿ	-	24	t	eu	-	-	-	-	-	-
Zhengzhou	ʒ	ɿ	-	24	t ^h	ou	-	-	-	-	-	-
Lanzhou	ʒ	ɿ	-	13	t ^h	ou	13	-	-	-	-	-
Yinchuan	ʒ	ɿ	-	13	t ^h	eu	-	-	-	-	-	-
Haerbin	ʒ	ɿ	-	53	t ^h	ou	-	-	-	-	-	-
Beijing	ʒ	ɿ	-	51	t ^h	ou	1	-	-	-	-	-
Huhehaote	ʒ	ɿ	-	55	t ^h	əu	31	-	-	-	-	-
Pingyao	ʒ	ʌ	?	53	t	eu	13	-	in	13	-	-



Automatic Cognate Detection

Excursus: Language-Specific Similarities

Language-Specific Similarity Measure

Excursus: Language-Specific Similarities

Language-Specific Similarity Measure

- recall the aforementioned distinction between **synchronic** and **diachronic** similarity

Excursus: Language-Specific Similarities

Language-Specific Similarity Measure

- recall the aforementioned distinction between **synchronic** and **diachronic** similarity
- **diachronic** similarity is the basic of all traditional approaches to cognate detection

Excursus: Language-Specific Similarities

Language-Specific Similarity Measure

- recall the aforementioned distinction between **synchronic** and **diachronic** similarity
- **diachronic** similarity is the basic of all traditional approaches to cognate detection
- diachronic sequence similarity is determined on the basis of systematic sound correspondences as opposed to similarity based on surface resemblances of phonetic segments

Excursus: Language-Specific Similarities

Language-Specific Similarity Measure

- recall the aforementioned distinction between **synchronic** and **diachronic** similarity
- **diachronic** similarity is the basic of all traditional approaches to cognate detection
- diachronic sequence similarity is determined on the basis of systematic sound correspondences as opposed to similarity based on surface resemblances of phonetic segments
- the most crucial aspect of correspondence-based (diachronic) similarity is that it is **language-specific**

Excursus: Language-Specific Similarities

Language-Specific Similarity Measure

- recall the aforementioned distinction between **synchronic** and **diachronic** similarity
- **diachronic** similarity is the basic of all traditional approaches to cognate detection
- diachronic sequence similarity is determined on the basis of systematic sound correspondences as opposed to similarity based on surface resemblances of phonetic segments
- the most crucial aspect of correspondence-based (diachronic) similarity is that it is **language-specific**

Meaning	German	Dutch	English
“tooth”	<i>Zahn</i> [ts a:n]	<i>tand</i> [t ant]	<i>tooth</i> [t u:θ]
“ten”	<i>zehn</i> [ts e:n]	<i>tien</i> [t i:n]	<i>ten</i> [t ɛn]
“tongue”	<i>Zunge</i> [ts ʊŋə]	<i>tong</i> [t ɔŋ]	<i>tongue</i> [t ʌŋ]

Excursus: Language-Specific Similarities

Language-Specific Similarity Measure

- recall the aforementioned distinction between **synchronic** and **diachronic** similarity
- **diachronic** similarity is the basic of all traditional approaches to cognate detection
- diachronic sequence similarity is determined on the basis of systematic sound correspondences as opposed to similarity based on surface resemblances of phonetic segments
- the most crucial aspect of correspondence-based (diachronic) similarity is that it is **language-specific**

Meaning	Shanghai	Beijing	Guangzhou
“nine”	[tɕ i ³⁵]	Beijing [tɕ iou ²¹⁴]	[k eu ³⁵]
“today”	[tɕ iŋ ⁵⁵ tsɔ ²¹]	Beijing [tɕ iə ⁵⁵]	[k em ⁵³ jet ²]
“rooster”	[koŋ ⁵⁵ tɕ i ²¹]	Beijing [kuŋ ⁵⁵ tɕ i ⁵⁵]	[k ei ⁵⁵ koŋ ⁵⁵]

Automatic Cognate Detection

LexStat (List 2012b)

Automatic Cognate Detection

LexStat (List 2012b)

- LexStat is a method for the automatic detection of cognates in multilingual word lists

Automatic Cognate Detection

LexStat (List 2012b)

- LexStat is a method for the automatic detection of cognates in multilingual word lists
- LexStat based its cognate judgments on language-specific, diachronic similarity

Automatic Cognate Detection

LexStat (List 2012b)

- LexStat is a method for the automatic detection of cognates in multilingual word lists
- LexStat based its cognate judgments on language-specific, diachronic similarity
- LexStat combines the most recent approaches to sequence comparison in computer science and biology with a novel approach to sequence modelling

Automatic Cognate Detection

LexStat (List 2012b)

- LexStat is a method for the automatic detection of cognates in multilingual word lists
- LexStat based its cognate judgments on language-specific, diachronic similarity
- LexStat combines the most recent approaches to sequence comparison in computer science and biology with a novel approach to sequence modelling
- LexStat largely outperforms alternative approaches and yields 90% precision and 84% recall on very divergent language families, such as Indo-European and Austronesian

Automatic Cognate Detection

LexStat (List 2012b)

- LexStat is a method for the automatic detection of cognates in multilingual word lists
- LexStat based its cognate judgments on language-specific, diachronic similarity
- LexStat combines the most recent approaches to sequence comparison in computer science and biology with a novel approach to sequence modelling
- LexStat largely outperforms alternative approaches and yields 90% precision and 84% recall on very divergent language families, such as Indo-European and Austronesian
- LexStat yields **transparent** decisions which can be directly examined by the researcher

Automatic Cognate Detection

ID	Items	German	English	Swedish
1	hand	hant	hænd	hand
2	woman	fraʊ	wʊmən	kvina
3	know	kɛnən	nəʊ	çɛna
3	know	visən	-	ve:ta
...

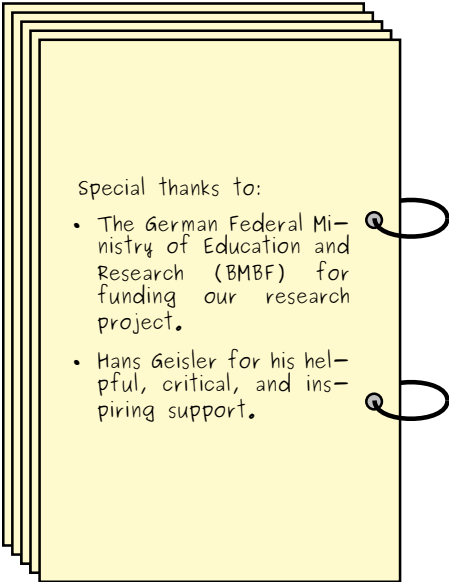
Automatic Cognate Detection

ID	Items	German	COG	English	COG	Swedish	COG
1	hand	hant	1	hænd	1	hand	1
2	woman	fraʊ	2	wʊmən	3	kvina	4
3	know	kɛnən	5	nəʊ	5	çɛna	5
3	know	visən	6	-	0	ve:ta	6
...

Automatic Cognate Detection

Basic Concept: <i>belly</i> (ID: 4)			
CogID	Language	Entry	Aligned Entry
6	Danish	ɔnʌliwʔ	--
7	German	baux	b au x
7	Dutch	bæyk	b æy k
7	Swedish	buk	b u k
7	Norwegian	bʉ:k	b ʉ: k
8	English	bɛlɪ	--
9	Swedish	ma:ge	m a: g e
9	Norwegian	mɑ:gə	m ɑ: g ə
9	Danish	mæ:və	m æ: v ə
10	Icelandic	kʰvɪ:ðyr	--





Special thanks to:

- The German Federal Ministry of Education and Research (BMBF) for funding our research project.
- Hans Geisler for his helpful, critical, and inspiring support.

