Multiple Sequence Alignment in Historical Linguistics

A Sound Class Based Approach

Johann-Mattis List[*]

[*] Institute for Romance Languages and Literature
Heinrich Heine University Düsseldorf

2011/06/25

Structure of the Talk

Sequences

Alignment Analyses

- Pairwise Sequence Alignment
- Multiple Sequence Alignment

Sequence Comparison in Historical Linguistics

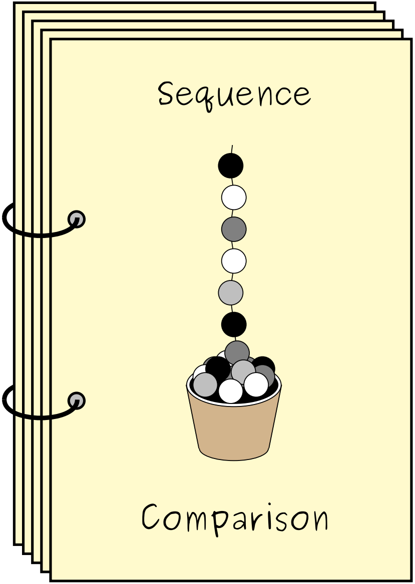
- Sequence Similarity
- Sound Classes

A New Method for Multiple Sequence Alignment

- Main Ideas
- Working Procedure
- Implementation

Performance of the Method

- Evaluation
- Results


Sequence
Comparison

Sequences

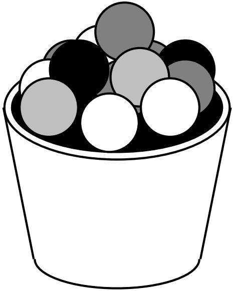
Definition 1

Sequences

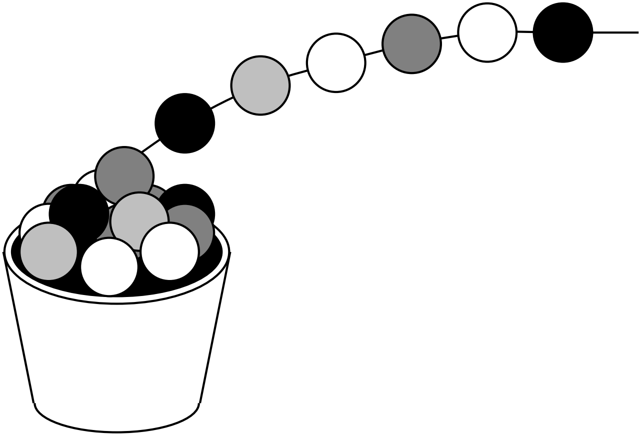
Definition 1

Given an *alphabet* (a non-empty finite set, whose elements are called *characters*), a *sequence* is an ordered list of characters drawn from the alphabet. The elements of sequences are called *segments*. (cf. Böckenbauer & Bongartz 2003: 30f)

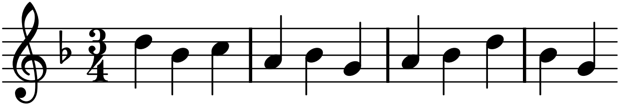
Sequences



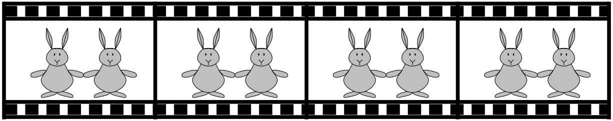
Sequences



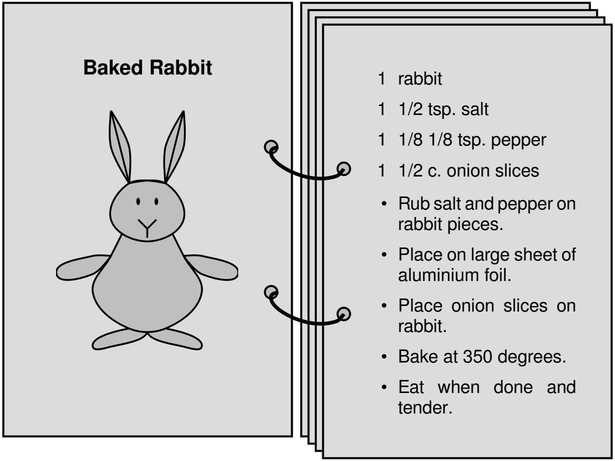
Sequences



Sequences



Sequences



Baked Rabbit

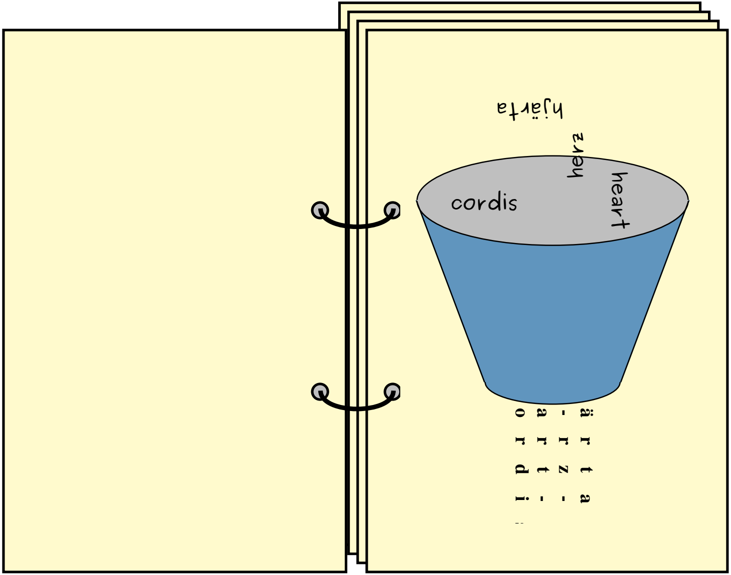
1 rabbit

1 1/2 tsp. salt

1 1/8 1/8 tsp. pepper

1 1/2 c. onion slices

- Rub salt and pepper on rabbit pieces.
- Place on large sheet of aluminium foil.
- Place onion slices on rabbit.
- Bake at 350 degrees.
- Eat when done and tender.


hjärta
herz
cordis
heart
ä r t a
- r z -
a r t -
o r d i

Alignment Analyses

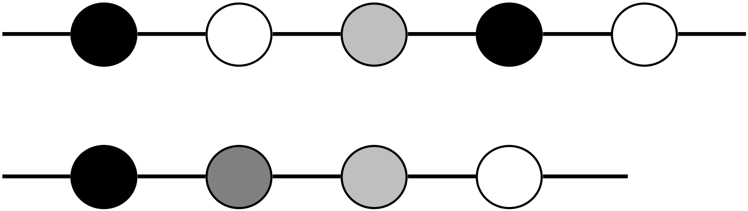
Definition 2

Alignment Analyses

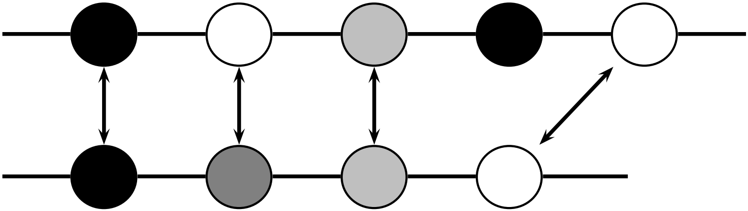
Definition 2

An *alignment* of two sequences *s* and *t* is a two-row matrix in which both sequences are aranged in such a way that all matching and mismatching segments occur in the same column, while empty cells, resulting from empty matches, are filled with gap symbols. (cf. Kruskal 1983)

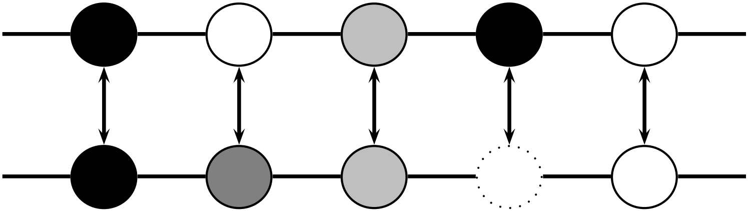
Alignment Analyses



Alignment Analyses



Alignment Analyses



Pairwise Sequence Alignment

Pairwise Sequence Alignment

→ Construct a matrix in which all segments of two sequences are confronted with each other and with gap characters.

Pairwise Sequence Alignment

- → Construct a matrix in which all segments of two sequences are confronted with each other and with gap characters.
- → Calculate the score of all subsequences recursively by filling the matrix from left to right and from top to bottom.

Pairwise Sequence Alignment

- → Construct a matrix in which all segments of two sequences are confronted with each other and with gap characters.
- → Calculate the score of all subsequences recursively by filling the matrix from left to right and from top to bottom.
- → Employ a scoring function in each recursion step, which evaluates, whether the characters in each cell should be matched with themselves or with gap characters.

Pairwise Sequence Alignment

- → Construct a matrix in which all segments of two sequences are confronted with each other and with gap characters.
- → Calculate the score of all subsequences recursively by filling the matrix from left to right and from top to bottom.
- → Employ a scoring function in each recursion step, which evaluates, whether the characters in each cell should be matched with themselves or with gap characters.
- → Retrieve the alignment by applying a traceback function which reconstructs the 'path of choices' (Durbin 2002) which led to the final value.

Pairwise Sequence Alignment

T	E	S	T	-	-	-	-
-	-	-	-	T	E	S	T

8

Pairwise Sequence Alignment

T	E	S	T	-	-	-
-	-	-	T	E	S	T

6

Pairwise Sequence Alignment

T	E	S	-	T	-	-	-
-	-	-	T	-	E	S	T

8

Pairwise Sequence Alignment

T	E	S	T	-	-	-
-	-	T	-	E	S	T

7

Pairwise Sequence Alignment

T	E	-	S	T	-	-	-
-	-	T	-	-	E	S	T

8

Pairwise Sequence Alignment

T	E	S	T	-	-	-
-	T	-	-	E	S	T

7

Pairwise Sequence Alignment

T	-	E	S	T	-	-	-
-	T	-	-	-	E	S	T

8

Pairwise Sequence Alignment

T	E	S	T	-	-	-
T	-	-	-	E	S	T

6

Pairwise Sequence Alignment

T	E	S	T	-	-
T	-	-	E	S	T

5

Pairwise Sequence Alignment

T	E	S	-	T	-	-
T	-	-	E	-	S	T

6

Pairwise Sequence Alignment

T	E	S	T	-	-
T	-	E	-	S	T

5

Pairwise Sequence Alignment

T	E	-	S	T	-	-
T	-	E	-	-	S	T

6

Pairwise Sequence Alignment

T	E	S	T	-	-
T	E	-	-	S	T

4

Pairwise Sequence Alignment

T	E	S	T	-
T	E	-	S	T

3

Pairwise Sequence Alignment

T	E	-	S	T	-
T	E	S	-	-	T

4

Pairwise Sequence Alignment

T	E	S	T	-
T	E	S	-	T

2

Pairwise Sequence Alignment

T	E	S	T
T	E	S	T

0

Multiple Sequence Alignment

Multiple Sequence Alignment

→ Align all sequences pairwise and store the scores in a matrix.

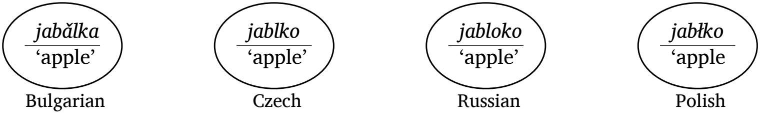
Multiple Sequence Alignment

- Align all sequences pairwise and store the scores in a matrix.
- Construct a guide tree from the matrix.

Multiple Sequence Alignment

- → Align all sequences pairwise and store the scores in a matrix.
- → Construct a guide tree from the matrix.
- → Align all sequences along the guide tree, going from its leaves to its root.

Multiple Sequence Alignment



jabălka ‘apple’
Bulgarian

jablko ‘apple’
Czech

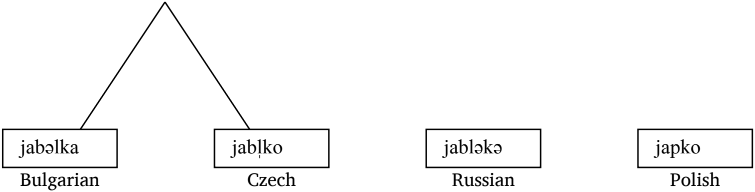
jabloko ‘apple’
Russian

jabłko ‘apple
Polish

Multiple Sequence Alignment

jabəlka	jabl̩ko	jabləkə	japko
Bulgarian	Czech	Russian	Polish

Multiple Sequence Alignment



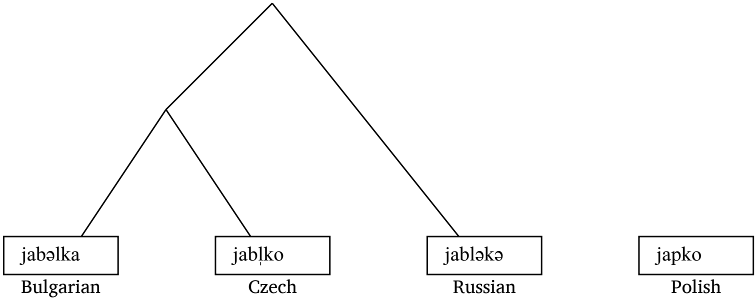
jabəlka
Bulgarian

jabl̩ko
Czech

jabləkə
Russian

japko
Polish

Multiple Sequence Alignment



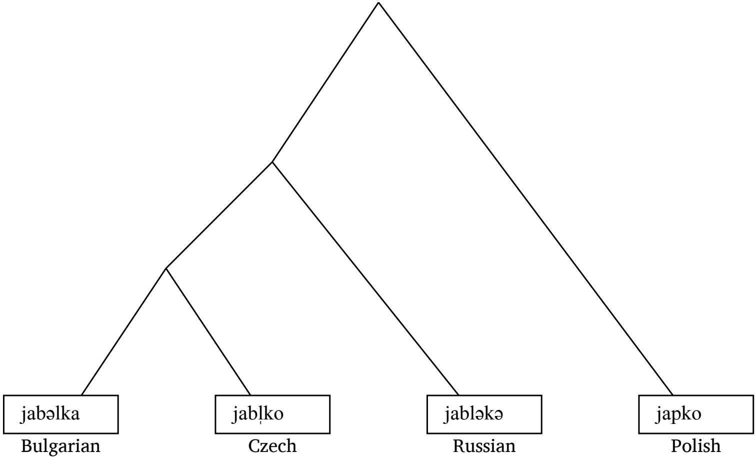
jabəlka
Bulgarian

jabl̩ko
Czech

jabləkə
Russian

japko
Polish

Multiple Sequence Alignment



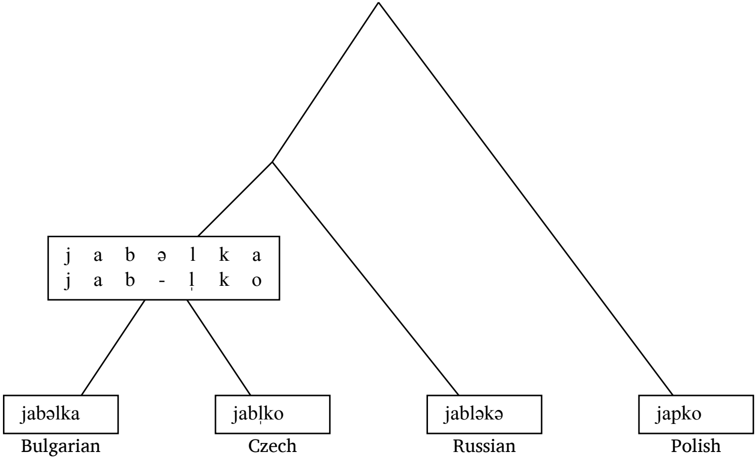
jabəlka
Bulgarian

jabl̩ko
Czech

jabləkə
Russian

japko
Polish

Multiple Sequence Alignment



j	a	b	ə	l	k	a
j	a	b	-	l̩	k	o

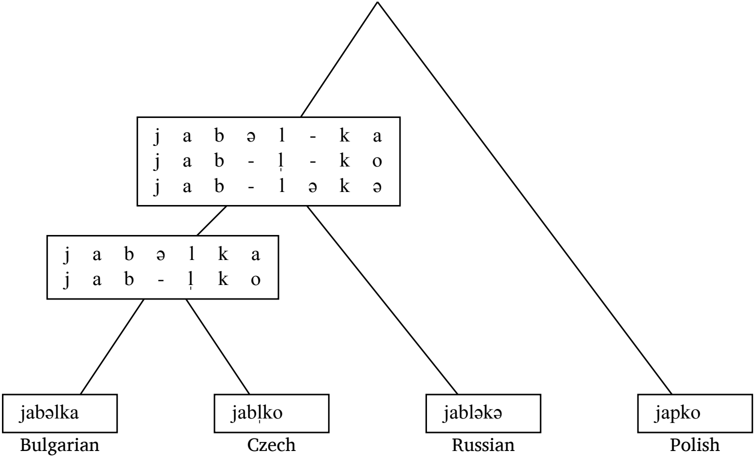
jabəlka
Bulgarian

jabl̩ko
Czech

jabləkə
Russian

japko
Polish

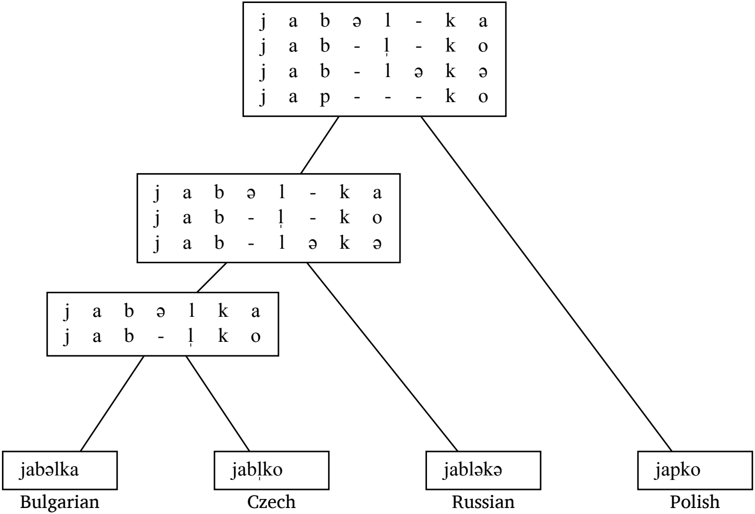
Multiple Sequence Alignment



j	a	b	ə	l	-	k	a
j	a	b	-	l̩	-	k	o
j	a	b	-	l	ə	k	ə

j	a	b	ə	l	k	a
j	a	b	-	l̩	k	o

Multiple Sequence Alignment

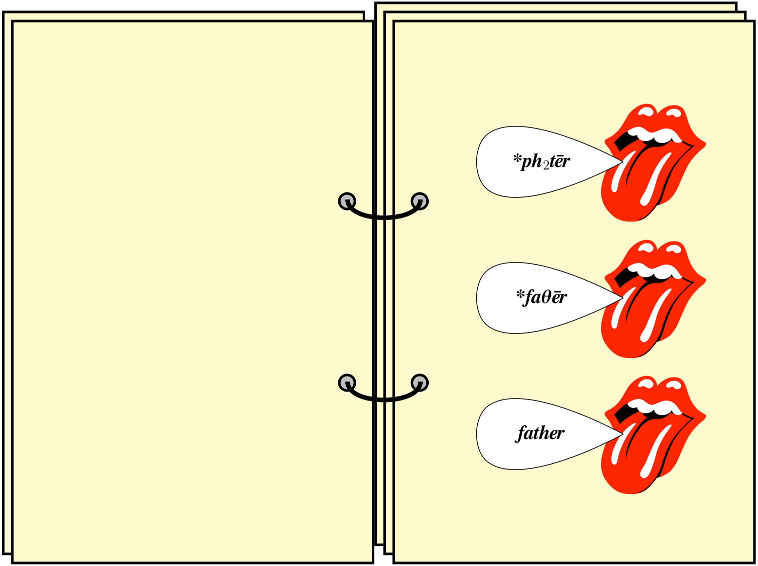


j	a	b	ə	l	-	k	a
j	a	b	-	l̩	-	k	o
j	a	b	-	l	ə	k	ə
j	a	p	-	-	-	k	o

j	a	b	ə	l	-	k	a
j	a	b	-	l̩	-	k	o
j	a	b	-	l	ə	k	ə

j	a	b	ə	l	k	a
j	a	b	-	l̩	k	o

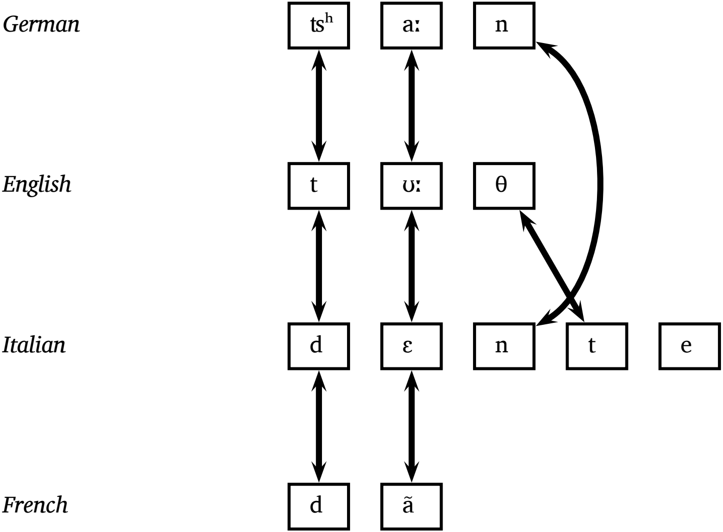
jabəlka	jabl̩ko	jabləkə	japko
Bulgarian	Czech	Russian	Polish


*ph$_2$tēr
*faθēr
father

Sequence Comparison in Historical Linguistics

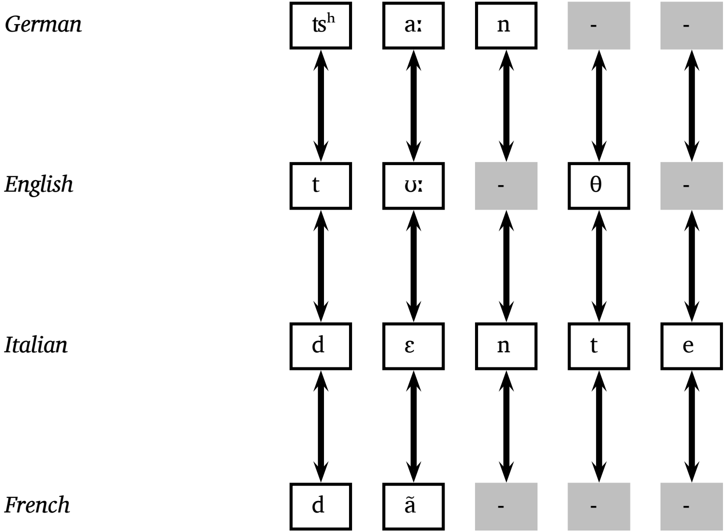
German	ʦʰ	aː	n		
English	t	ʊː	θ		
Italian	d	ɛ	n	t	e
French	d	ɑ̃			

Sequence Comparison in Historical Linguistics



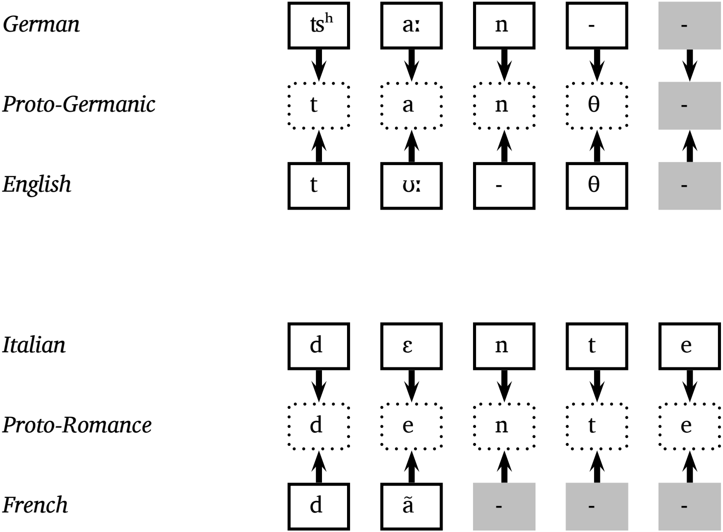
German	ʦ^{h}	aː	n		
English	t	ʊː	θ		
Italian	d	ɛ	n	t	e
French	d	ɑ̃			

Sequence Comparison in Historical Linguistics



German	ʦʰ	aː	n	-	-
English	t	ʊː	-	θ	-
Italian	d	ɛ	n	t	e
French	d	ã	-	-	-

Sequence Comparison in Historical Linguistics



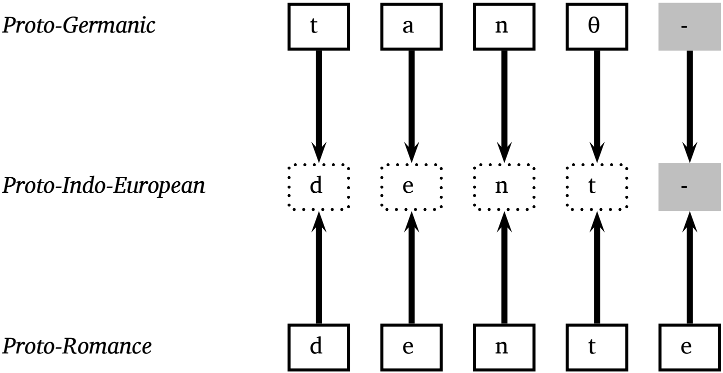
German	tsʰ	aː	n	-	-
Proto-Germanic	t	a	n	θ	-
English	t	ʊː	-	θ	-

Italian	d	ɛ	n	t	e
Proto-Romance	d	e	n	t	e
French	d	ã	-	-	-

Sequence Comparison in Historical Linguistics

Proto-Germanic	t	a	n	θ	-
Proto-Romance	d	e	n	t	e

Sequence Comparison in Historical Linguistics

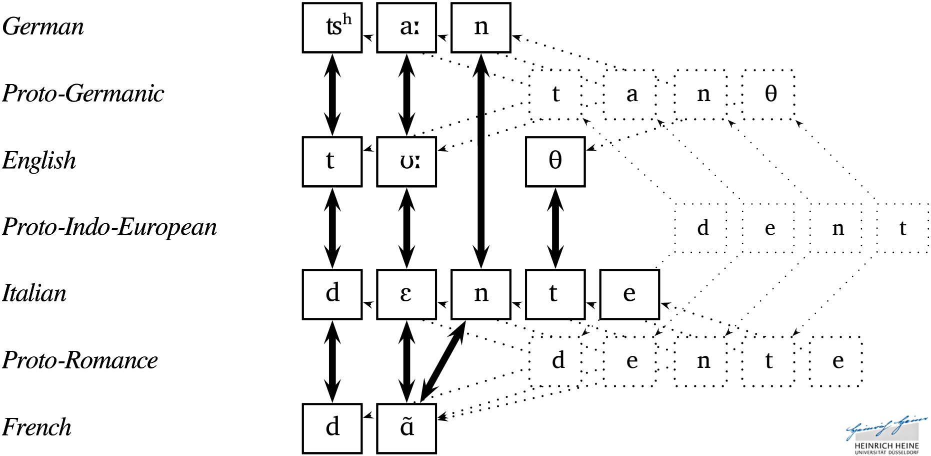


Proto-Germanic	t	a	n	θ	-
Proto-Indo-European	d	e	n	t	-
Proto-Romance	d	e	n	t	e

Sequence Comparison in Historical Linguistics

Proto-Indo-European	d	e	n	t

Sequence Comparison in Historical Linguistics



German tsʰ aː n

Proto-Germanic t a n θ

English t ʊː θ

Proto-Indo-European d e n t

Italian d ɛ n t e

Proto-Romance d e n t e

French d ɑ̃

Sequence Similarity

Synchronic Sequence Similarity

Diachronic Sequence Similarity

Sequence Similarity

Synchronic Sequence Similarity

Sequences are judged to be similar if the segments of the sequences are phonetically similar (‘phenotypic resemblence’, Lass 1997).

Diachronic Sequence Similarity

Sequence Similarity

Synchronic Sequence Similarity

Sequences are judged to be similar if the segments of the sequences are phonetically similar (‘phenotypic resemblence’, Lass 1997).

Diachronic Sequence Similarity

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

Sequence Similarity

Synchronic Sequence Similarity

Diachronic Sequence Similarity

Sequence Similarity

Synchronic Sequence Similarity

Greek	mati	‘eye’	≈	Malay	mata	‘eye’
Greek	θεɔs	‘god’	≈	Spanish	diɔs	‘god’

Diachronic Sequence Similarity

Sequence Similarity

Synchronic Sequence Similarity

Greek	mati	‘eye’	≈	Malay	mata	‘eye’
Greek	θɛɔs	‘god’	≈	Spanish	diɔs	‘god’

Diachronic Sequence Similarity

German	ʦʰaːn	‘tooth’	≈	English	tʊːθ	‘tooth’
Spanish	eʧo	‘fact’	≈	French	fɛ	‘fact’

Sound Classes

Sound Classes

→ ‘[Even] the most divergent languages show examples of phonetic change which are remarkably similar’ (Arlotto 1972: 77).

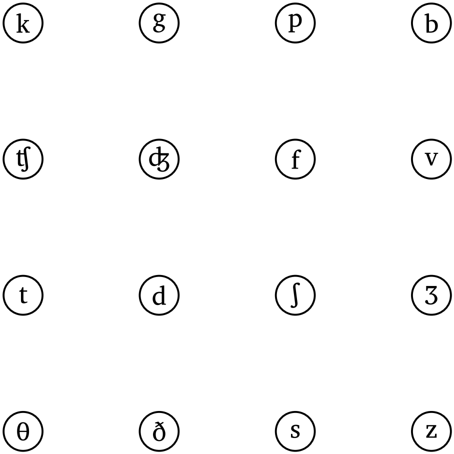
Sound Classes

- ‘[Even] the most divergent languages show examples of phonetic change which are remarkably similar’ (Arlotto 1972: 77).
- Sounds which often occur in correspondence relations in genetically related languages can be clustered into *classes* (cf. Dolgopolsky 1986, Burlak & Starostin 2005).

Sound Classes

- ‘[Even] the most divergent languages show examples of phonetic change which are remarkably similar’ (Arlotto 1972: 77).
- Sounds which often occur in correspondence relations in genetically related languages can be clustered into *classes* (cf. Dolgopolsky 1986, Burlak & Starostin 2005).
- In contrast to the pure notion of synchronic and diachronic similarity, *sound classes* incorporate phonetic detail **and** systematic correspondence patterns within a probabilistic framework.

Sound Classes



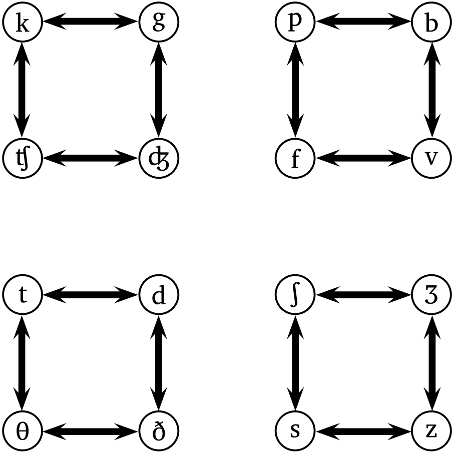
k g p b

ʧ ʤ f v

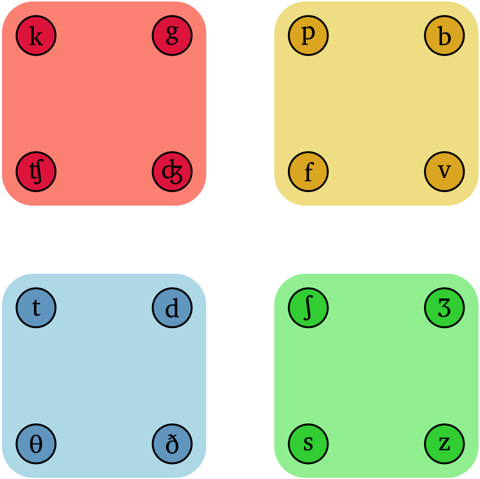
t d ʃ ʒ

θ ð s z

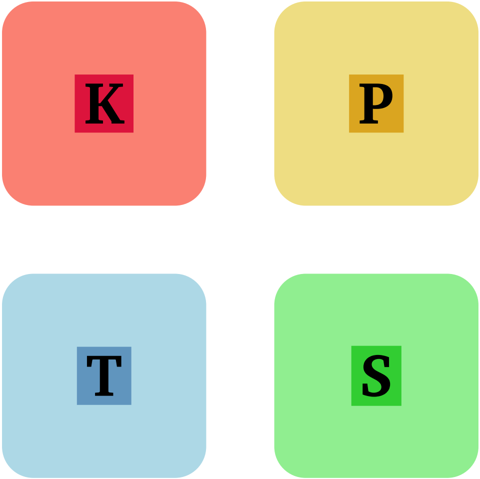
Sound Classes



Sound Classes



Sound Classes



K

P

T

S

Current Sound Class Approaches

Dolgopolsky (1986)

Holman et al. (2011)

Current Sound Class Approaches

Dolgopolsky (1986)

Based on an empirical basis, speech sounds are divided into ten types, distinguished ‘in such a way that phonetic correspondences inside a “type” are more regular than those between different types’ (Dolgopolsky 1986: 35).

Holman et al. (2011)

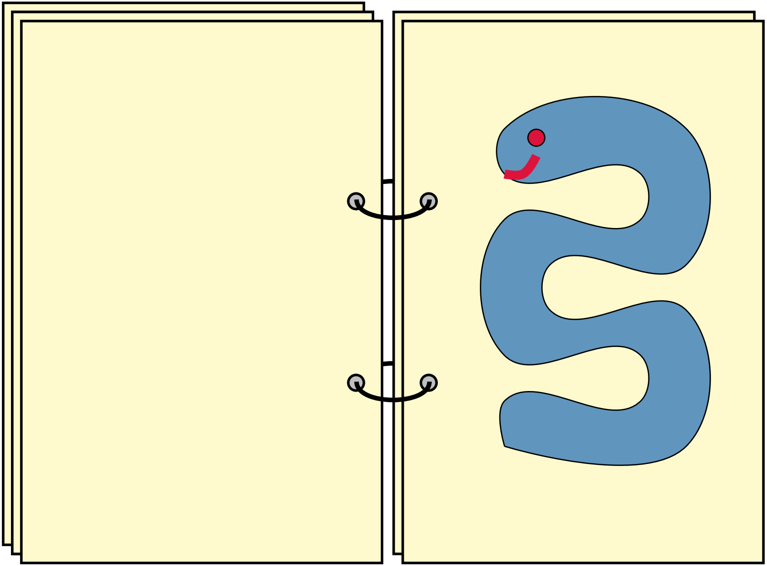
Current Sound Class Approaches

Dolgopolsky (1986)

Based on an empirical basis, speech sounds are divided into ten types, distinguished ‘in such a way that phonetic correspondences inside a “type” are more regular than those between different types’ (Dolgopolsky 1986: 35).

Holman et al. (2011)

Sounds are represented in ASJP code (Brown et al. 2008), a transcription system, which reduces the full range of the IPA alphabet to 41 symbols. An automatic approach for the calculation of the frequency of sound correspondences is applied to a large database (Wichmann et al. 2010), giving the possibility to determine transition probabilities between the 41 sound classes.



Main Ideas

Sound Classes

Scoring Functions

Position-Specific Scoring

Swap Check

Main Ideas

Sound Classes

Phonetic sequences are internally represented as sound classes.

Scoring Functions

Position-Specific Scoring

Swap Check

Main Ideas

Sound Classes

Phonetic sequences are internally represented as sound classes.

Scoring Functions

Scoring functions define specific transition probabilities among sound classes.

Position-Specific Scoring

Swap Check

Main Ideas

Sound Classes

Phonetic sequences are internally represented as sound classes.

Scoring Functions

Scoring functions define specific transition probabilities among sound classes.

Position-Specific Scoring

Substitution scores vary according to prosodic context.

Swap Check

Main Ideas

Sound Classes

Phonetic sequences are internally represented as sound classes.

Scoring Functions

Scoring functions define specific transition probabilities among sound classes.

Position-Specific Scoring

Substitution scores vary according to prosodic context.

Swap Check

Alignments are automatically searched for swapped sites.

Position-Specific Scoring

Position-Specific Scoring

→ It is assumed that sound change occurs more frequently in prosodically *weak* positions of phonetic sequences (cf. Geisler 1992).

Position-Specific Scoring

- It is assumed that sound change occurs more frequently in prosodically *weak* positions of phonetic sequences (cf. Geisler 1992).
- Given the sonority structure of a phonetic sequence, one can, apart from the *initial* and *final* positions, distinguish positions of *ascending*, *maximum* and *descending* sonority.

Position-Specific Scoring

- It is assumed that sound change occurs more frequently in prosodically *weak* positions of phonetic sequences (cf. Geisler 1992).
- Given the sonority structure of a phonetic sequence, one can, apart from the *initial* and *final* positions, distinguish positions of *ascending*, *maximum* and *descending* sonority.
- These positions can be ordered in a *hierarchy of strength* (initial > ascending > descending > maximum > final).

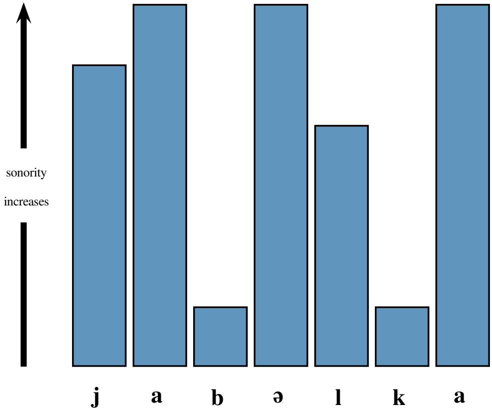
Position-Specific Scoring

- It is assumed that sound change occurs more frequently in prosodically *weak* positions of phonetic sequences (cf. Geisler 1992).
- Given the sonority structure of a phonetic sequence, one can, apart from the *initial* and *final* positions, distinguish positions of *ascending*, *maximum* and *descending* sonority.
- These positions can be ordered in a *hierarchy of strength* (initial > ascending > descending > maximum > final).
- Based on the relative strength of all sites in a phonetic sequence, substitution scores and gap penalties are modified by scaling factors, favoring changes in weaker positions and aggravating them in stronger positions.

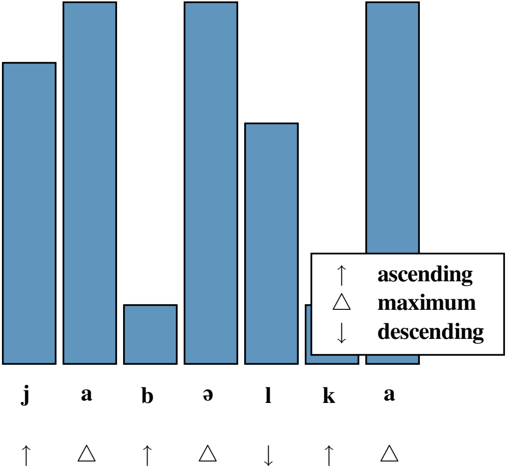
Position-Specific Scoring

j	a	b	ə	l	k	a

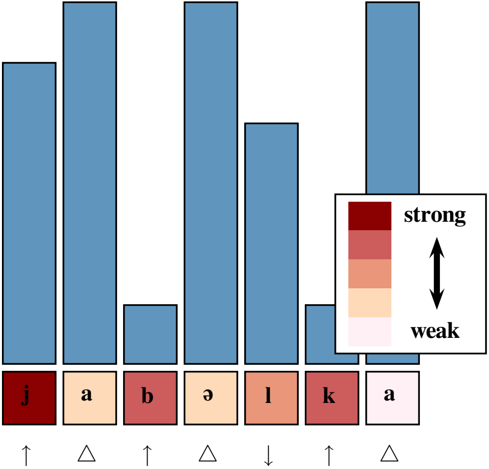
Position-Specific Scoring



Position-Specific Scoring



Position-Specific Scoring



Working Procedure

INPUT SEQUENCES

jabl̩ko
jabəlka
jabləkə
japkɔ

Working Procedure

SOUND CLASS CONVERSION

jabl̩ko	⟶	`yablko`
jabəlka	⟶	`yab3lka`
jabləkə	⟶	`yabl3k3`
japkɔ	⟶	`yapko`

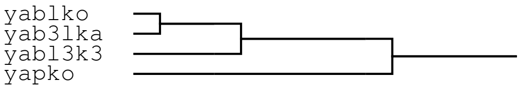
Working Procedure

DISTANCE CALCULATION

yablko	0.00	0.14	0.34	0.12
yab3lka	0.14	0.00	0.46	0.28
yabl3k3	0.34	0.46	0.00	0.44
yapko	0.12	0.28	0.44	0.00

Working Procedure

CLUSTER ANALYSIS



Working Procedure

PROGRESSIVE ALIGNMENT

```
yablko
yab3lka
yabl3k3
yapko
```

Working Procedure

PROGRESSIVE ALIGNMENT

```
y   a   b   -   l   k   o
y   a   b   3   l   k   a
yabl3k3
yapko
```

Working Procedure

PROGRESSIVE ALIGNMENT

y	a	b	-	l	-	k	o
y	a	b	3	l	-	k	a
y	a	b	-	l	3	k	3

yapko

Working Procedure

PROGRESSIVE ALIGNMENT

y	a	b	-	l	-	k	o
y	a	b	3	l	-	k	a
y	a	b	-	l	3	k	3
y	a	p	-	-	-	k	o

Working Procedure

SWAP CHECK

y	a	b	-	l	-	k	o
y	a	b	3	l	-	k	a
y	a	b	-	l	3	k	3
y	a	p	-	-	-	k	o

Working Procedure

IPA CONVERSION

y	a	b	...	→	j	a	b	. . .
y	a	b	...	→	j	a	b	. . .
y	a	b	...	→	j	a	b	. . .
y	a	p	...	→	j	a	p	. . .

Working Procedure

OUTPUT MSA

j	a	b	-	l̩	k	o
j	a	b	ə	l	k	a
j	a	b	l	ə	k	ə
j	a	p	-	-	k	ɔ

Implementation

The algorithm is implemented as part of the LingPy library (List 2011, see `http://lingulist.de/lingpy/`). LingPy is a suite of open source Python modules for sequence comparison, distance analyses, data operations and visualization methods in quantitative historical linguistics.

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	-

Evaluation

Evaluation

→ In biological analyses, the performance of alignment algorithms is traditionally tested by comparing manually edited alignments (*reference alignment*) with those produced by the respective algorithms (*test alignment*).

Evaluation

- In biological analyses, the performance of alignment algorithms is traditionally tested by comparing manually edited alignments (*reference alignment*) with those produced by the respective algorithms (*test alignment*).
- There exist different evaluation measures for determining the goodness of an algorithm.

Evaluation

→ In biological analyses, the performance of alignment algorithms is traditionally tested by comparing manually edited alignments (*reference alignment*) with those produced by the respective algorithms (*test alignment*).

→ There exist different evaluation measures for determining the goodness of an algorithm.

(a) The *percentage of identical columns* score (PIC) calculates how many columns match in the reference and the test alignment.

Evaluation

→ In biological analyses, the performance of alignment algorithms is traditionally tested by comparing manually edited alignments (*reference alignment*) with those produced by the respective algorithms (*test alignment*).

→ There exist different evaluation measures for determining the goodness of an algorithm.

(a) The *percentage of identical columns* score (PIC) calculates how many columns match in the reference and the test alignment.

(b) The *percentage of identical rows* score (PIR) calculates how many rows match in the reference and the test alignment.

Evaluation

→ In biological analyses, the performance of alignment algorithms is traditionally tested by comparing manually edited alignments (*reference alignment*) with those produced by the respective algorithms (*test alignment*).

→ There exist different evaluation measures for determining the goodness of an algorithm.

(a) The *percentage of identical columns* score (PIC) calculates how many columns match in the reference and the test alignment.

(b) The *percentage of identical rows* score (PIR) calculates how many rows match in the reference and the test alignment.

(c) The *sum-of-pairs* score (SOP) calculates the size of the intersection of aligned pairs of residues in the reference and the test alignment divided by the size of aligned pairs of residues in the reference alignment (cf. Thompson et al. 1999).

Evaluation

→ In biological analyses, the performance of alignment algorithms is traditionally tested by comparing manually edited alignments (*reference alignment*) with those produced by the respective algorithms (*test alignment*).

→ There exist different evaluation measures for determining the goodness of an algorithm.

(a) The *percentage of identical columns* score (PIC) calculates how many columns match in the reference and the test alignment.

(b) The *percentage of identical rows* score (PIR) calculates how many rows match in the reference and the test alignment.

(c) The *sum-of-pairs* score (SOP) calculates the size of the intersection of aligned pairs of residues in the reference and the test alignment divided by the size of aligned pairs of residues in the reference alignment (cf. Thompson et al. 1999).

(d) The *modified rand index* (MRI) checks ‘whether the same elements are together in the [test] alignment and the [reference] alignment’ (Prokić et al. 2009: 21).

The BulDial Gold Standard (Prokić et al. 2009)

The BulDial Gold Standard (Prokić et al. 2009)

→ 152 manually edited MSAs

The BulDial Gold Standard (Prokić et al. 2009)

- → 152 manually edited MSAs
- → 192 taxa (dialect points)

The BulDial Gold Standard (Prokić et al. 2009)

- 152 manually edited MSAs
- 192 taxa (dialect points)
- ca. 30,000 sequences

Two Test Models

	DOLGO	**ASJP**
Sound Classes	Dolgopolsky (1986)	ASJP-Code
No. of Symbols	11	41
Scoring Function	simple matching	Brown et al. (2011)
Default Gap Pen.	-6	-12
Main Algorithm	LingPy	LingPy

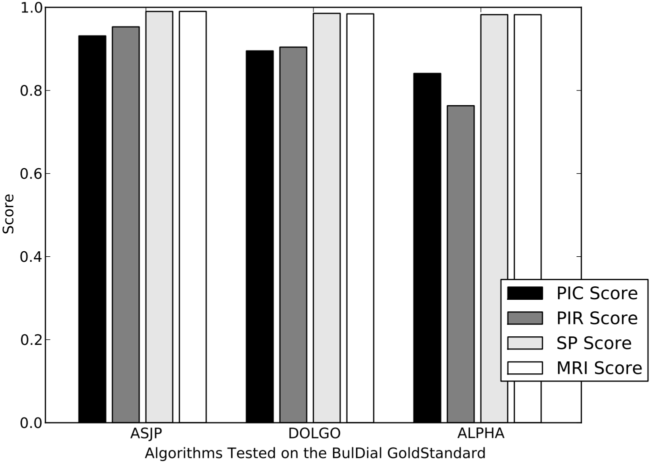
Two Test Models

	ALPHA (modif. by Prokić et al. 2009)
Sound Classes	no
No. of Symbols	full IPA
Scoring Function	CV distinction
Default Gap Pen.	unknown
Main Algorithm	ALPHAMALIG

Results

	ASJP	DOLGO	ALPHA
Perfect Alignments	132 (87%)	123 (81%)	103 (69%)
Perc. of Ident. Col.	0.9313	0.8952	0.8409
Per. of Ident. Rows	0.9531	0.9043	0.7632
Sum of Pairs	0.9901	0.9855	0.9825
Modified Rand Index	0.9902	0.9844	0.9824

Results


1.0
0.8
0.6
0.4
0.2
0.0
Score
ASJP
DOLGO
ALPHA
Algorithms Tested on the BulDial GoldStandard
■ PIC Score
▨ PIR Score
□ SP Score
□ MRI Score

Advantages of LingPy

Advantages of LingPy

→ The method for swap detection identifies 19 of 21 swapped sites in LingPy-ASJP, 13 of them are aligned properly.

Advantages of LingPy

→ The method for swap detection identifies 19 of 21 swapped sites in LingPy-ASJP, 13 of them are aligned properly.

→ The scoring function in LingPy-ASJP is based on a large empirical basis, allowing fine distinctions along with the extended model of sound classes.

Advantages of LingPy

- The method for swap detection identifies 19 of 21 swapped sites in LingPy-ASJP, 13 of them are aligned properly.
- The scoring function in LingPy-ASJP is based on a large empirical basis, allowing fine distinctions along with the extended model of sound classes.
- The application of prosodic profiles enhances both the calculation of the guide tree and the alignment process.

Examples

Examples

ALPHAMALIG

Aldomirovci	v	-	r	ɑ	-	tʃ	ɑ	m
Asparuhovo	v	-	r	ɤ	ʃ	t	ɑ	m
Panagjurishte	v	ɤ	r	-	ʃ	t	ə	m
Rakovica	v	-	r̩	-	ʃ	t	ɑ	m
Stambolovo	v	-	r	ɤ	ç	t	ə	m

Examples

ALPHAMALIG

Aldomirovci	v	-	r	ɑ	-	ʧ	ɑ	m
Asparuhovo	v	-	r	ɤ	ʃ	t	ɑ	m
Panagjurishte	v	ɤ	r	-	ʃ	t	ə	m
Rakovica	v	-	r̩	-	ʃ	t	ɑ	m
Stambolovo	v	-	r	ɤ	ç	t	ə	m

LingPy-ASJP

Aldomirovci	v	r	ɑ	-	ʧ	ɑ	m
Asparuhovo	v	r	ɤ	ʃ	t	ɑ	m
Panagjurishte	v	ɤ	r	ʃ	t	ə	m
Rakovica	v	r̩	-	ʃ	t	ɑ	m
Stambolovo	v	r	ɤ	ç	t	ə	m

Examples

Examples

LingPy-DOLGO

Aldomirovci	u	n	e	t	r	e
Asparuhovo	-	v	ɤ	t	r^j	ə
Babjak	f	n	e	t	r	e
Bachkovo	-	v	ɑ	t	r^j	ə
Bagrenci	u	n	e	t	r	e

Examples

LingPy-DOLGO

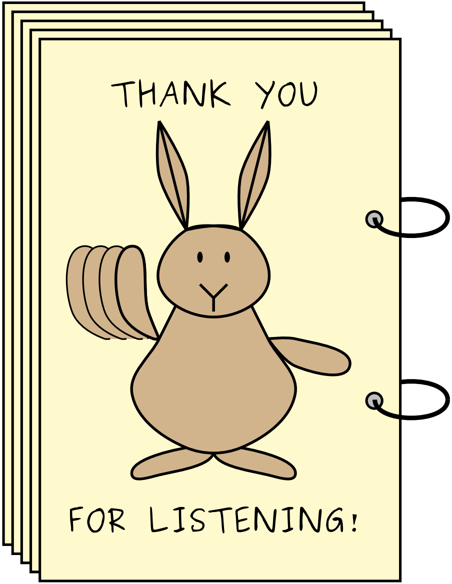
Aldomirovci	u	n	e	t	r	e
Asparuhovo	-	v	ɤ	t	rʲ	ə
Babjak	f	n	e	t	r	e
Bachkovo	-	v	ɑ	t	rʲ	ə
Bagrenci	u	n	e	t	r	e

LingPy-ASJP

Aldomirovci	u	n	e	t	r	e
Asparuhovo	v	-	ɤ	t	rʲ	ə
Babjak	f	n	e	t	r	e
Bachkovo	v	-	ɑ	t	rʲ	ə
Bagrenci	u	n	e	t	r	e

Special thanks to:

- Jelena Prokić, Martijn Wieling, and John Nerbonne for providing me with the BulDial Gold Standard and the source codes for calculating CDE and MRI scores.
- The German Federal Ministry of Education and Research (BMBF) for funding our research project.
- Hans Geisler for his helpful, critical, and inspiring support.


THANK YOU
FOR LISTENING!