Validating Pronunciation Measurements

Wilbert Heeringa
University of Groningen, Faculty of Arts, Scandinavian Department

LOT Winter School 2012
University of Tilburg, January 10, 2012

Overview

- Measuring pronunciation differences with Levenshein distance
- Validating pronunciation measurements
- Application to Dutch

Measuring pronunciation differences with Levenshein distance

Algorithm

- *Levenshtein distance* was applied for comparison of Irish dialects by Kessler in 1995. Later it was applied to Dutch, Sardinian, Norwegian, American English, German and Bulgarian dialects, and Bantu languages.
- Calculate the cost of changing one string into another.
- Example: *afternoon* may be pronounced as [ˈæəftəˌnʉˑn] in the dialect of Savannah and as [ˌæftərˈnuˑn] in the dialect of Lancaster.
- Change the first pronounciation into the other.

æəftənʉn	delete ə	1
æftənʉn	insert r	1
æftərnʉn	subst. ʉ/u	1
æftərnun		
		3

Algorithm

- Many sequence operations map [æəftənʉn] $\longrightarrow$ [æftərnun]. Levenshtein distance $=$ cost of cheapest mapping.
- The sum of the operations is divided by the length of the longest alignment which gives the minimum cost. The longest alignment has the greatest number of matches.
- Example:

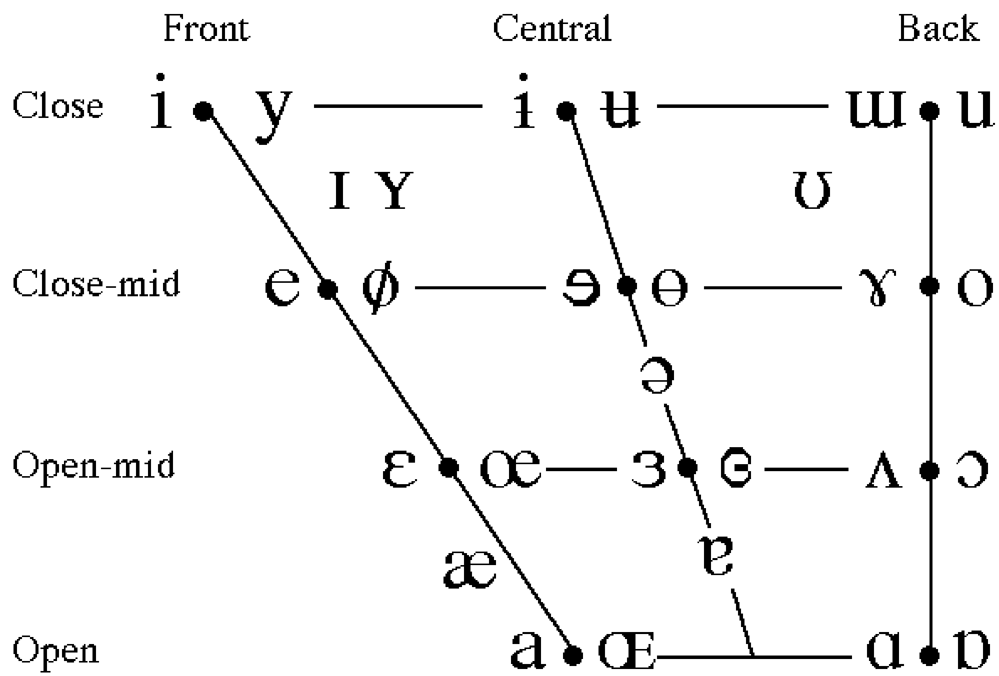
æ	ə	f	t	ə	∅	n	ʉ	n
æ	∅	f	t	ə	r	n	u	n
	1				1		1	

  A total cost of 3 divided by a length of 9 gives a word distance of 0.33 or 33%.
- Using m words the distance between two dialects is equal to the average of m Levenshtein distances.

Factors (1)

- Operation weights: segment distances.
  - Phones: binary distances
  - Features: gradual distances, e.g. IPA features, vowels are defined by advancement, height, rounding, and consonants are defined by place of articulation, manner of articulation and voicing.
  - Acoustic: gradual distances measured between acoustic representations (spectrograms, Barkfilters) of a set of predefined IPA samples

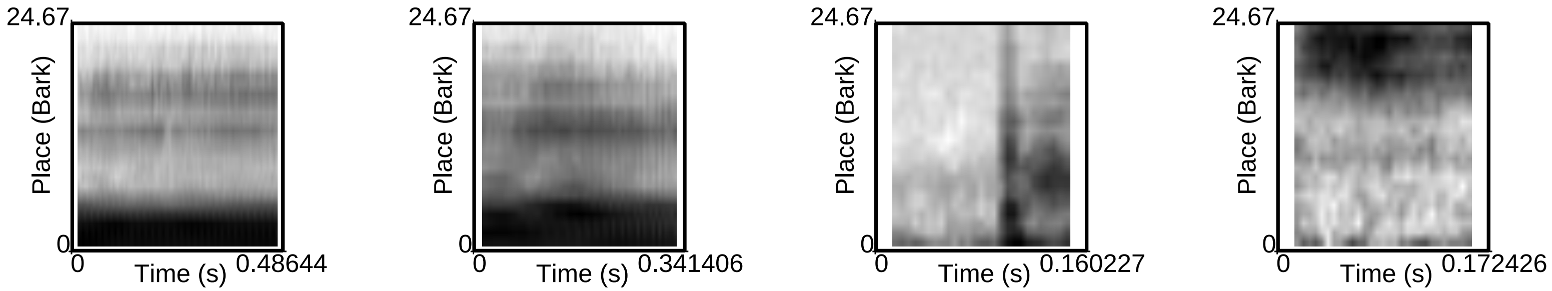
Operation weights: feature-based segment distances



Vowel distances in the Almeida & Braun system: distances of 1 point: ε vs. æ (hight), ε vs. ɜ (advancement), ε vs. œ (round).

Operation weights: acoustic-based segment distances

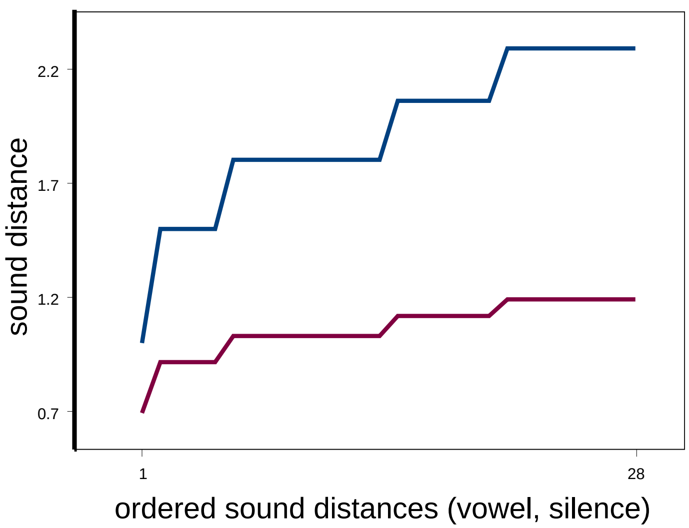
- Feature systems mostly not based on physical measurements.
- Samples of all IPA segments are found on the audio tape or CD published by the *Department of Phonetics and Linguistics* of the *University College London*: *The Sounds of the International Phonetic Alphabet* (1995).
- Calculate distances between the samples using their spectrograms or formant tracks.
- Intensity is processed, durations are made equal.



i e p s

Factors (2)

- Linear or logarithmic gradual segment distances:



Factors (3)

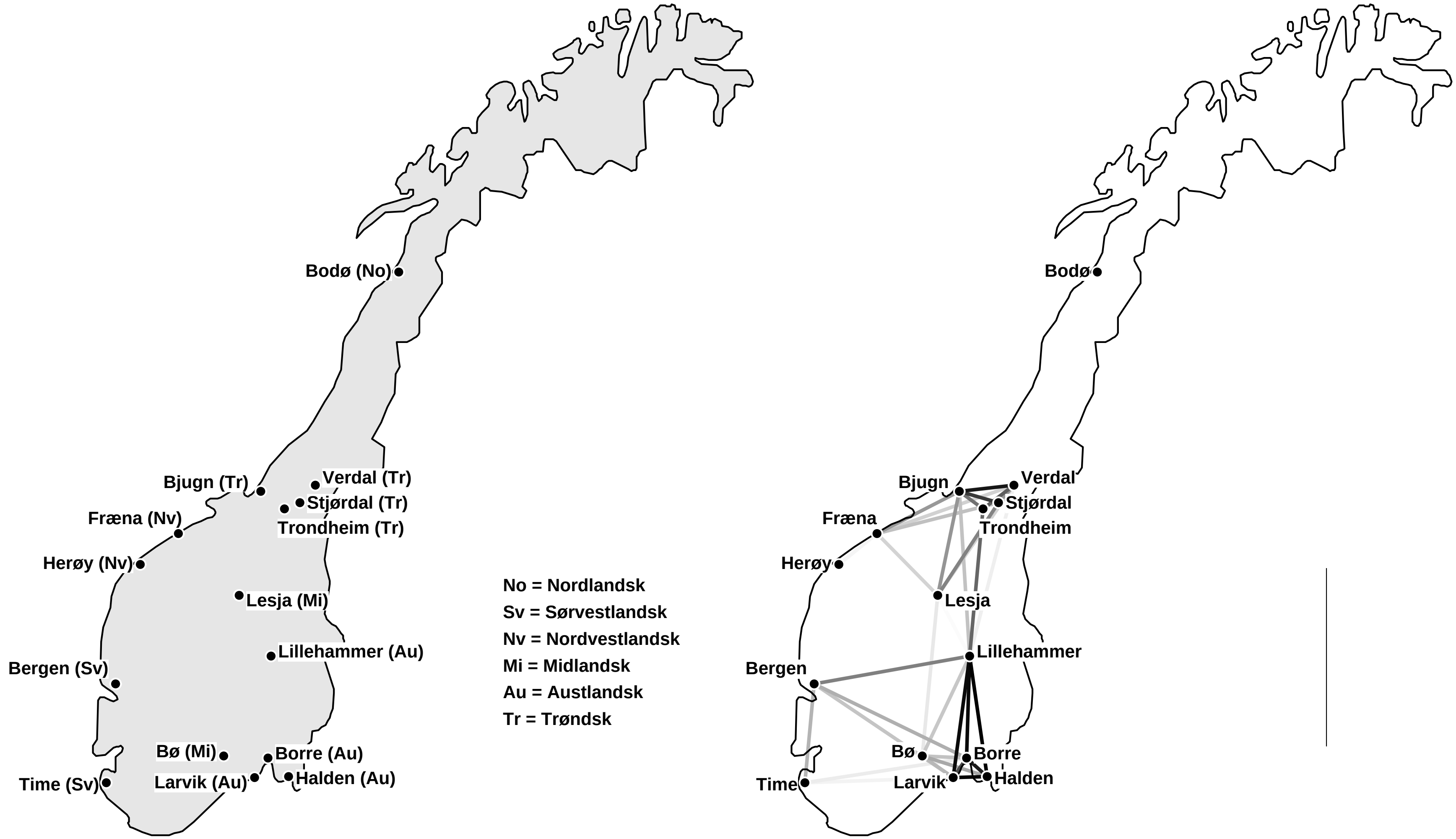
- Free alignment / linguistic alignment. In a linguistic alignment we assure that the minimum cost is based on a alignment in which:
  - a vowel matches with a vowel
  - a consonant matches with a consonant
  - the [j] or [w] matches with a vowel
  - the [i] or [u] matches with a consonant
  - the schwa matches with a sonorant
- Effect of context: unigrams, bigrams, trigrams, xbigrams.
- All-word comparison / same-word comparison.

Application

- Material compiled by Jørn Almberg.
- Translations in Norwegian dialects of the fable 'The North Wind and the Sun'.
- Audio files *and* transcriptions available via:

  `http://www.ling.hf.ntnu.no/nos/`

- We selected 15 dialects.
- Each dialect text usually consists of 58 different words.
- Multiple pronunciations of one word are processed.
- Phonetic, morphological and lexical differences are processed with Levenshtein distance (all-word comparison).



Left: distribution of 15 dialects in the Norwegian language area. Right: the averaged Levensthein distances between the dialects. Darker lines connect closer points, lighter lines more remote ones. Lines longer than the line on the right are omitted.

Consistency

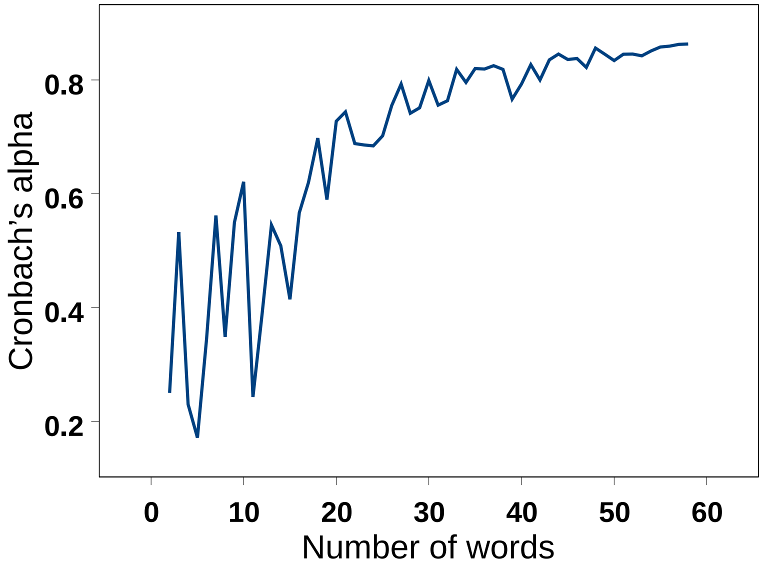
- Cronbach's α is a popular method to measure consistency or reliability.
- When calculating the distances between n_v varieties on the basis of n_w words, n_w matrices are obtained, each containing the distances between the n_v varieties on the basis of the pronunciations of one word.
- The average inter-correlation $\bar{r}$ among the words is calculated as:

$$\bar{r} = \frac{\displaystyle\sum_{i=2}^{n_w} \sum_{j=1}^{i-1} r(w_i, w_j)}{\frac{n_w \times (n_w - 1)}{2}}$$

- Cronbach's α can be written as a function of the number of words and the average inter-correlation among the words:

$$\alpha = \frac{n_w \times \bar{r}}{1 + (n_w - 1) \times \bar{r}}$$

Consistency

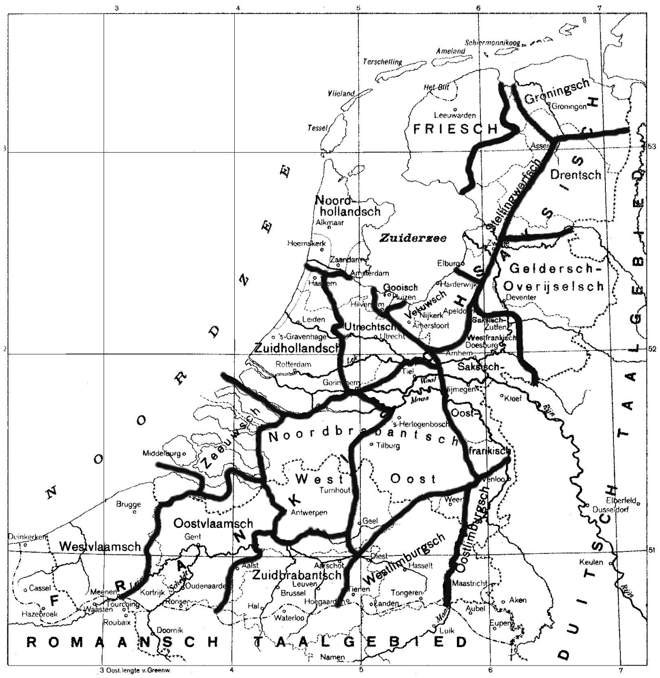


Cronbach's α values for random subsets of 2 through 58 words. From 25 words on α is always higher than 0.70. For 58 words α is equal to 0.86.

Validating
Pronunciation Measurements

Validation

- Qantitative comparison to consensus. Left: map of Lecoutere (1921), right: map of Daan (1969).



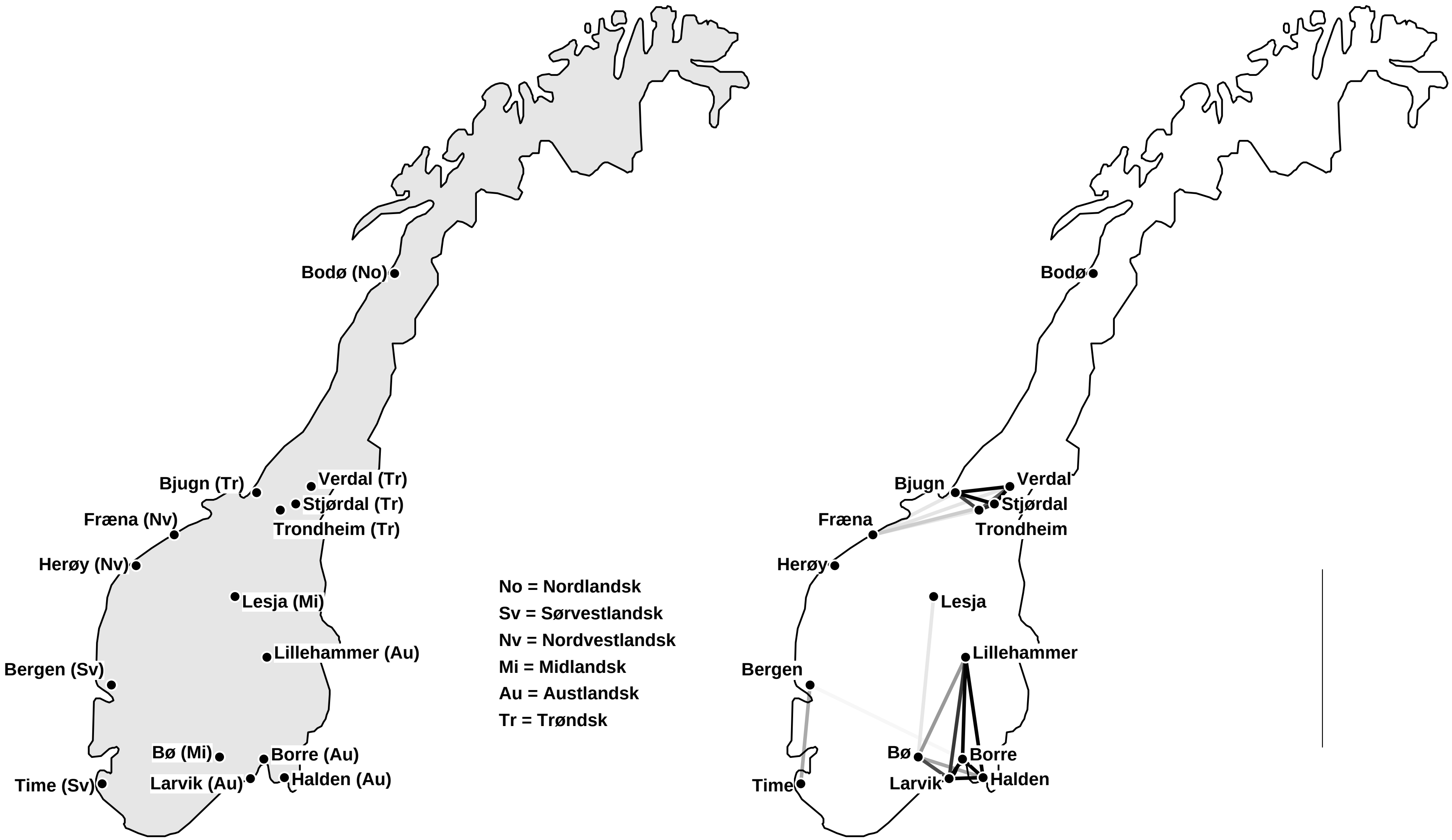
Validation

- Results of computational methods are compared to results of a perception experiment of Charlotte Gooskens.
- In each of the 15 locations, a group of 16 to 27 high school pupils listened to all 15 texts.
- The texts were presented in a randomized order.
- Task: each pupil notes for each text the distance of the corresponding dialect compared to his own dialect.
- Scale from 1 (similar to own dialect) to 10 (not similar to own dialect).
- Final result: a 15×15 perceptual distance matrix.
- We correlate the distance matrices of the various computational methods with the perceptual distance matrix.

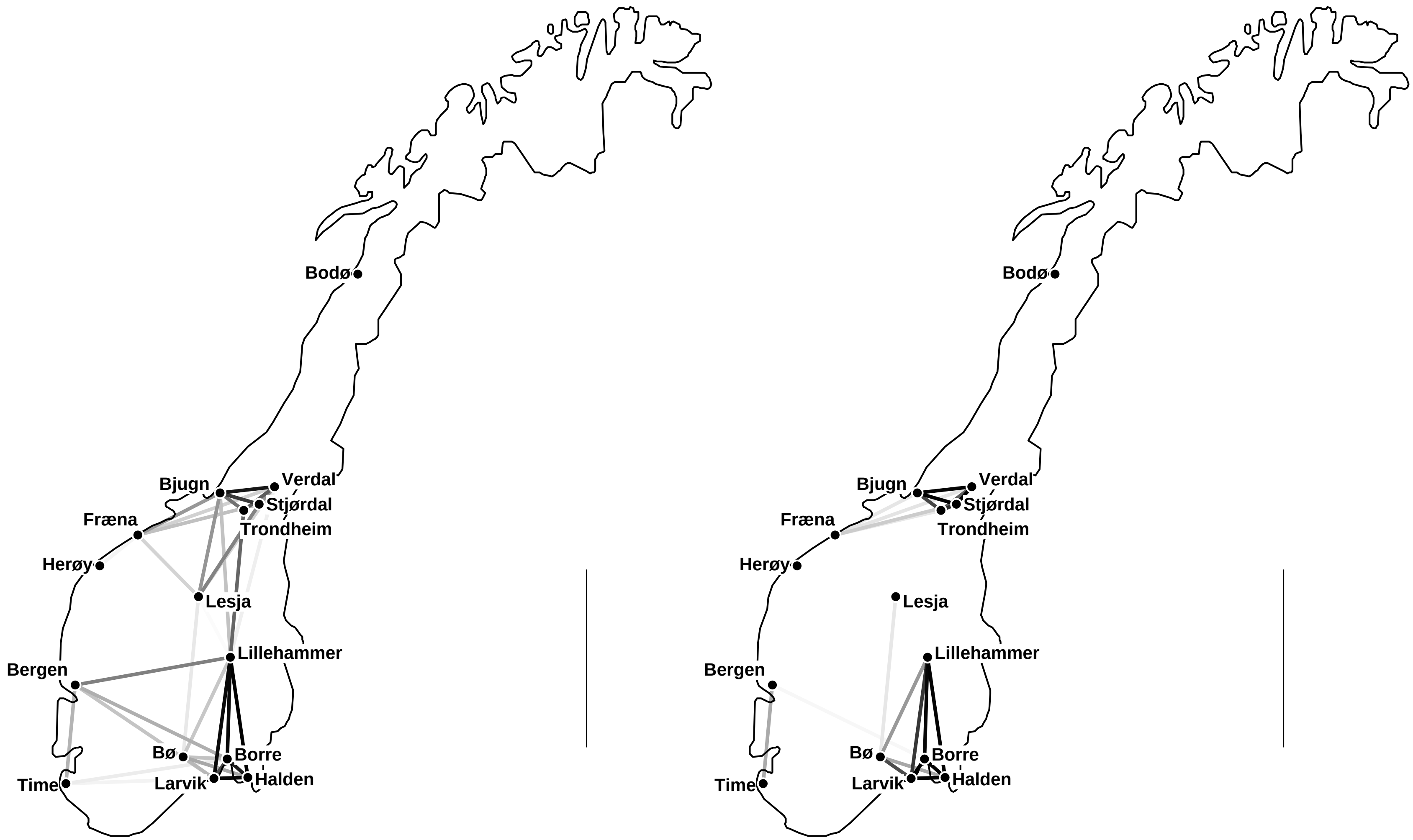
Validation

	Be	Bj	Bo	Bø	Bo	Fr	Ha	He	La	Le	Li	St	Ti	Tr	Ve
Bergen	1.7	9.0	8.2	8.0	7.7	7.7	8.2	6.9	8.0	8.9	8.5	8.4	4.8	8.5	8.0
Bjugn	9.1	3.4	6.4	8.2	9.2	5.8	8.3	8.0	8.4	7.3	9.1	2.2	8.0	3.3	2.8
Bodø	8.7	7.9	1.5	8.3	8.3	6.6	7.9	7.8	7.3	8.0	8.7	6.6	8.1	6.2	6.3
Bø	8.1	7.8	7.5	1.0	7.7	8.1	4.9	7.8	5.3	6.0	5.1	7.1	6.3	8.2	8.6
Borre	6.1	8.8	7.8	6.5	1.7	8.5	1.8	7.5	1.6	7.5	2.0	7.2	7.5	8.5	9.1
Fræna	9.0	7.5	7.1	8.4	8.8	3.1	8.1	7.8	8.5	7.2	9.0	6.6	7.4	6.1	7.6
Halden	7.0	8.2	8.0	6.8	4.0	8.1	2.8	7.9	2.8	6.6	3.0	7.4	7.0	8.0	8.3
Herøy	8.6	9.3	8.4	8.5	9.1	7.0	8.6	1.2	9.3	9.3	9.4	8.5	7.5	7.5	8.2
Larvik	7.4	8.7	7.6	4.0	4.0	7.7	3.2	5.6	3.4	7.1	4.6	8.2	6.8	8.3	7.5
Lesja	8.5	7.6	7.8	7.4	8.2	7.3	7.6	7.7	7.6	1.0	7.1	6.9	7.2	7.7	8.2
Lillehammer	6.7	8.3	8.1	6.2	4.4	8.0	3.1	7.5	4.1	7.3	2.7	7.6	6.8	8.7	8.1
Stjørdal	8.7	3.7	6.8	7.7	8.1	6.0	7.5	7.7	8.3	7.1	8.3	2.0	7.7	3.8	3.4
Time	7.0	9.3	8.4	8.1	8.4	8.3	8.0	7.2	8.2	9.1	8.8	8.8	1.8	8.8	9.0
Trondheim	7.8	5.8	6.7	7.5	6.4	7.3	6.0	7.1	5.9	7.9	6.3	4.4	7.6	3.3	6.8
Verdal	8.8	3.4	6.4	8.2	8.4	5.7	7.2	7.9	7.9	7.4	8.4	1.8	7.9	3.1	2.6

Perceptual distances among 15 Norwegian dialeact varieties. Distance pairs A–B / B–A are averaged in the validation analyses.



Left: distribution of 15 dialects in the Norwegian language area. Right: the perceptual distances between the dialects. Darker lines connect closer points, lighter lines more remote ones. Lines longer than the line on the right are omitted.



Left: the averaged Levensthein distances between the dialects. Right: the perceptual distances between the dialects. Darker lines connect closer points, lighter lines more remote ones. Lines longer than the line on the right are omitted. $r = 0.80$ (r=0.67).

Two factors

- Segment representations: phones / features / acoustic*
- Segment distance metric: linear / logarithmic*

* = Levenshtein only.

Three methods

- Corpus frequency method:
  - Hoppenbrouwers and Hoppenbrouwers (1988)
  - Dialect distance: sum of *phone* or *feature* frequency differences.
- Frequency per word method:
  - Nerbonne and Heeringa (1998).
  - Word distance: sum of *phone* of *feature* frequency differences.
  - Words processed as linguistic units.
- Levenshtein distance:
  - Used by Kessler (1995) for finding distances between Irish Gaelic dialects.
  - Sensitive to segment order in a word.

Validation

- We validated different dialect comparison methods, segment representations and segment distance metrics.
- The effect is shown in the average correlation coefficients of computational distances with respect to perceptual distances on the basis of 15 Norwegian varieties:

	Corp. freq.	Freq. word	Lev. lin.	Lev. log.
phones	0.66	0.66	0.67	0.67
features	0.46	0.59	0.62	0.64
acoustic			0.64	0.66

Validation

- In a study with Peter Kleiweg, Charlotte Gooskens and John Nerbonne we studied the effect of alignment (no alignment, free alignment, linguistically based alignment) and normalization (distance divided by the length of the alignment or not).
- Phonetic and morphological differences are processed with Levenshtein distance (same-word comparison).
- Stress, tonemes, suprasegmentals and diacritics are not processed. Affricates are processed as a sequence of consonants.

Validation

- Recent results for the effect of aligment and normalization:

		without length normalization	with length normalization
no	alignment	0.7011	0.6658
free	alignment	0.7078	0.6725
linguistic	alignment	0.7060	0.6730

Validation

- In the same study, the effect of context was examined. Besides unigrams we considered bigrams, trigrams and xbigrams.
- Results from the paper for the effect of context:

	without length normalization	with length normalization
unigram	0.67	0.66
bigram	0.68	0.67
trigram	0.70	0.68
xbigram	0.70	0.69

Conclusions

- **Comparison method**:
  corp. freq. meth. $<$ freq. per word meth. $<$ Lev. dist.
- **Segment representation**:
  features $<$ acoustic $<$ phones.
- **Segment distance metric**:
  linear $<$ logarithmic.
- **Alignment**:
  - without normalization:
    no alignment $<$ ling. alignment $<$ free alignment;
  - with normalization:
    no alignment $<$ free alignment $<$ ling. alignment.
- **Normalization**:
  with normalization $<$ without normalization.
- **Context**: unigrams $<$ bigrams $<$ trigrams $=$ xbigrams.

Choice

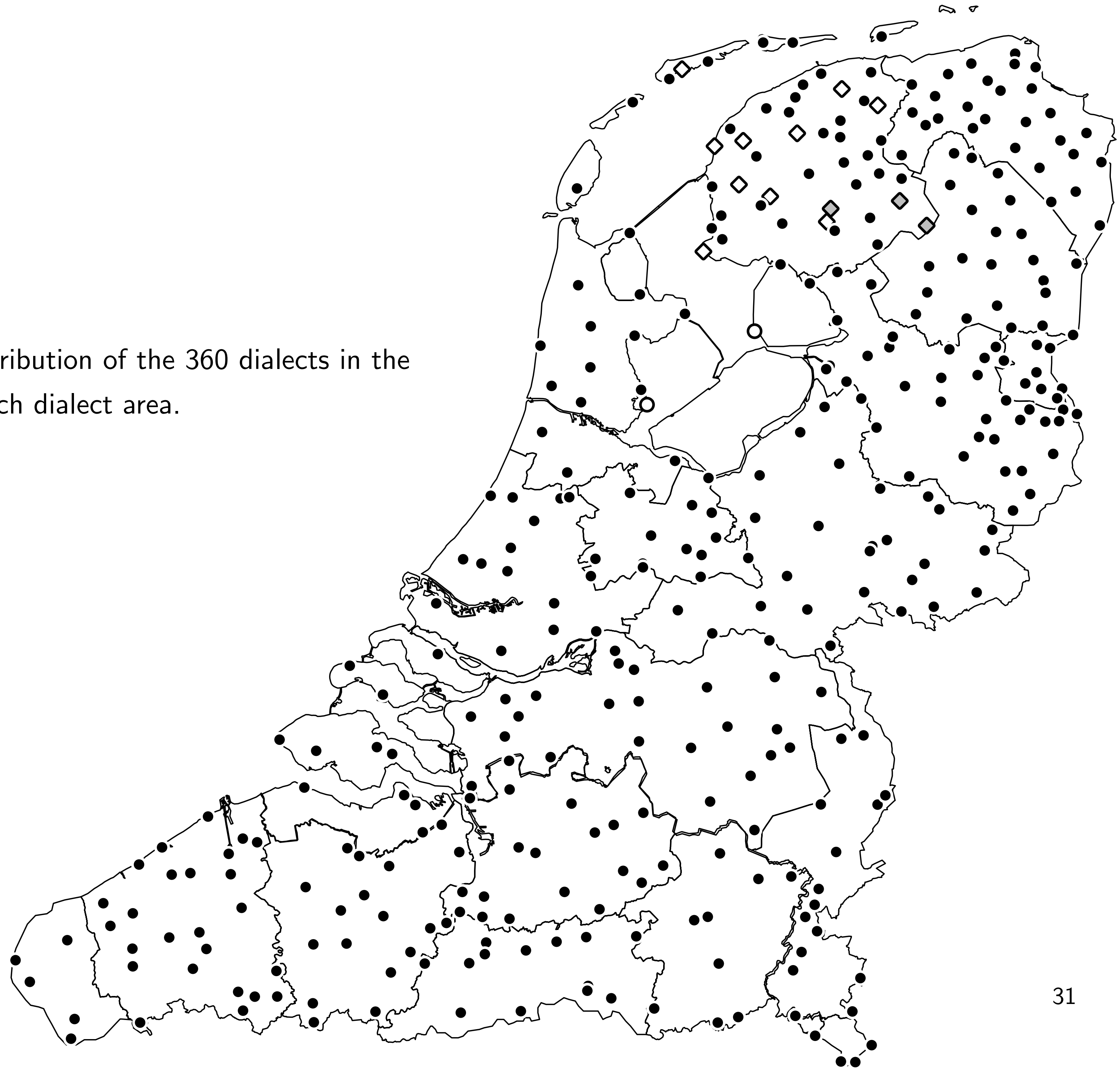
- We use:
  - Levenshtein distance, based on unigrams
  - Logarithmic acoustic segment distances
  - Linguistic alignments without alignment normalization

Application to Dutch

Data source

- Reeks Nederlandse Dialectatlassen
- Compiled by E. Blancquaert and W. Pée
- Texts from 1922–1975
- 1956 dialects, 139 sentences each
- we selected 360 dialects, 125 words

Distribution of the 360 dialects in the Dutch dialect area.



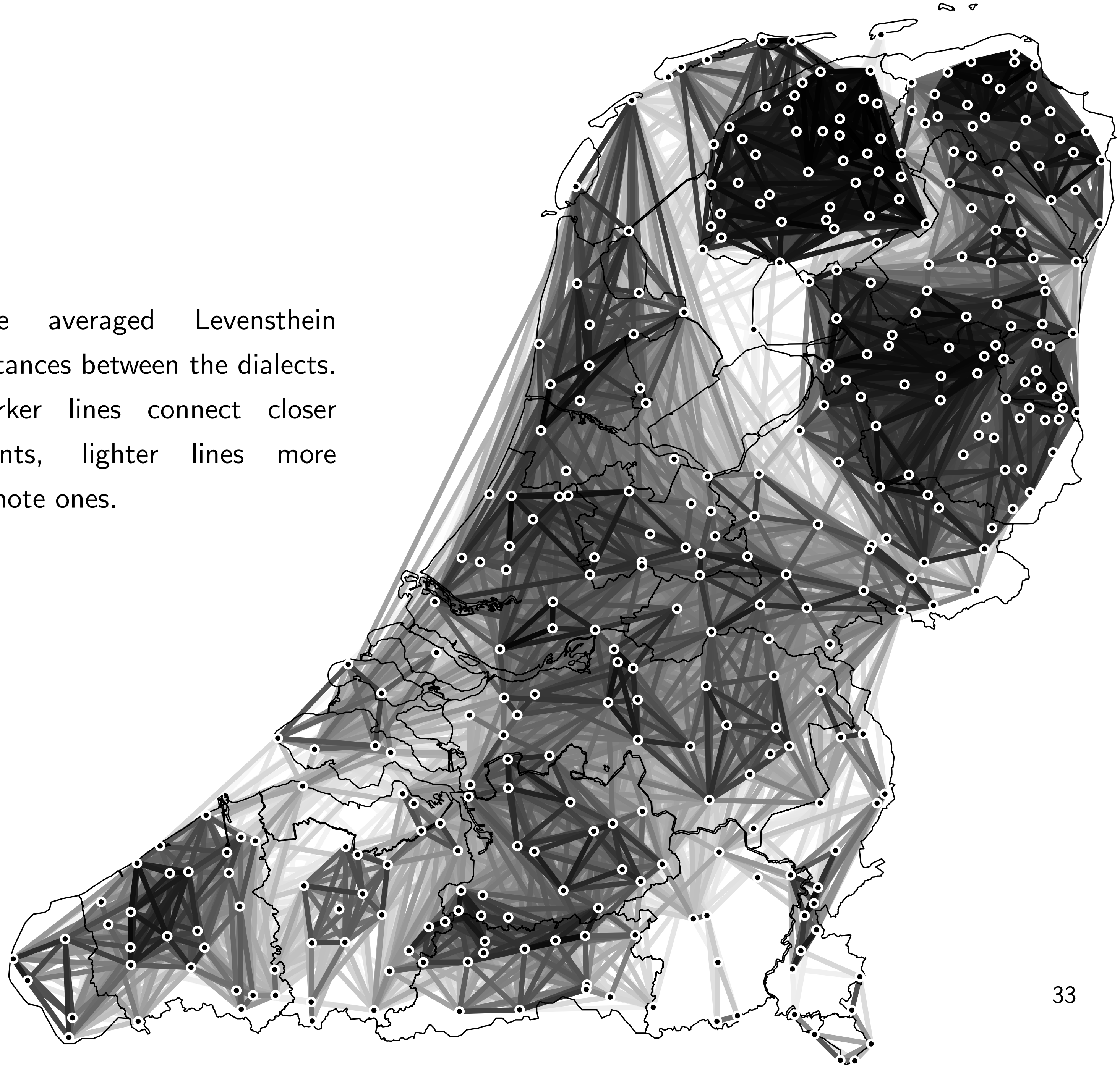
Aggregating pronunciation distances

- Simplified example: Find distance between Middelstum and Ommen on the basis of 6 words. Diacritics are ignored, all operation costs have a weight of 1.

	Middelstum	Ommen		
schip	sxɪp	sxɪp	0	4
pet	pɛt	pɛtə	1	4
geroepen	rɔupm	ərupm	2	6
springen	sprɪŋ	sprɪŋkt	2	7
kelder	kɛlər	kɛldər	1	6
huis	hus	hys	1	3
			7	30

- Distance: $(7/30) \times 100 = 23\%$.
- From the RND questionaire we selected 125 words. Since only word pairs are considered which have the same lexemes (same-word comparisons), the number of word pairs per dialect pair may vary.

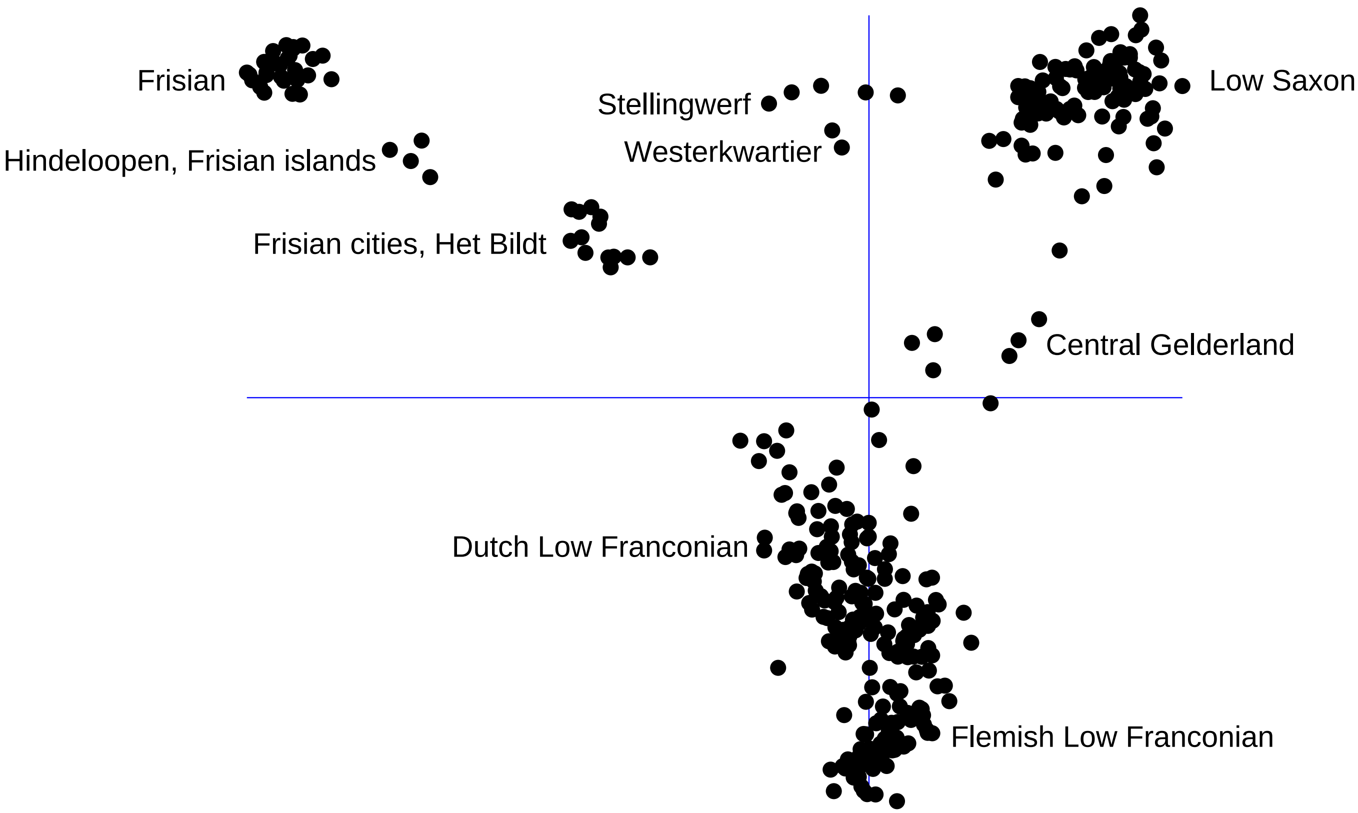
The averaged Levensthein distances between the dialects. Darker lines connect closer points, lighter lines more remote ones.



Multidimensional scaling

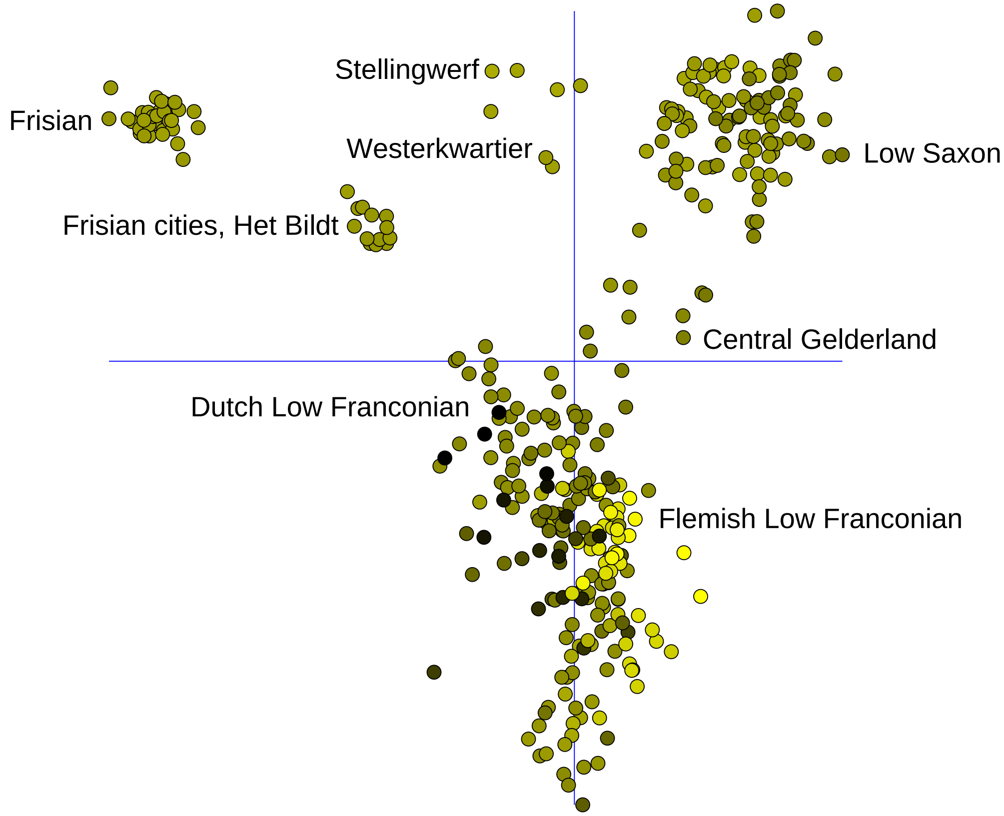
- Given a geographic map, distances between locations can be measured.
- Multidimensional scaling: given distances, locations on a map can be inferred.
- In our case: from $n \times n$ distances we infer coordinates in 2- or 3-dimensional space. So n dimensions are reduced to two or three.
- We used: Kruskal's Non-metric Multidimensional Scaling.
- Create a color map where each dialect gets an unique color: map 3 major MDS dimensions to red, green and blue.
- Color space between dialect points: interpolate using Inverse Distance Weighting.

Multidimensionale schaling



Using MDS the 360 dimensions are reduced to 2. The Y-coordinates represent the first and X-coordinates inversely represent the second dimension.

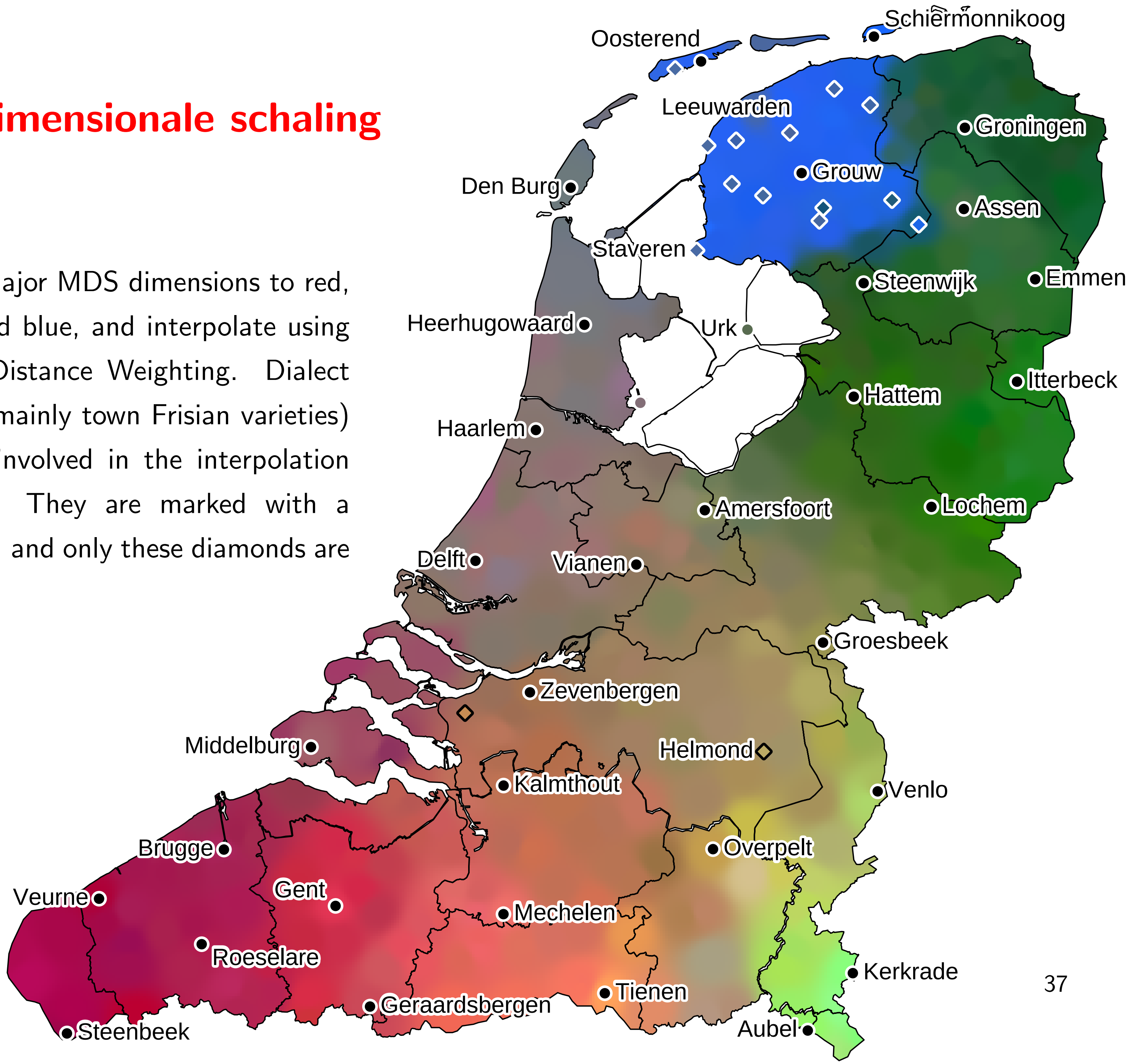
Multidimensionale schaling

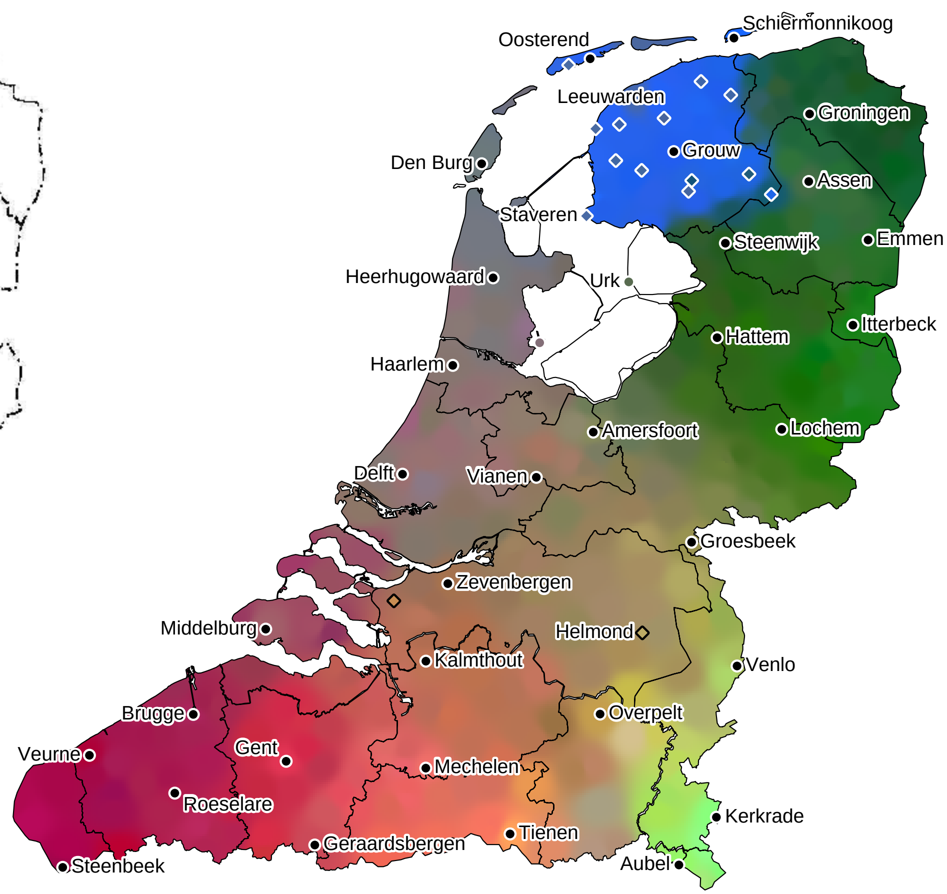


Using MDS the 360 dimensions are reduced to 3. Y-coordinates represent the first, X-coordinates inversely represent the second, and greytone represents the third dimension (distinct in the South).

Multidimensionale schaling

Map 3 major MDS dimensions to red, green and blue, and interpolate using Inverse Distance Weighting. Dialect islands (mainly town Frisian varieties) are not involved in the interpolation process. They are marked with a diamond, and only these diamonds are colored.





Left: dialect map of De Schutter, the Communis Opinio at the end of the 20th century. Right: color map based on 3 multidimensional scaling dimensions.

Final Remarks

- The slides are available at:
  http://www.let.rug.nl/~heeringa/Gabmap/slides.pdf.

- More about dialectometry in Groningen and Amsterdam can be found at:
  http://www.dialectometry.net/

- Gabmap is available at:
  http://gabmap.meertens.knaw.nl/.

- The maps in this presentation are produced with RuG/L04, developed by Peter Kleiweg and available at:
  http://www.let.rug.nl/kleiweg/L04/.