

# Phonetic and lexical predictors of intelligibility of Scandinavian varieties.

Karin Beijering (University of Groningen)  
Charlotte Gooskens (University of Groningen)  
Wilbert Heeringa (Meertens Institute)

The Thirteenth International Conference on Methods in Dialectology, 4 – 8 august 2008, Leeds.  
Workshop “Measuring linguistic relations between closely related varieties”

# outline

- > background
- > material / method
- > intelligibility
- > phonetic distance
- > lexical distance
- > correlations / regression analyses
- > conclusions

# background

- > aggregate phonetic distance successful predictor of intelligibility of standard Scandinavian languages,  $r = -.82$  and  $p < 0.01$ , (Gooskens 2007).
- > phonetic distance has not been split up for vowels and consonants / phonetic sublevels with regard to intelligibility.
- > lexical distance is another influential factor in intelligibility.

# research question 1

- > What is the relative contribution of (full) phonetic distance and lexical distance to intelligibility?

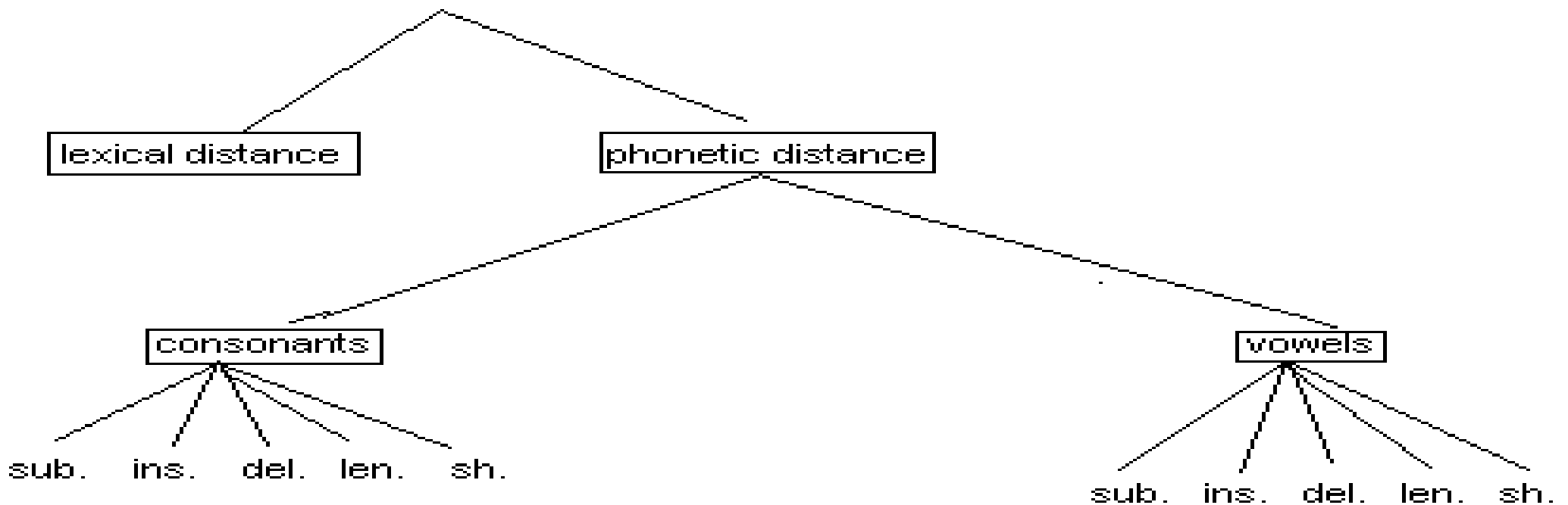
# research question 2

- > What is the relative contribution of consonant distances and vowel distances to intelligibility?

# research question 3

- > What is the relative contribution of substitution, insertion, deletion, lengthening and shortening of both vowels and consonants to intelligibility?

# intelligibility and linguistic distances



# consonants vs vowels

- > consonants are more important for the semantic identity of a word than vowels; they function as reference points in words.
- > word recognition in English depends more on correct consonant identification than on the correct identification of vowels (Van Ooijen 1994 and references).



# Ashby and Maidment (2005)

Mary has a little lamb.

- replace all vowels by [ɜ:]:

Mɜ:rɜ: hɜ:s ɜ: lɜ:tlɜ: lɜ:mb.

- replace all consonants by [d]:

Dady dad a diddde dadd.

# hypothesis

- > the occurrence of deviant segments in the consonantal structure of a word is more disturbing for intelligibility than changes in the vowels of a word.

# material

> *The North Wind and the Sun.*

- well-known text in phonetic research.
- recordings and transcriptions in 18 Scandinavian language varieties.
- average of 98 words.

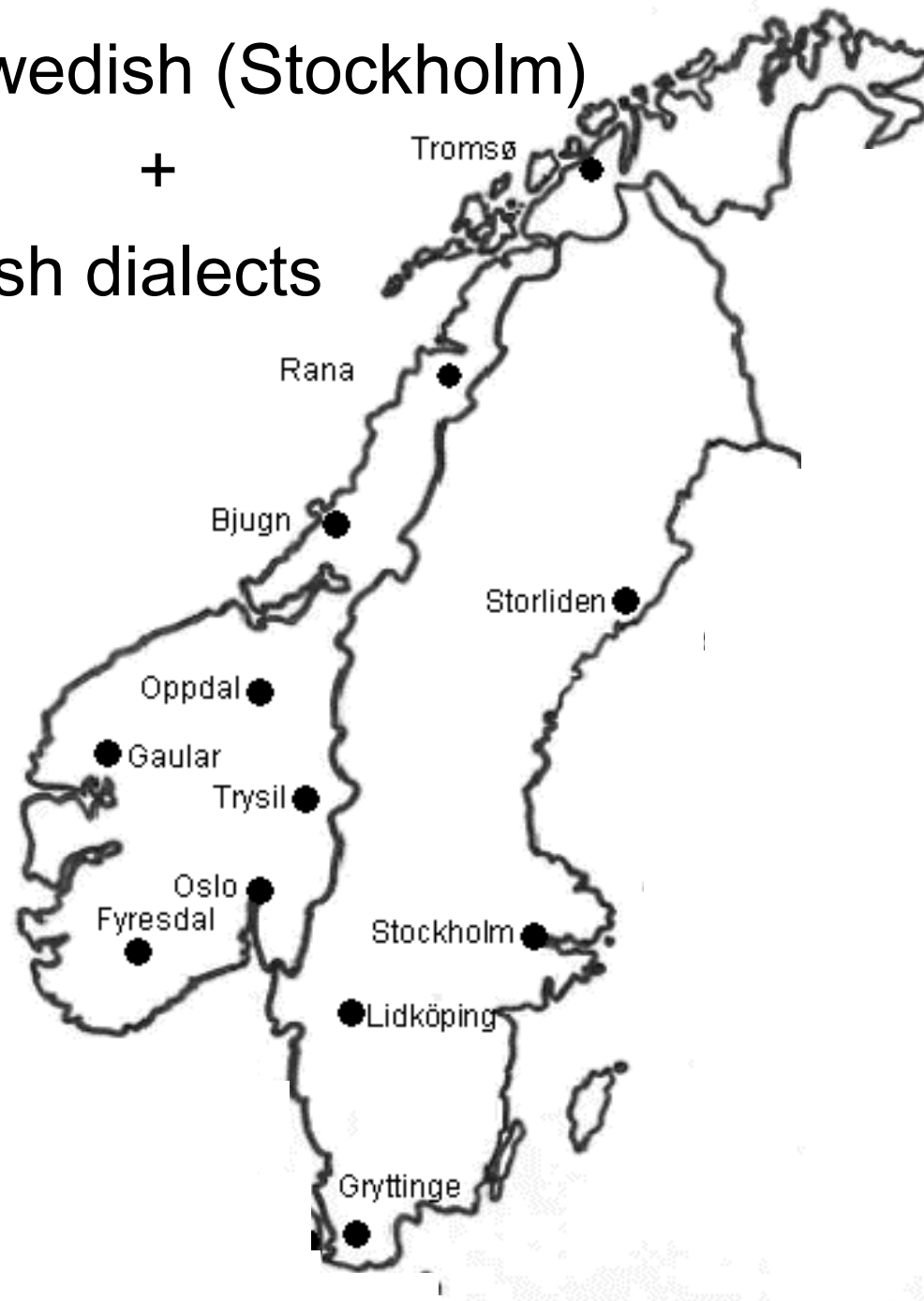
# 8 Norwegian varieties



# Standard Swedish (Stockholm)

+

## 3 Swedish dialects



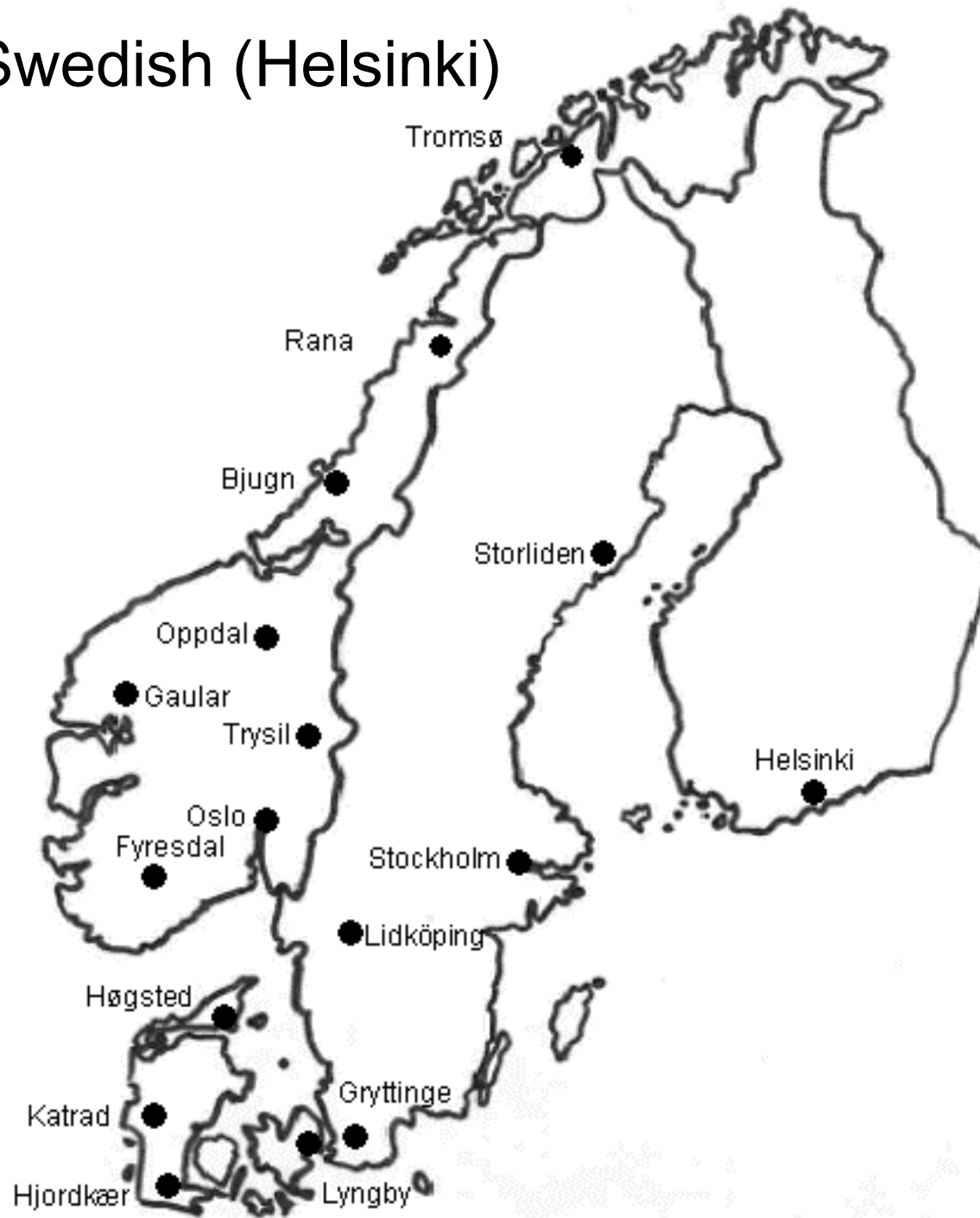
# Standard Danish (Lyngby)

+

## 3 Danish dialects



# Finland Swedish (Helsinki)



# Faroese (Torshavn)





# material

- > *The Scandinavian languages and dialects.*
- closely related language varieties.
- Standard Mainland Scandinavian languages are mutually intelligible.
- little research on dialect intelligibility in Scandinavia.

# intelligibility experiment

- how intelligible are the 18 Scandinavian language varieties for speakers of Standard Danish?
- **listening experiment** on the basis of recordings of the fable *The North Wind and the Sun* in all Scandinavian varieties.
- **test:** 6 sentences of the fable in 6 different varieties (random order)  
e.g. sentence 3 Tromsø, sentence 5 Lidköping , sentence 2 Oslo  
etc.

# intelligibility experiment

- **listeners:** 351 high school pupils from Copenhagen, aged between 15 and 20 (average 17,6).
- **task:** translate the spoken text into Standard Danish.
- **intelligibility score** = percentage of correctly translated words.

# phonetic distance

- > transcriptions of the fable *The North Wind and the Sun* in all Scandinavian varieties.
- > cognate wordpairs between Standard Danish and each of the other varieties.
- > Levenshtein algorithm.

# Levenshtein algorithm

> 5 operations to change a pronunciation in one variety into a corresponding pronunciation in another variety:

- **substitution**
- **insertion**
- **deletion**
- **lengthening**
- **shortening**

# Levenshtein algorithm

Standard DA [ʔe:ni] versus Standard SE [e:nɪgɑ]  
 'in agreement'

---

Standard DA

ʔeni

delete ʔ

eni

substitute i by ɪ

enɪ

insert g

enɪg

insert a

Standard SE

enɪgɑ

---

# Levenshtein algorithm

divide total number of operations by length of alignment:

	1	2	3	4	5	6
Standard DA	?	e	n	i		
Standard SE		e	n	I	g	
	del			sub	ins	in

phonetic distance:  $4/6 = 0.67$  or 67%

# Levenshtein algorithm

- > many different versions (cf. Heeringa 2004).
  
- > present study:
  - gradual segment distances.
  - spectrogram distances on the basis of IPA samples.
  - e.g. the distance between [ɑ] and [a] is smaller than between [a] and [i].



# Levenshtein algorithm

- > full phonetic distances [all speech segments].
- > consonant / vowel distance: sum of all consonant / vowel sublevel distances.
- > phonetic sublevel distance (e.g. consonant substitution): total number of consonant substitutions/length of alignment.
- > the larger the phonetic distance, the less intelligible another Scandinavian variety will be for a speaker of Standard Danish.

# lexical distance

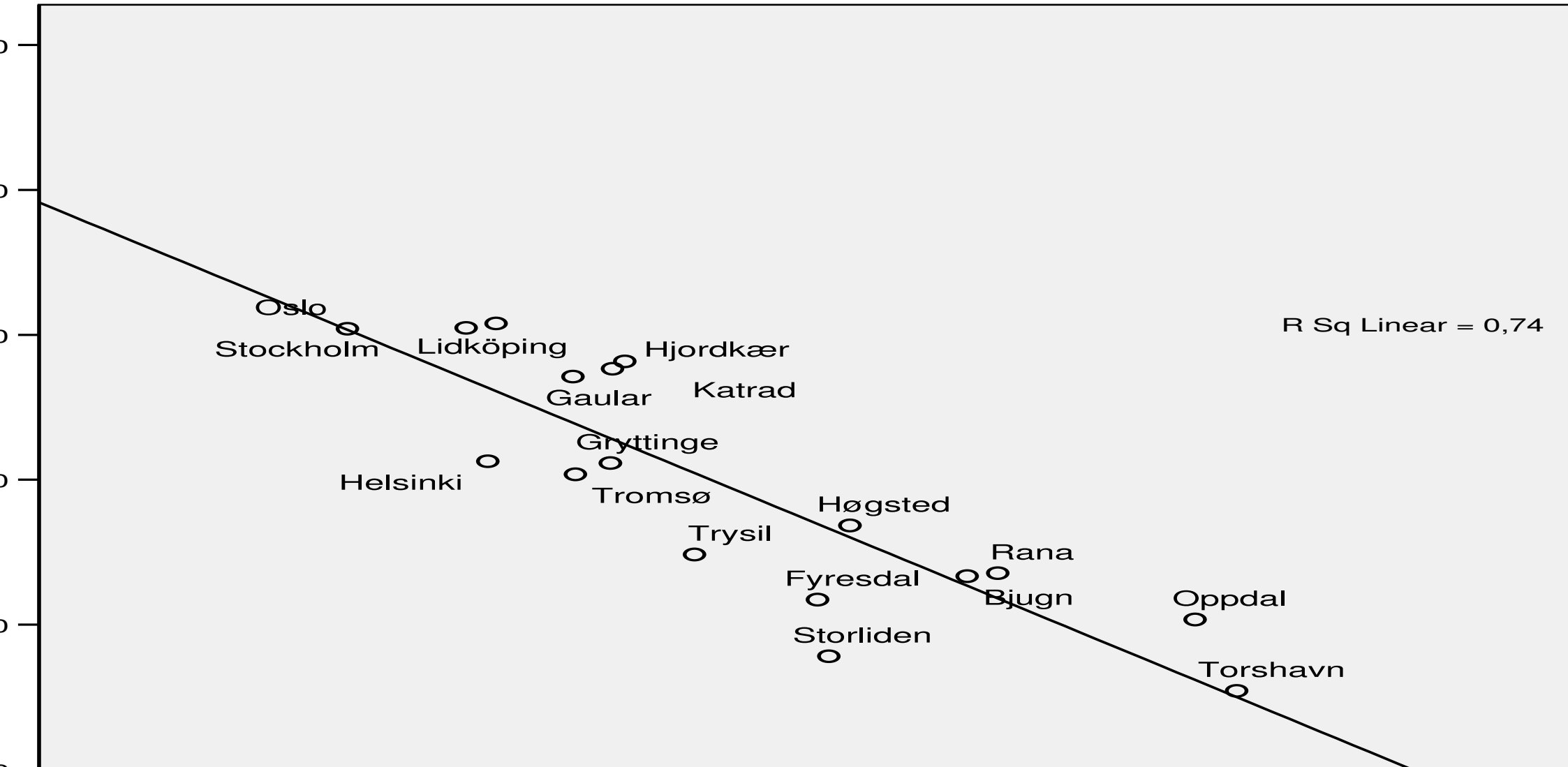
- > percentage of non-cognates words between Standard Danish and each of the other varieties.
- > the larger the lexical distance, the less intelligible another Scandinavian variety will be for a speaker of Standard Danish.

# research question 1

- > What is the relative contribution of (full) phonetic distance and lexical distance to intelligibility?

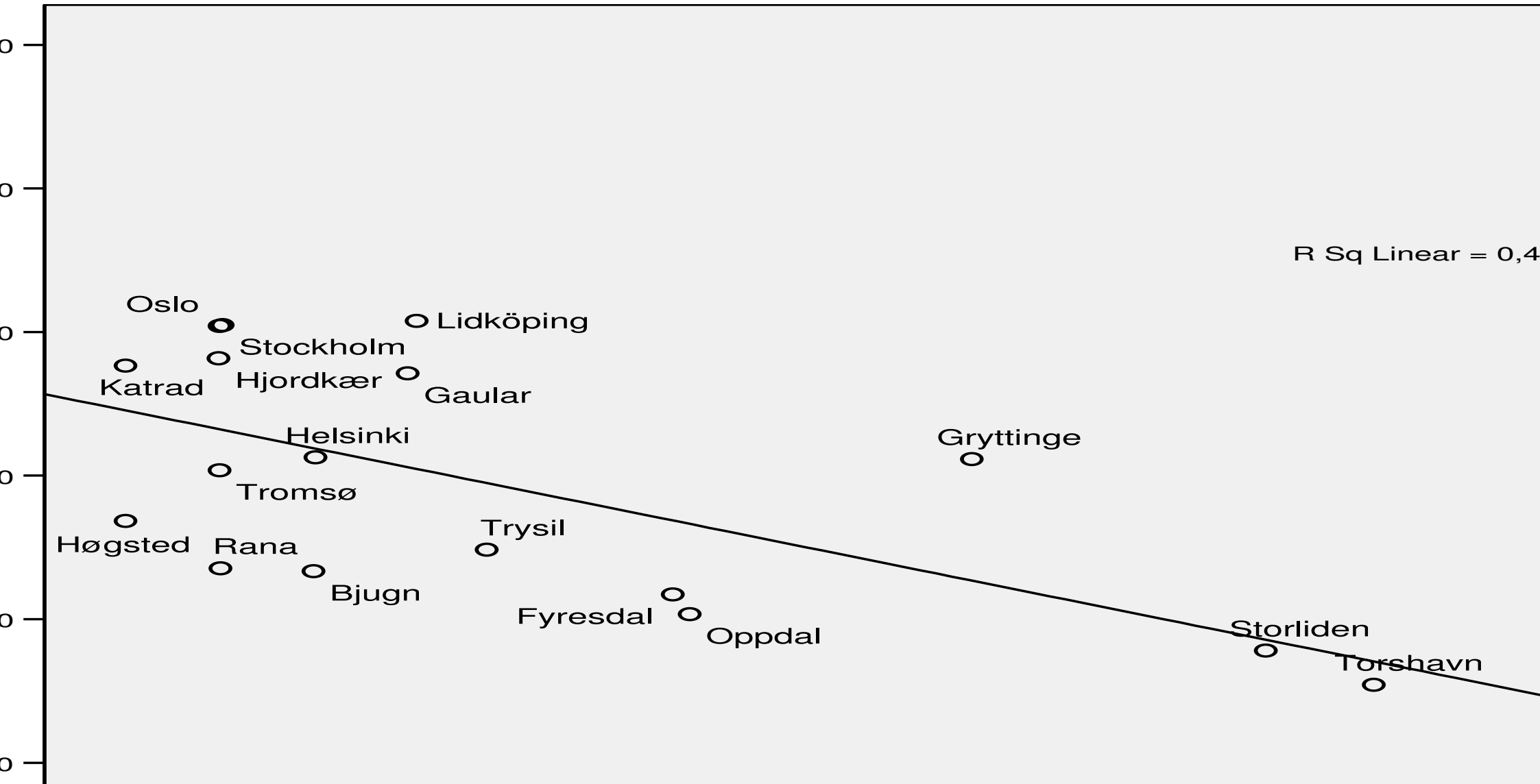
## full phonetic distance and intelligibility

$$r = -.86 \text{ and } p < 0.01$$



# lexical distance and intelligibility

$r = -.64$  and  $p < 0.01$



<b>linguistic distance</b>	<b>intelligibility</b>
full phonetic distance	-.86**
lexical distance	-.64**
full phonetic distance + lexical distance	-.90**

\*\* . Correlation is significant at the 0.01 level (2-tailed)

Stepwise regression analysis excludes: lexical distance

# research question 2

- > What is the relative contribution of consonant distances and vowel distances to intelligibility?

<b>phonetic distance</b>	<b>intelligibility</b>
vowel distances	-.65**
consonant distances	-.67**

\*\* . Correlation is significant at the 0.01 level (2-tailed)



# research question 3

- > What is the relative contribution of substitution, insertion, deletion, lengthening and shortening of both vowels and consonants to intelligibility?

<b>phonetic distance</b>	<b>intelligibility</b>
vowel substitution	.11
vowel insertion	-.39
vowel deletion	-.44
vowel lengthening	-.42
vowel shortening	-.49*

\*.Correlation is significant at the 0.05 level (2-tailed)

<b>phonetic distance</b>	<b>intelligibility</b>
consonant substitution	-.57*
consonant insertion	-.20
consonant deletion	-.22
consonant lengthening	.07
consonant shortening	-.22

\*.Correlation is significant at the 0.05 level (2-tailed)

# regression analyses

method	entered variables	model	significance
overall	all phonetic sublevels	all phonetic sublevels	$r = .93$ $p = .050$
pairwise	all phonetic sublevels	consonant sub., vowel ins., vowel short.	$r = .87$ $p = .000$

# conclusions

- > full phonetic distance is a good predictor of intelligibility:  
 $r = -.86^{**}$
- > lexical distance is a moderately good predictor of intelligibility:  $r = -.64^{**}$
- > the relative contribution of consonant distances and vowel distances to intelligibility is about equal:  
 $r = -.67^{**}$  versus  $r = -.65^{**}$

# conclusions

- > consonant substitutions:  $r = -.57^*$ 
  - change in reference points of a word.
  
- > vowel shortening:  $r = -.49^*$ 
  - loss of information about the vowel (quantity);  
vowel length is phonemically distinctive.
  
- > vowel insertion:  $r = -.39$ 
  - insertions change syllable structure / may  
result in diphthongs.

Thank you for  
your attention