rijksuniversiteit
groningen

Mutual intelligibility between closely related languages

Charlotte Gooskens

Renée van Bezooijen

Charles University in Prague, 7 October 2009

Overview

- background
- intelligibility testing
- factors determining intelligibility
- four investigations
- future research

Background

Background

- ***Linguistic determinants of mutual intelligibility in Scandinavia***
- Financed by NWO (The Netherlands Organisation for Scientific Research)
- 1 January 2006 – 1 January 2011
- Members of the project group:
  - Nanna Haug Hilton
  - Anja Schüppert
  - Renée van Bezooijen
  - Charlotte Gooskens
  - student assistants
  - Sebastian Kürschner
  - Vincent van Heuven
- http://www.let.rug.nl/~gooskens/project/

Background

Semicommunication

- Haugen (1966)
- ≈ nonconvergent/asymmetric/bilingual discourse, receptive bilingualism
- Speakers of different but related languages each speak their own language and still comprehend each others' languages
- Mutual intelligibility is sometimes imperfect and asymmetric

Background

Observed semicommunication

- Danish - Norwegian - Swedish (Haugen 1966, Maurud 1976….)
- Czech - Slovakian (Budovičá 1987)
- Czech - Polish (Hansen 1987)
- Spanish - Portuguese (Coseriu 1988, Jensen 1989, Zeevaert 2002)
- Italian - Spanish (Hansen 1987)
- German - Dutch (Haz 2002)
- Frisian - Dutch (Feitsma 1986)
- Croatian - Serbian (Haugen 1990)
- Hindi - Urdu (Haugen 1990)
- Icelandic - Faeroese (Braunmuller & Zeevaert 2001)
- Macedonian - Bulgarian (Haugen 1990)
- Russian - Bulgarian (Braunmuller & Zeevaert 2001)

Background

Central questions

- How can the mutual intelligibility between closely related languages be measured?
- How can the relevant (extra-)linguistic factors be measured?
- To what extent are the (extra-)linguistic factors predictors of intelligibility?

Intelligibility

Measuring intelligibility

- **Opinion testing:**
  How well does the listener **think** he understands the other language variety?

- **Functional testing:**
  How well does the listener **actually** understand the other language variety?

- **Observations:**
  How well do people understand each other in **real** language situations?

Measuring intelligibility

Type of research	Type of speech	Method
opinion test	• spontaneous speech • read texts • isolated words • …	• open questions • multiple choice • translations • reaction times • …
functional test	• spontaneous speech • read texts • isolated words • …	• opinion scales
observations	• spontaneous speech	• counting phenomena

Factors explaining intelligibility

Factors explaining intelligibility

- Non-linguistic
- Linguistic (distances)

Factors explaining intelligibility

Non-linguistic

- attitude
- contact
- orthography

Explaining factor: attitude

What do you think of the Danish language?

beautiful ☐ ☐ ☐ ☐ ☐ ugly

What do you think of the Danes?

kind ☐ ☐ ☐ ☐ ☐ unkind

Would you like to live in Denmark?

yes ☐ maybe ☐ no ☐

Explaining factor: attitude

- It is plausible that in real life attitudes play a role in intelligibility
- Still statistic relationships have hardly been found
- The test situation is likely to block attitudes

Explaining factor: contact

I watch Danish television...

- once a week ☐
- once a month ☐
- once a year ☐
- less often ☐

I meet Danes...

I am in Denmark...

Explaining factor: contact

- It is certain that contact plays a role in intelligibility
- Still statistic relationships have hardly been found
- Scandinavians have little contact

Explaining factor: orthography

Danish	Swedish	
hund [hun]	*hund* [hund]	'dog'

Factors explaining intelligibility

Linguistic distances

- **pronunciation**
- **lexicon**
- morphology
- syntax

Explaining factor: phonetic distance

Levenshtein algorithm

- Heeringa (2004)
- Measures the phonetic distance between related language varieties
- Compares the sounds of cognate word pairs
- Counts how many sounds minimally must be substituted, added or removed in order to change the sounds of one word into the sounds of another word
- Total distance is obtained by summing word distances

Explaining factor: phonetic distance

Example Levenshtein distance:

Danish *ligne* vs. Swedish *likna* ‘be like’:

l	iː		n	ə
l	iː	k	n	a
		1		1

(1+1=2)/5 = 40% difference

Explaining factor: lexical distance

Percentage of non-cognates

Example of a non-cognate:

› Danish *pige* vs. Swedish *flicka* ‘girl’

Investigations

Investigations

- Text comprehension with open questions
- Intelligibility of isolated words
- Intelligibility among children
- The role of syntax in intelligibility

Text comprehension with open questions

Text comprehension with open questions

- The project *Internordisk sprogforståelse*, INS (*Inter-Nordic comprehension*)
- Supported by the Nordic Cultural Fund
- Test groups from all Nordic countries
- Intelligibility of the three mainland Scandinavian languages (Danish, Norwegian and Swedish) is tested
- Questionnaire about attitudes and contact

Text comprehension with open questions

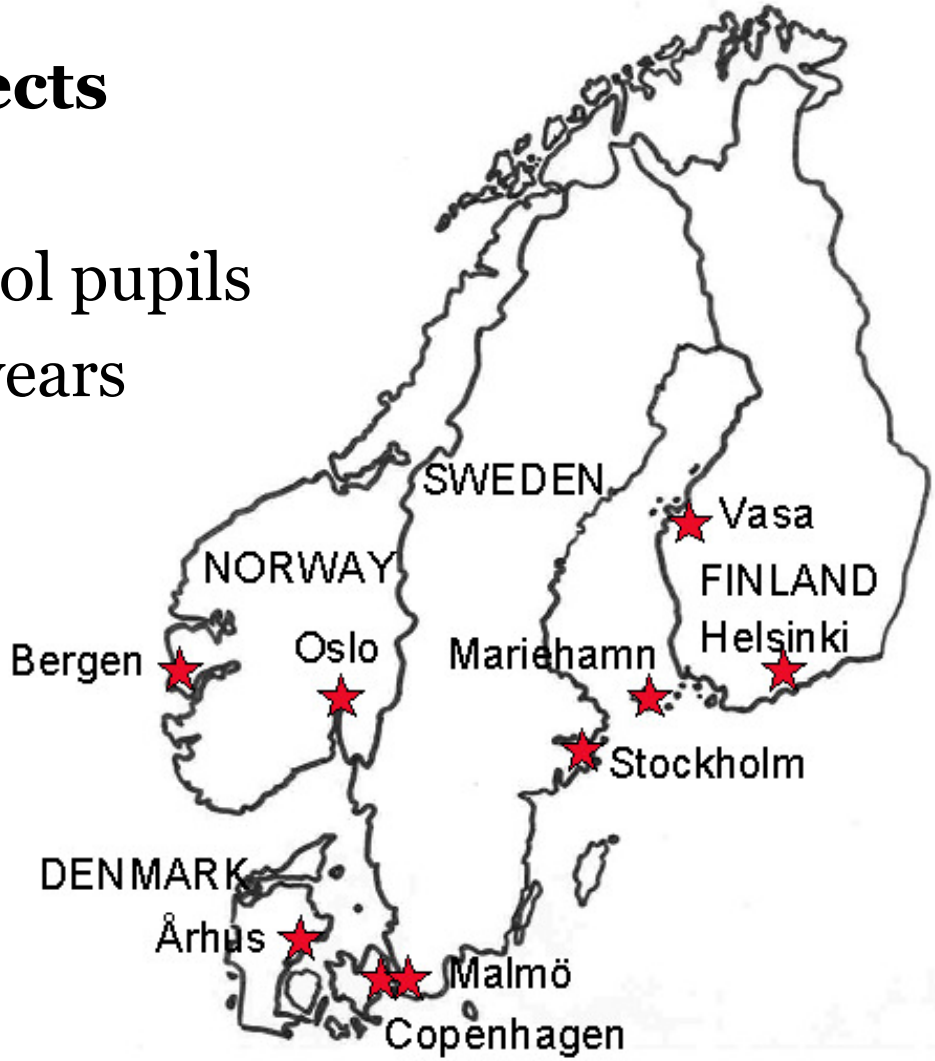
Our contribution to INS:

- To investigate to what extent differences in inter-Scandinavian intelligibility can be explained by linguistic distances
- To compare results to intelligibility between speakers of Dutch, Frisian and Afrikaans

Text comprehension with open questions

Scandinavian subjects

- 690 secondary school pupils
- between 16 and 19 years



Text comprehension with open questions

Design of Scandinavian study

Subjects from	Test language		
	Danish	Norwegian	Swedish
Denmark:			
Århus	-	33	59
Copenhagen	-	62	44
Norway:			
Bergen	47	-	40
Oslo	57	-	84
Sweden:			
Malmö	44	43	-
Stockholm	41	47	-
Finland:			
Mariehamn	22	25	-
Vasa	-	12	-
Helsinki	9	21	-

Text comprehension with open questions

West-Germanic subjects

- 81 secondary school pupils
- between 16 and 17 years

Text comprehension with open questions

Design of West-Germanic study

Subjects:	Test language		
	Dutch	Frisian	Afrikaans
Dutch	-	16	16
Frisian	-	-	17
Afrikaans	15	17	-

Text comprehension with open questions

Method

- News item (250 words)
- Read aloud in test language
- 5 open questions
- Intelligibility was expressed as percentage of correctly answered questions per test group

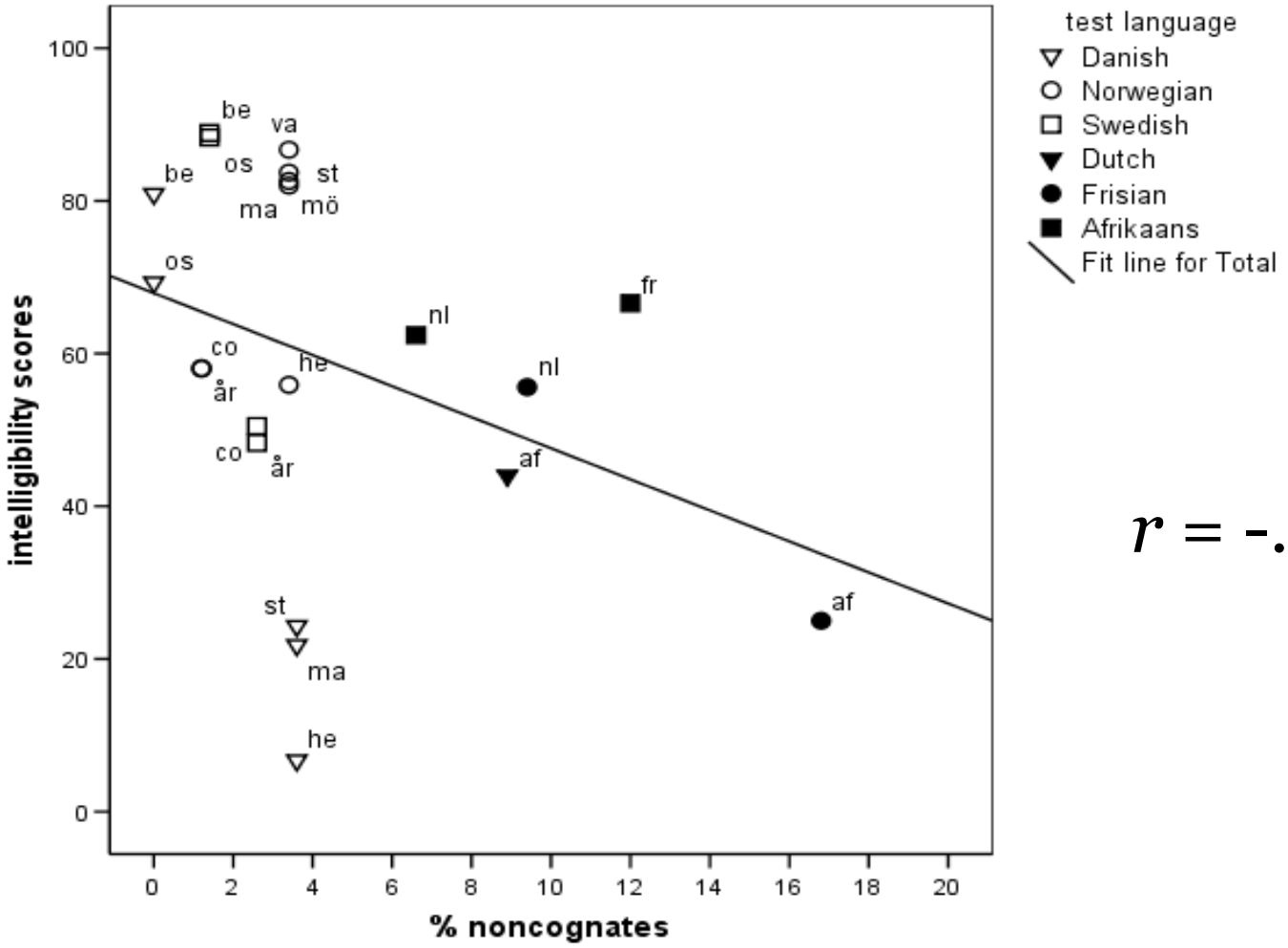
Text comprehension with open questions

Linguistic distances

- Extra recordings from each of the nine towns in Scandinavia
- All Scandinavian and West-Germanic recordings were transcribed phonetically
- Phonetic and lexical distances were calculated

Text comprehension with open questions

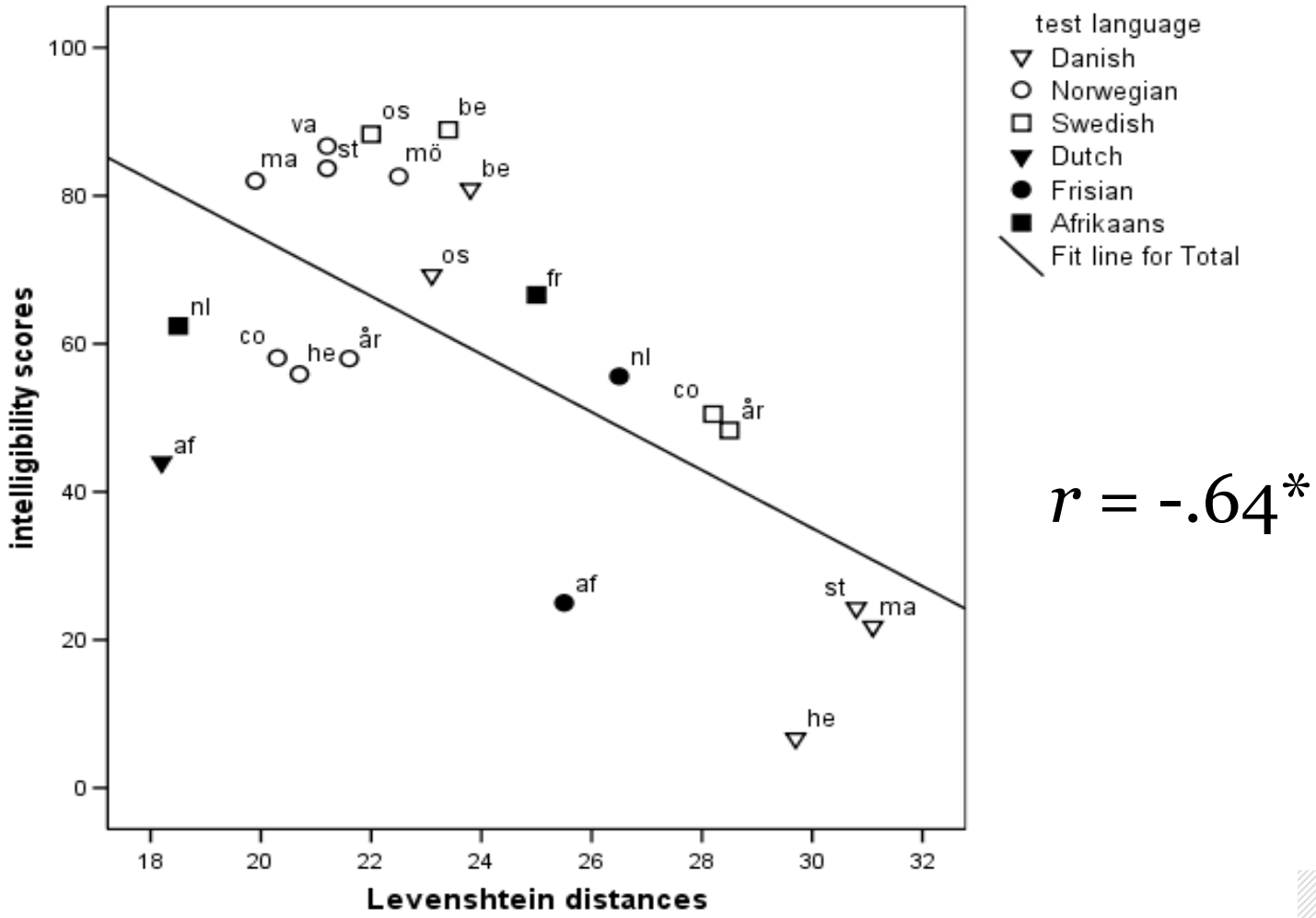
Correlation between intelligibility and lexical distances



$r = -.36$

Text comprehension with open questions

Correlation between intelligibility and phonetic distances



$r = -.64^{*}$

Text comprehension with open questions

Large advantage:

› Realistic (ecologically valid)

Disadvantages:

› Difficult to assess representativity of stimulus text
› Single unintelligible words may have a large effect on intelligibility
› Difficult to construct open questions
› Difficult to decide when a question is correctly answered
› Many varieties needed in order to be able to correlate intelligibility with linguistic distances
› Provides only an overall impression of the role of phonetic and lexical distances for the intelligibility

Intelligibility of isolated words

Intelligibility of isolated words

- Internet experiment involving seven languages in the Germanic language area

- Word comprehension

Intelligibility of isolated words

Research questions

- How well do speakers of various Germanic languages understand each other's vocabularies?
- Which linguistic factors play a role in the intelligibility?

Intelligibility of isolated words

Test words

- 384 nouns
- Randomly selected from a list of 2575 highly frequent spoken words
- Translated into seven languages
- Recordings of standard language speakers

Intelligibility of isolated words

Test words are non-cognates or cognates

Example non-cognate:

Du. *lichaam*, De. *krop*, 'body'
Fr. *fyts*, Zw. *cykel*, De. *Fahrrad* 'bicycle'

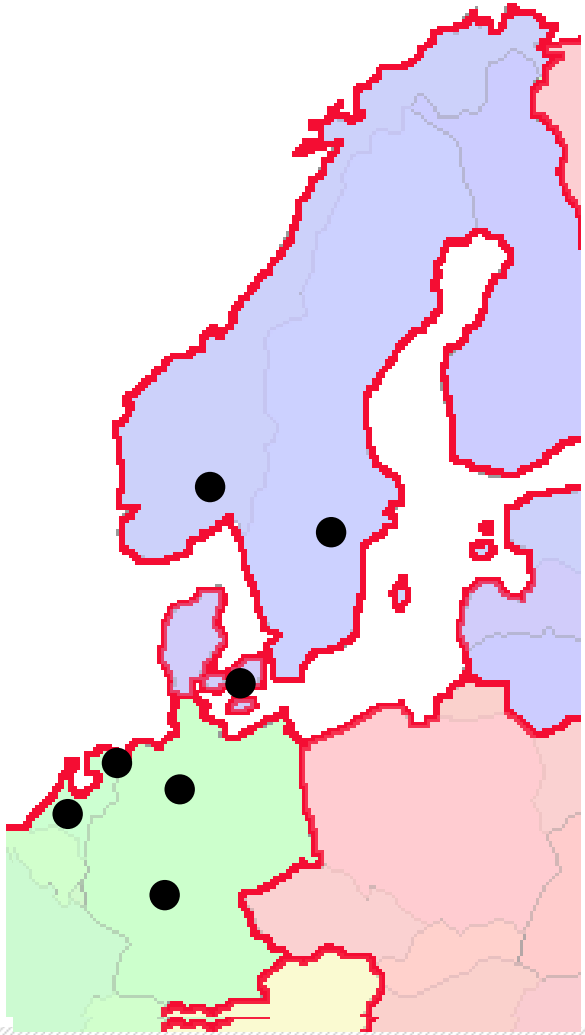
Examples cognates:

Fr. *strategy*, Du. *strategie*, De. *strategi* 'strategy'
Ge. *Weg*, Sw. *väg*, Fr. *wei* 'road'

Intelligibility of isolated words

Recordings

- Swedish (Sw)
- Norwegian (No)
- Danish (Da)
- German (Ge)
- Low German (LG)
- Frisian (Fr)
- Dutch (Du)



Intelligibility of isolated words

Listeners	Test language						
	Da	Sw	No	LG	Ge	Fr	Du
Sw	x		x				
No	x	x					
Da		x	x	x	x	x	x
Fr	x						
Du	x			x	x		

Intelligibility of isolated words

Listeners	Test language						
	Da	Sw	No	LG	Ge	Fr	Du
Sw	x		x				
No	x	x					
Da		x	x	x	x	x	x
Fr	x						
Du	x			x	x		

Intelligibility of isolated words

Listeners	Test language						
	Da	Sw	No	LG	Ge	Fr	Du
Sw	x		x				
No	x	x					
Da		x	x	x	x	x	x
Fr	x						
Du	x			x	x		

Intelligibility of isolated words

Listeners	Test language						
	Da	Sw	No	LG	Ge	Fr	Du
Sw	x		x				
No	x	x					
Da		x	x	x	x	x	x
Fr	x						
Du	x			x	x		

Intelligibility of isolated words

Dutch	*Ik weet dat hij naar huis komt*
Frisian	*Ik wit dat er thús komt*
German	*Ich weiß dass er nach Hause kommt*
Danish	*Jeg ved at han kommer hjem*
Swedish	*Jag vet att han kommer hem*
Norwegian	*Jeg veit at han kommer hjem*

‘I know that he comes home’

Intelligibility of isolated words

Listeners

- 1400 high school pupils
- 15-19 years
- Speaking test language at home

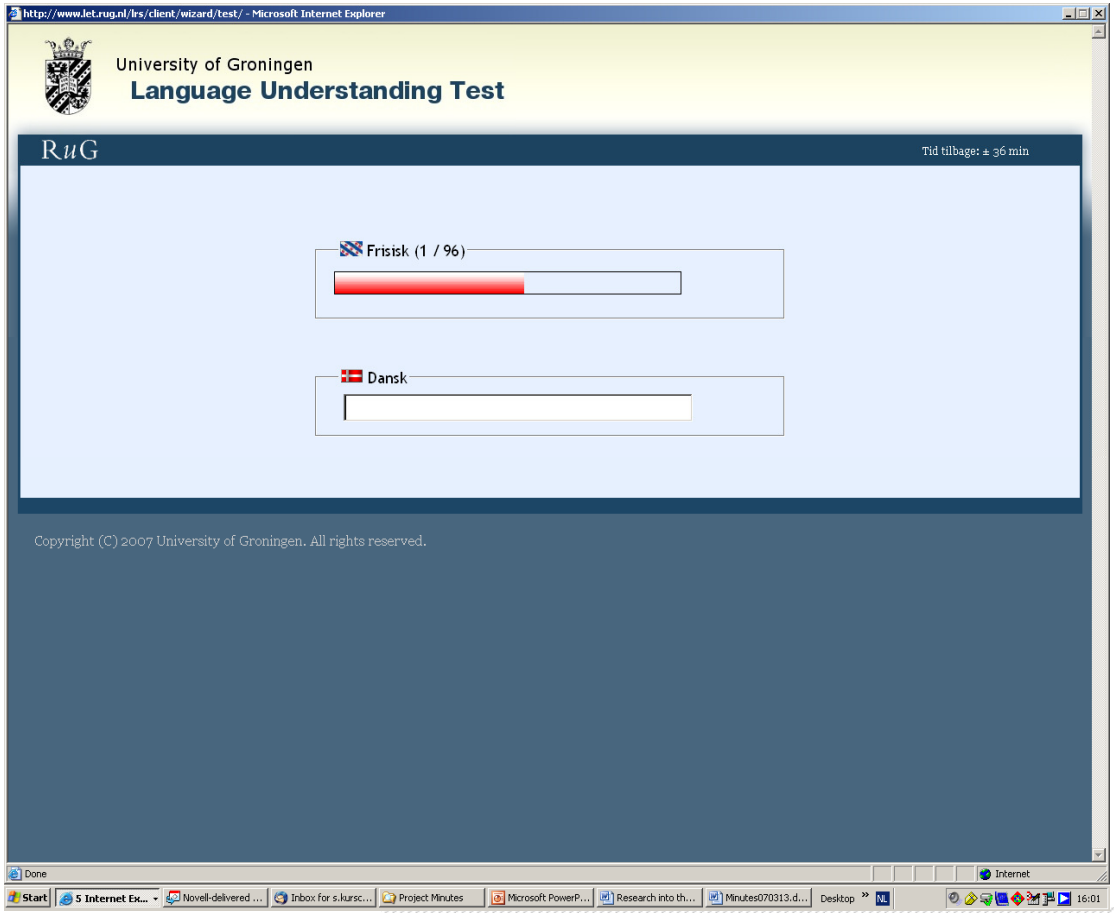
Intelligibility of isolated words

Procedure

- Internet-based
- Subjects listened to 96 words via head phones
- Translations into mother tongue

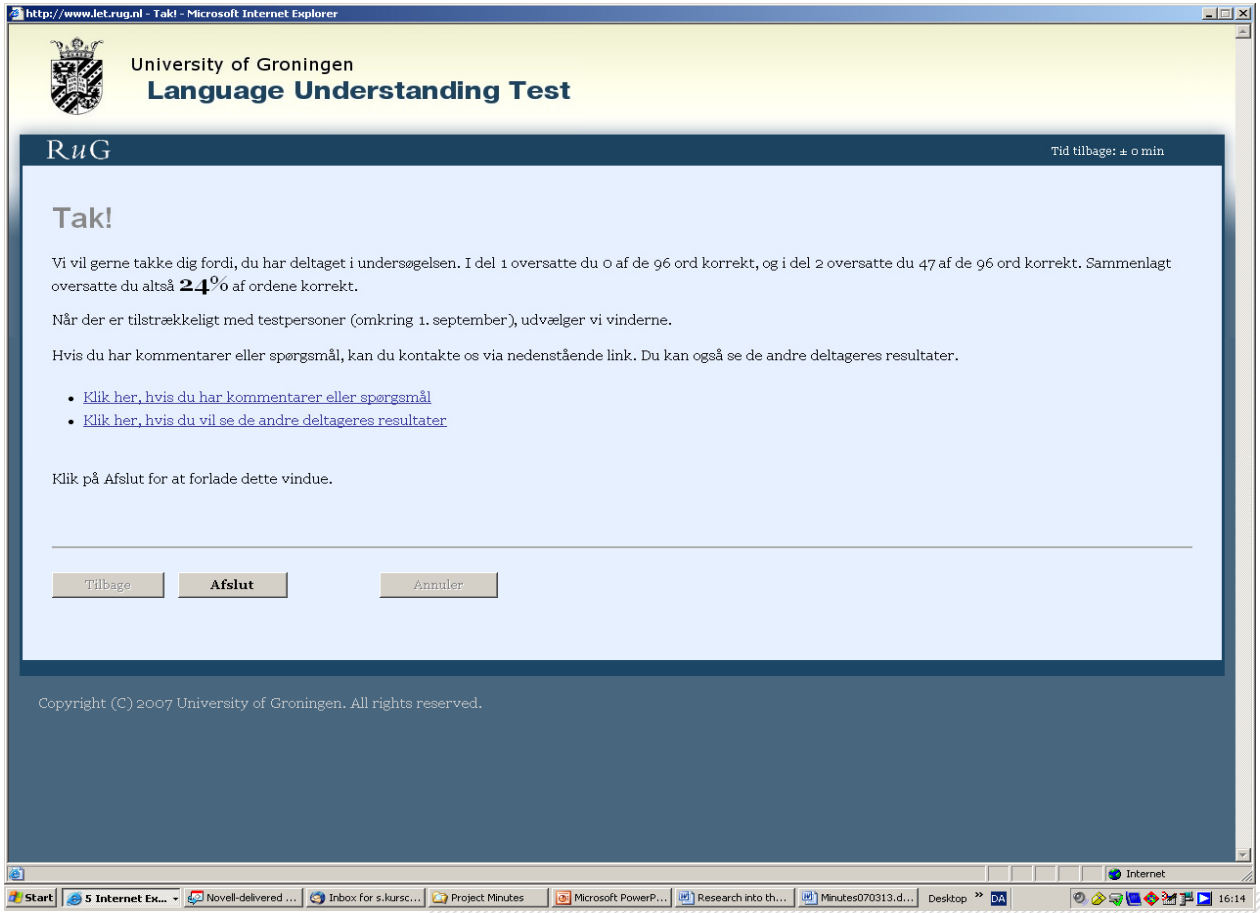
- http://www.let.rug.nl/lrs
  login: germanic
  password: guest

Internet-experiment



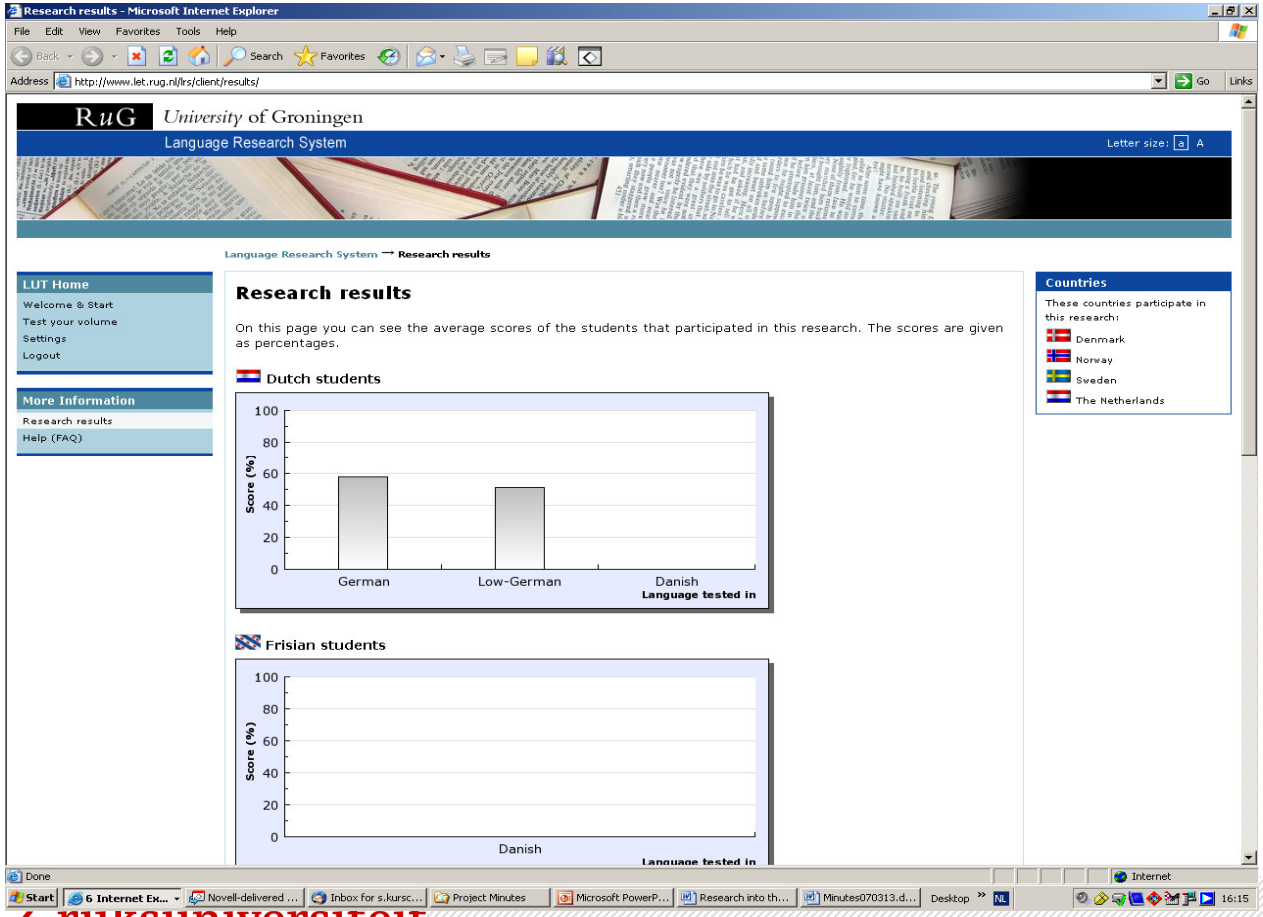
Internet-experiment

The result of the test person is shown:



Internet-experiment

The results of all test persons are shown:



Intelligibility of isolated words

Calculations

Intelligibility = percentage of correctly translated words

- Ignoring spelling mistakes

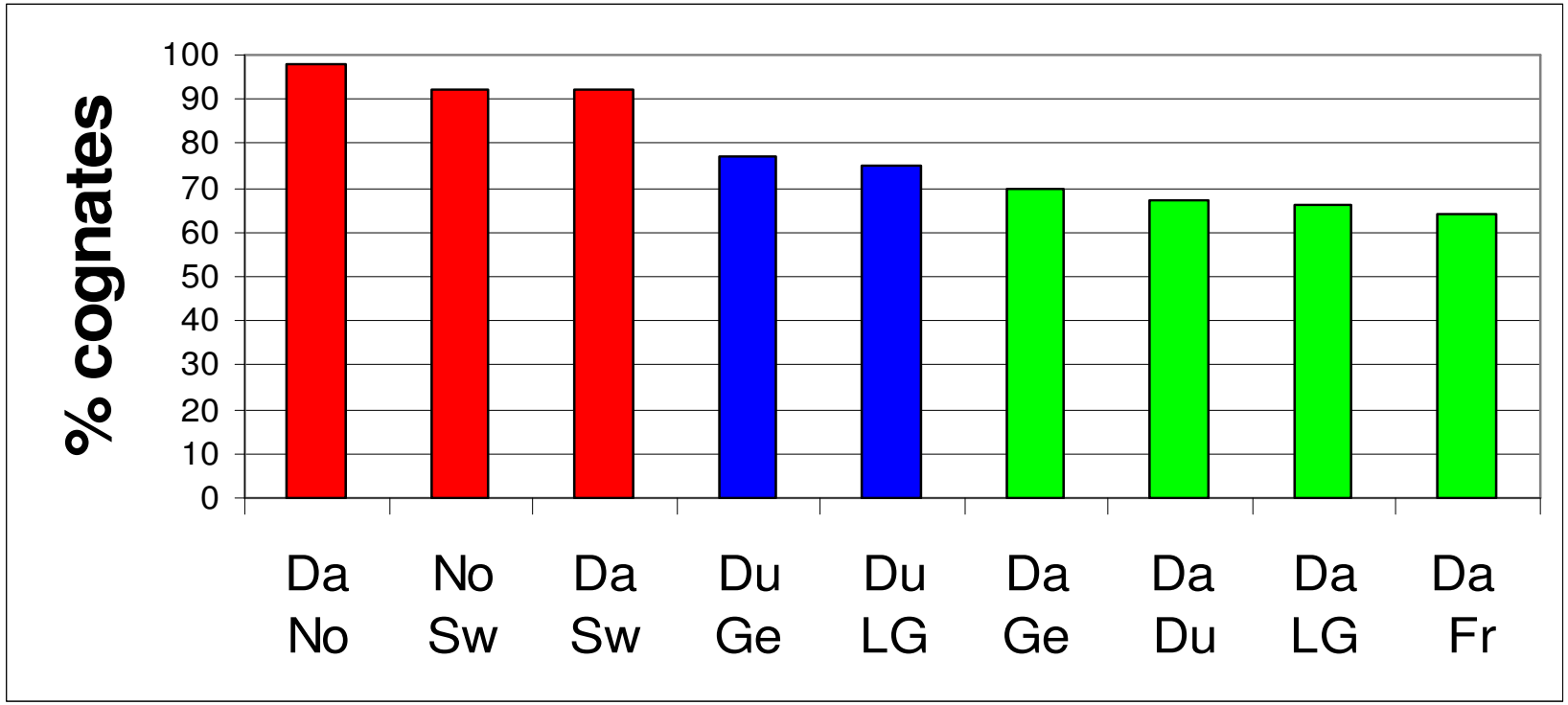
  ex. Dutch *kultuur* instead of *cultuur* for Danish *kultur* 'culture'

- Allowing alternative translations

  ex. Dutch *winkel* or *boetiek* for Low German *Laden* 'shop'

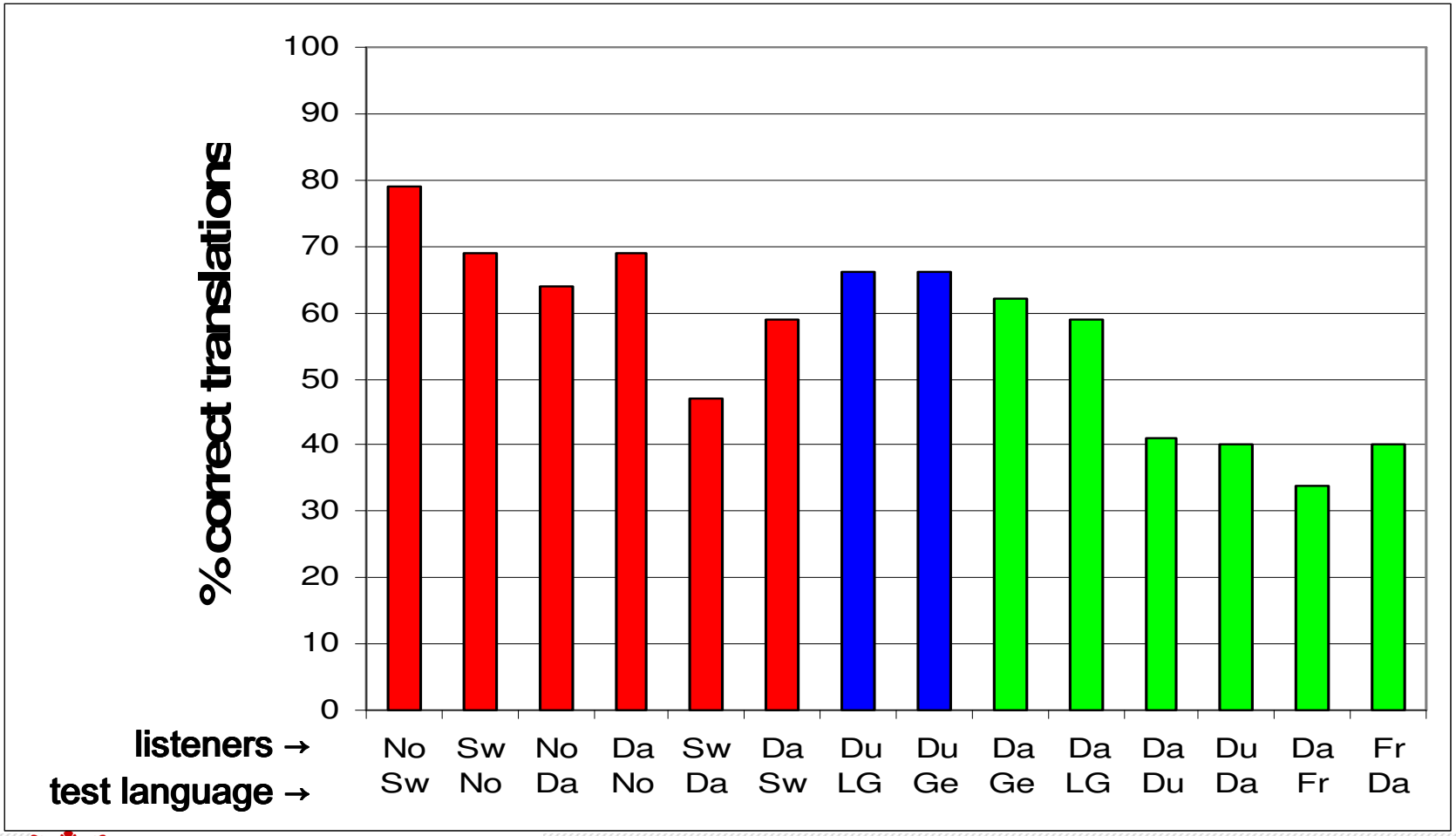
Intelligibility of isolated words

% cognates per language pair



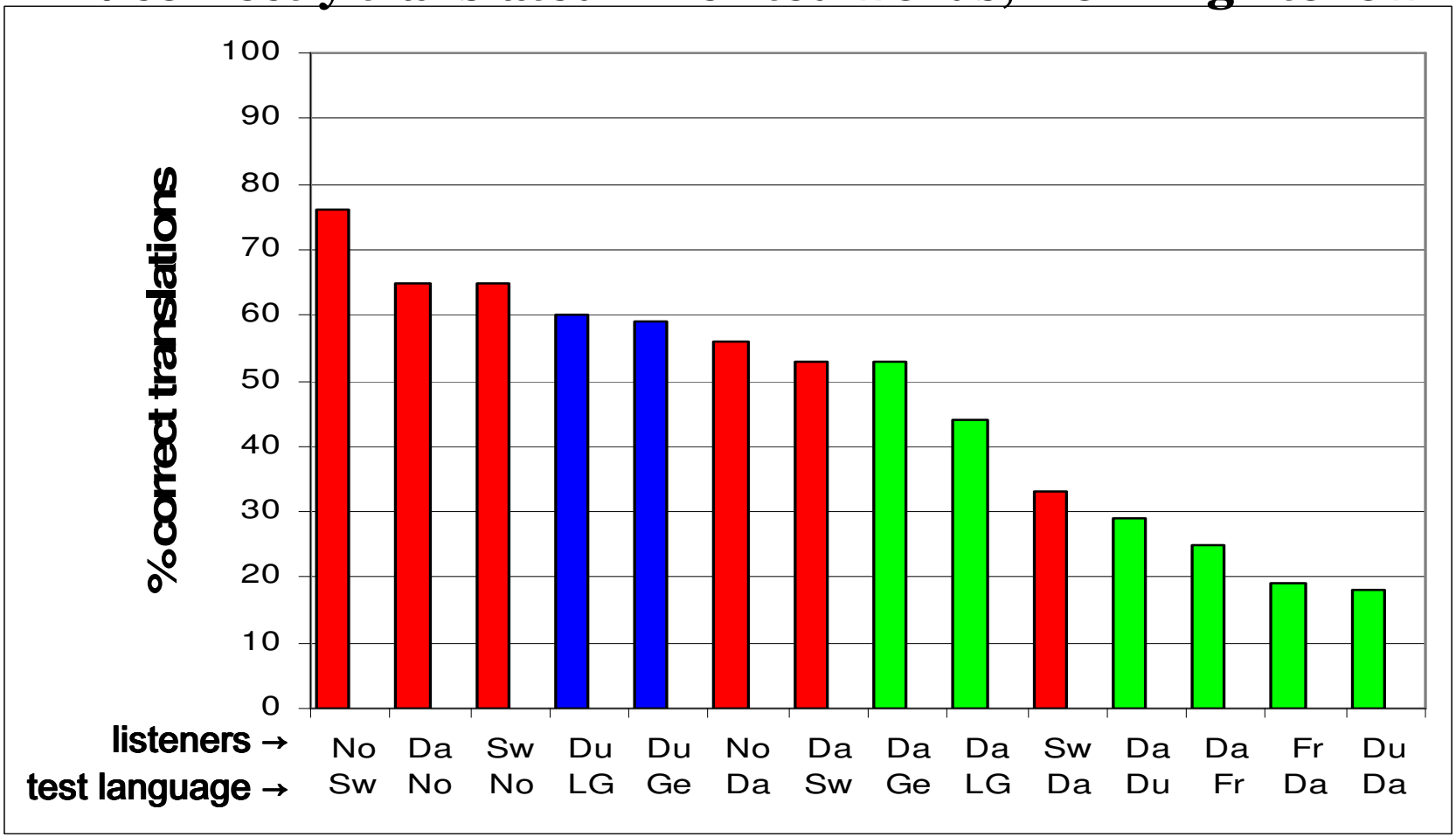
Intelligibility of isolated words

% correctly translated cognates:



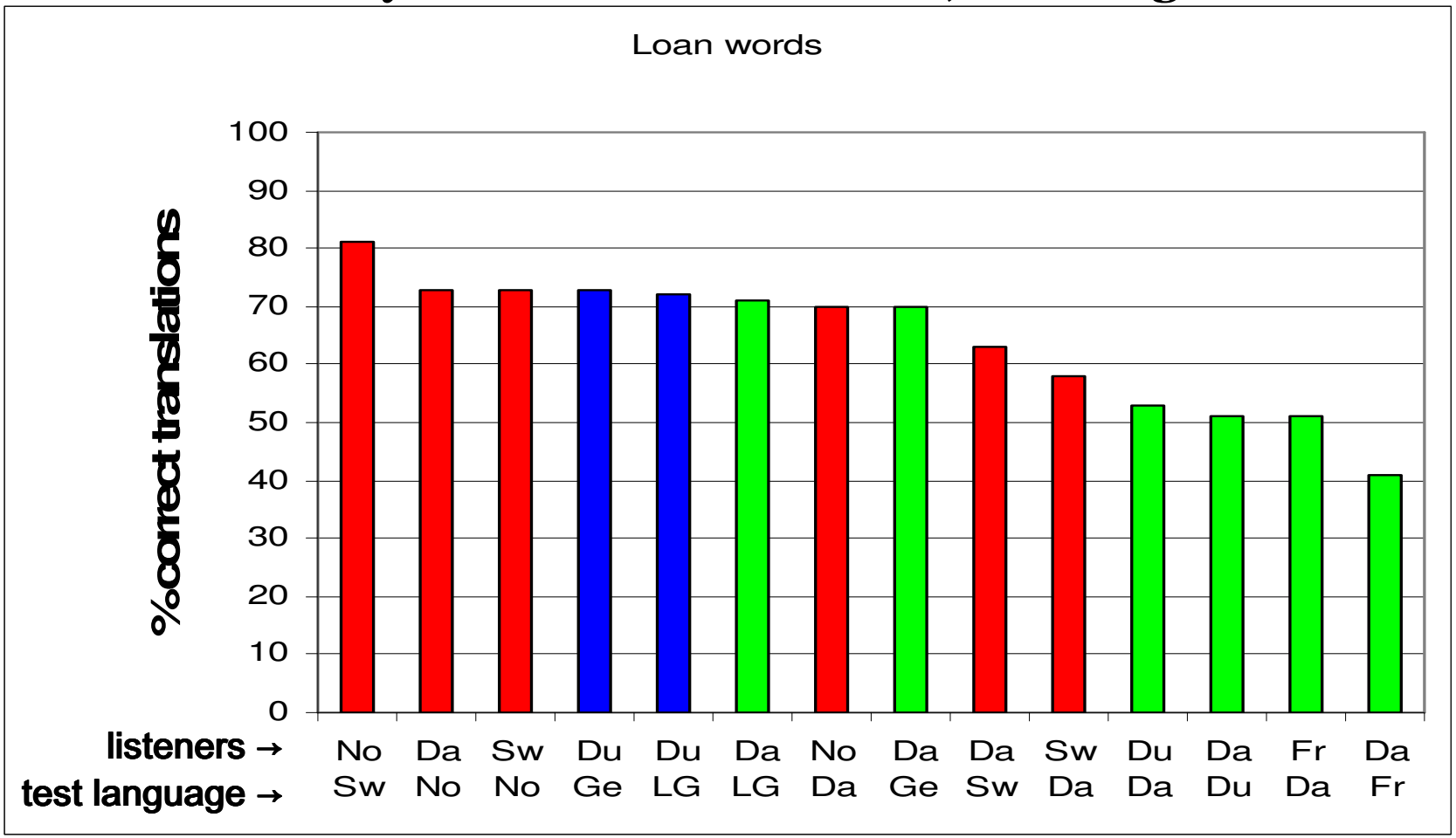
Intelligibility of isolated words

% correctly translated inherited words, from high to low



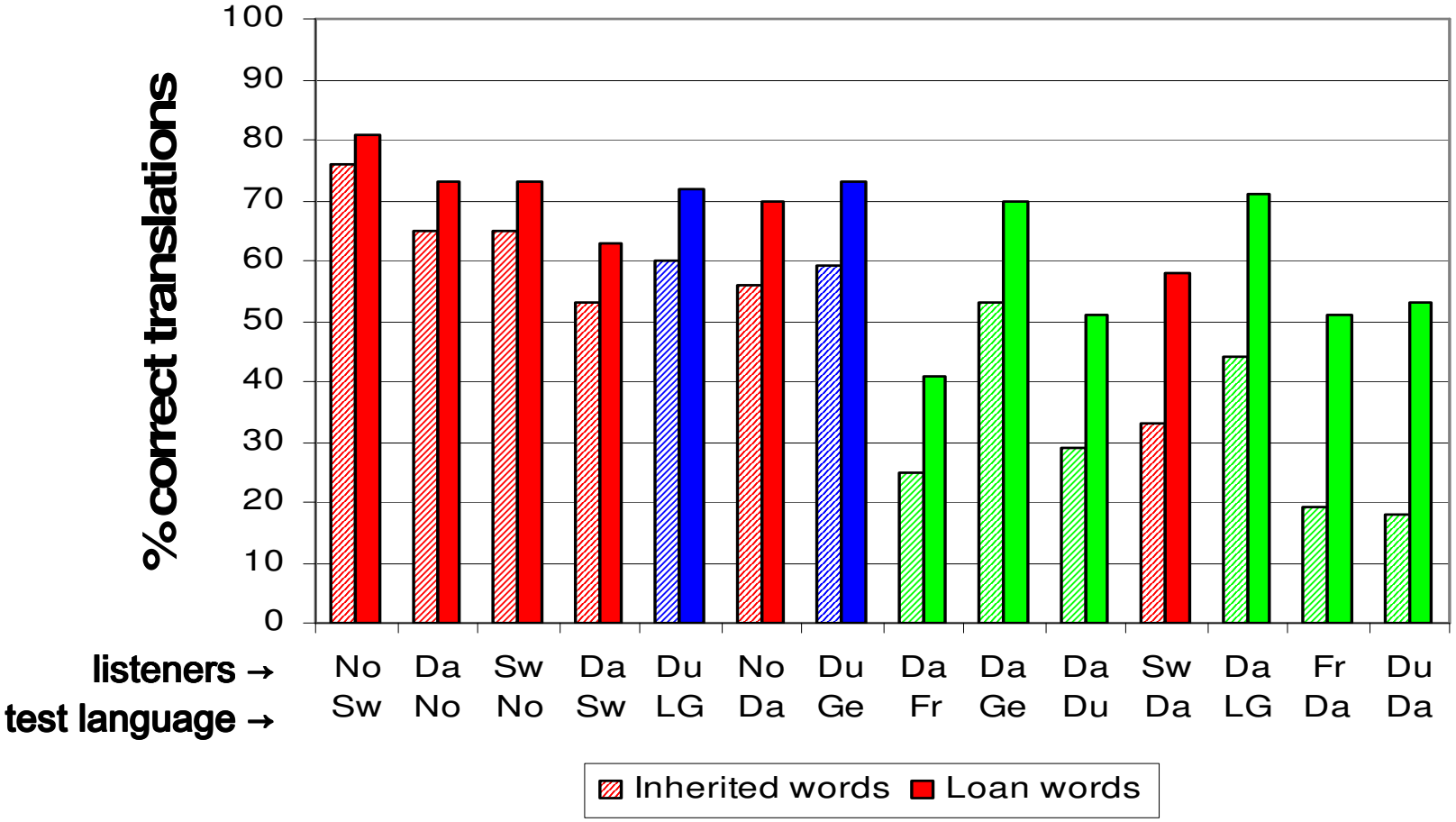
Intelligibility of isolated words

% correctly translated loan words, from high to low



Intelligibility of isolated words

From small to large difference between inherited and loan words



Intelligibility of isolated words

11 factors considered for prediction of intelligibility of Swedish words by Danish listeners

- **Levenshtein distance**
- **Foreign sounds**
- **Word length**
- Word stress differences
- **Differences in number of syllables**
- Lexical tones
- **Stød**
- **Neighbourhood density**
- Etymology (native words versus loan words)
- **Orthography**
- **Word frequency**

Intelligibility of isolated words

Analysis of errors can give information about listener strategies

- Phonetic confusions of sounds
  - e.g. Sw. /k/ is often perceived as /g/ by Danes
  - Sw. *klass*, Da. *klasse* 'class' is translated into Da. *glas* 'glass'
- Influence of neighbour words
  - e.g. Sw. *kør*, Da. *kor* 'choir' is often translated into Da. *kør* 'drive'
- Interference from foreign languages
  - e.g. Sw. *hot*, Da. *trussel* 'threat' is often translated into Da. *varm* 'hot'

Intelligibility among children

Intelligibility among children

Ph.D.-project Anja Schüppert:

Nonlinguistic factors can be neutralised by testing young children:

- Cannot read
- Have no knowledge of foreign languages
- Have had little contact with neighbouring countries
- Have less strong attitudes towards neighbouring languages

Intelligibility among children

Subjects

- 19 Danish children from Odense, 4-6 years old
- 26 Swedish children from Vaxjö, 4-6 years old

- 20 Danish adults from Odense
- 19 Swedish adults from Vaxjö

Intelligibility among children

Stimulus material

- 50 nouns (cognates) that are frequent in child language and early acquired
- Read alound in Danish and Swedish

Intelligibility among children

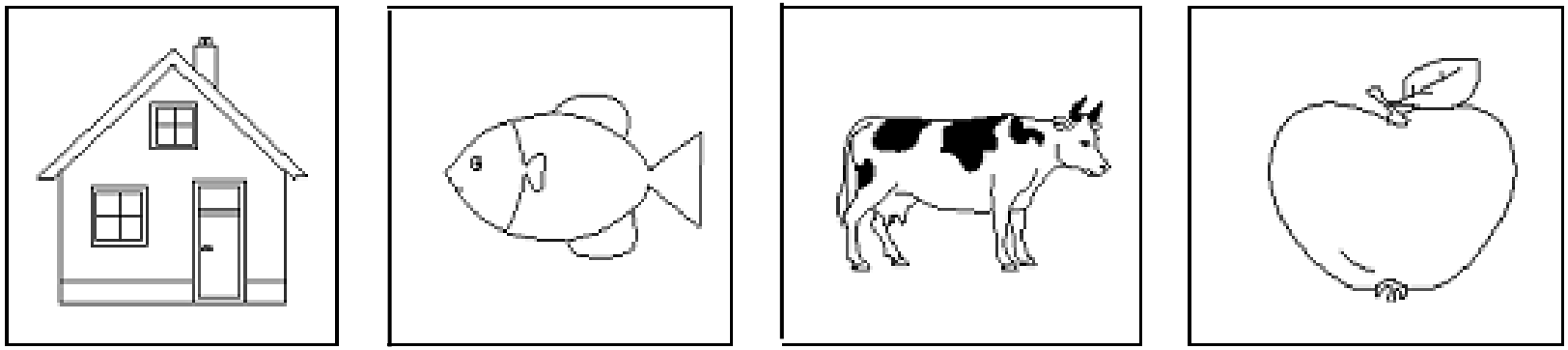
Test

- Danes listen to Swedish words and Swedes listen to Danish words
- Per word four pictures are presented
- Subject points to the picture corresponding to the test word on a touch screen
- Response time is measured

Intelligibility among children

Example

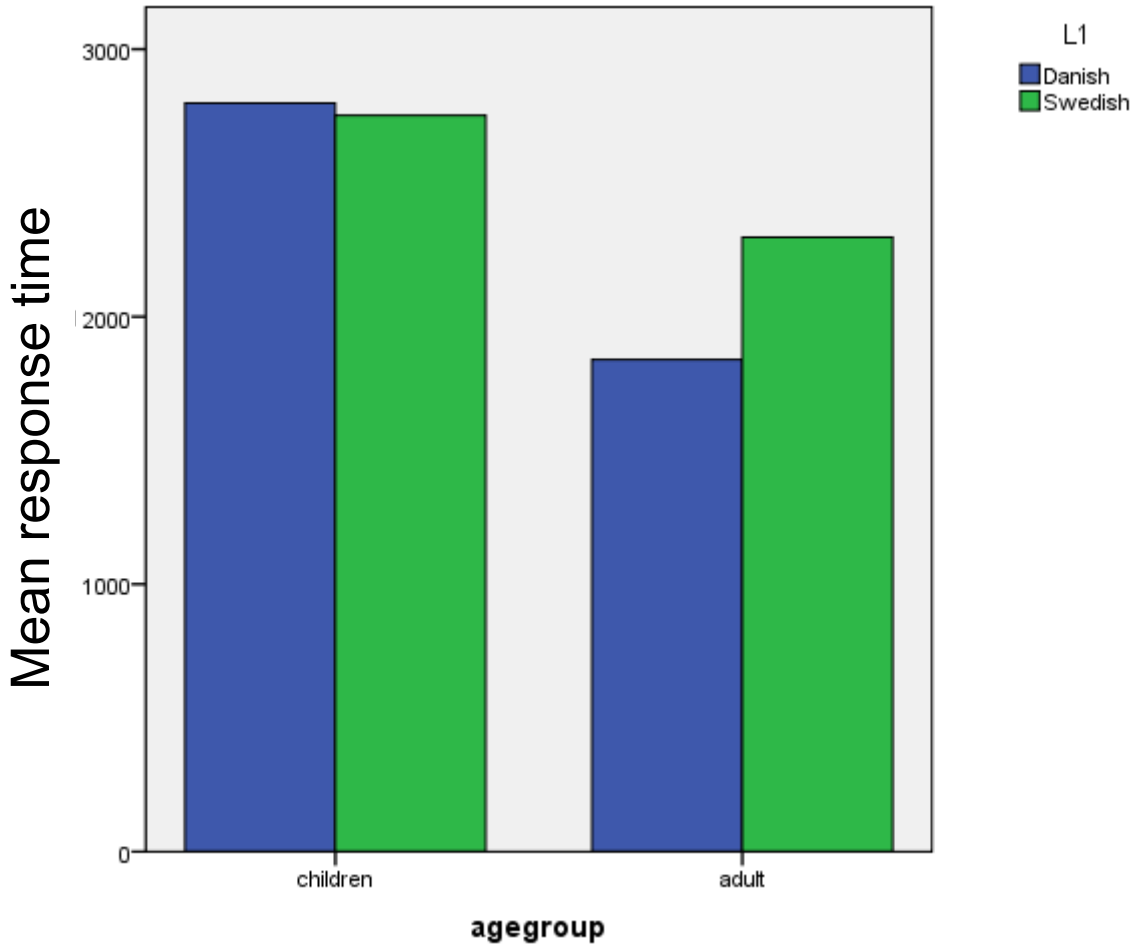
Swedish subject hears Danish *æble* ‘apple’ and sees the following pictures



The subject chooses rightmost picture

Intelligibility among children

Intelligibity results



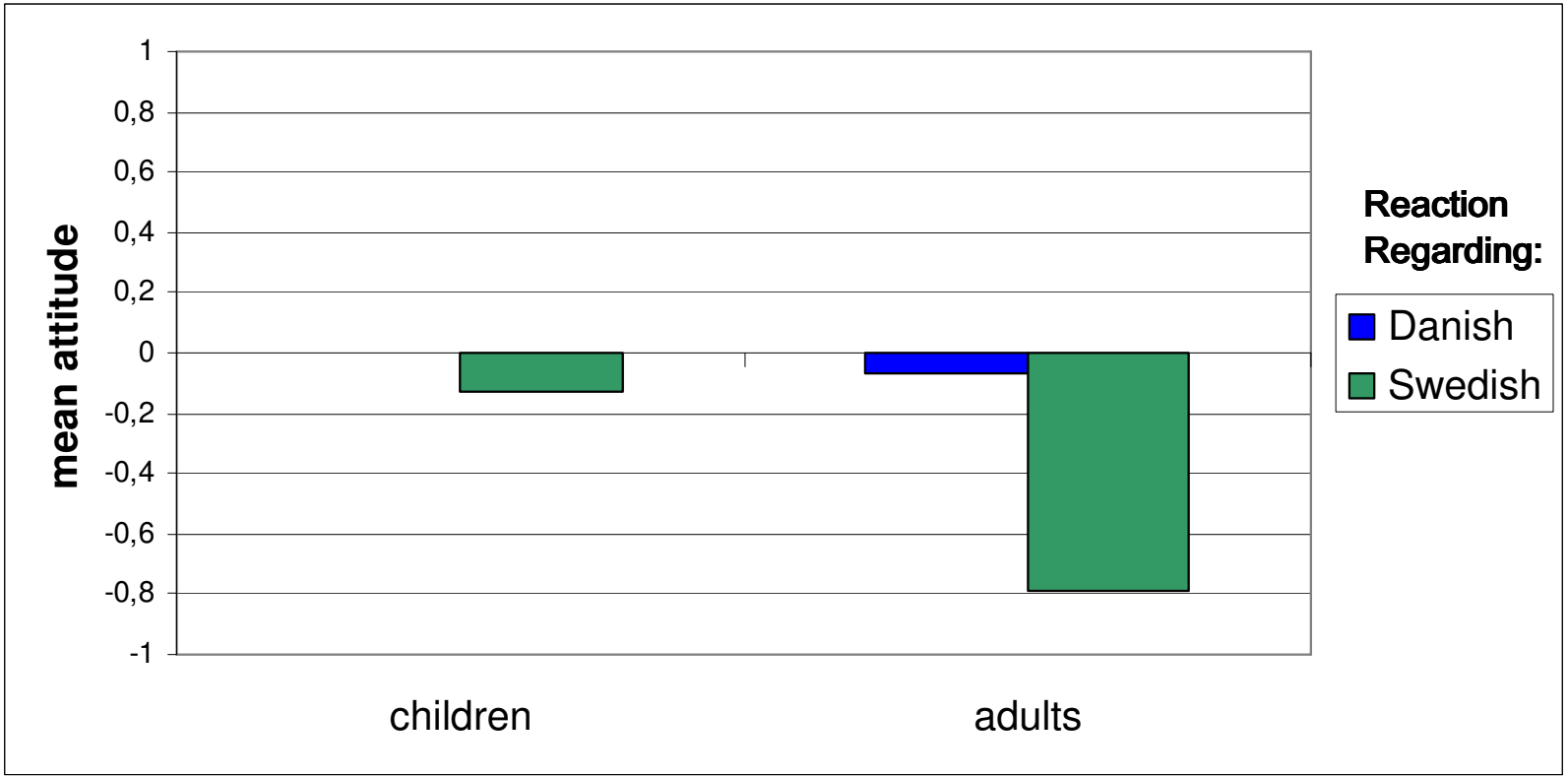
Intelligibility among children

Attitude elicitation

- Children and adults were asked if they thought the language sounded…
- …less nice than their native language (-1)
- …as nice as their native language (0)
- …nicer than their native language (1)

Intelligibility among children

Attitude results



Intelligibility among children

Plans

Measure intelligbility and attitudes among children of various ages

The role of syntax in intelligibility

The role of syntax in intelligibility

Research questions

- Do syntactic differences play a role in the intelligibility of a closely related language?
- What is the relative influence of syntactic versus phonetic differences on intelligibility?

The role of syntax in intelligibility

Research design

- Investigation tests the effect of idiosyncratic Norwegian syntactic constructions on Danes' comprehension of Norwegian.

- e.g. difference in particle placement

Norwegian:	Han	tok	av	brillene
	subj	*verb*	*part*	*obj*
Danish:	Han	tog	brillerne	af
	subj	*verb*	*obj*	*part*

'he took off the glasses'

The role of syntax in intelligibility

Research design

› Listeners are asked to decide whether the content of a sentence is plausible or not (binary choice)

e.g. implausible: *elefanten slog ordet op*
'the elephant looked up the word'

e.g. plausible: *journalisten skrev en artikel ud*
'the journalist printed the article'

› Response time is measured in addition to number of correct answers

The role of syntax in intelligibility

One informant hears:

	Without Noise		With Noise	
	Danish syntax	Norwegian syntax	Danish syntax	Norwegian syntax
Danish phonology	12	12	12	12
Norwegian phonology	12	12	12	12

The role of syntax in intelligibility

Research design

›6 types of syntactic constructions
›8 sentences per construction
›2 plausability conditions
›4 different linguistic conditions
›2 noise conditions

768 total sentence count

The role of syntax in intelligibility

Plans

- Experiment still in design stage, but will be completed and conducted during the coming year with informants in Denmark

- If results show that syntactic differences do indeed impede intelligibility the project can be extended to listener groups in Norway and Sweden.

Future research

Future research

› Develop a general model of intelligibility among closely related languages