Mutual intelligibility of Dutch and German cognates

by humans and computers

Universiteit Leiden

rijksuniversiteit groningen

Vincent J. van Heuven, Charlotte Gooskens, Renée van Bezooijen

Introduction: research theme

- Measuring mutual intelligibility
  - Judgments (opinion testing)
  - Functional tests (word recognition, dictation)
- Explaining mutual intelligibility
  - From linguistic difference/similarity
  - From extra-linguistic factors

Introduction: asymmetry

- Asymmetry in mutual intelligibility
  - Brazilian-Portuguese listeners understand Argentinian-Spanish better than *vice versa* (Jensen 1989)
  - South-Chinese dialect speakers understand Northern (Mandarin) dialects better than *vice versa* (Cheng 1997, Tang & van Heuven 2009)
  - Danes understand Swedes better than *vice versa* (Gooskens et al. 2010)

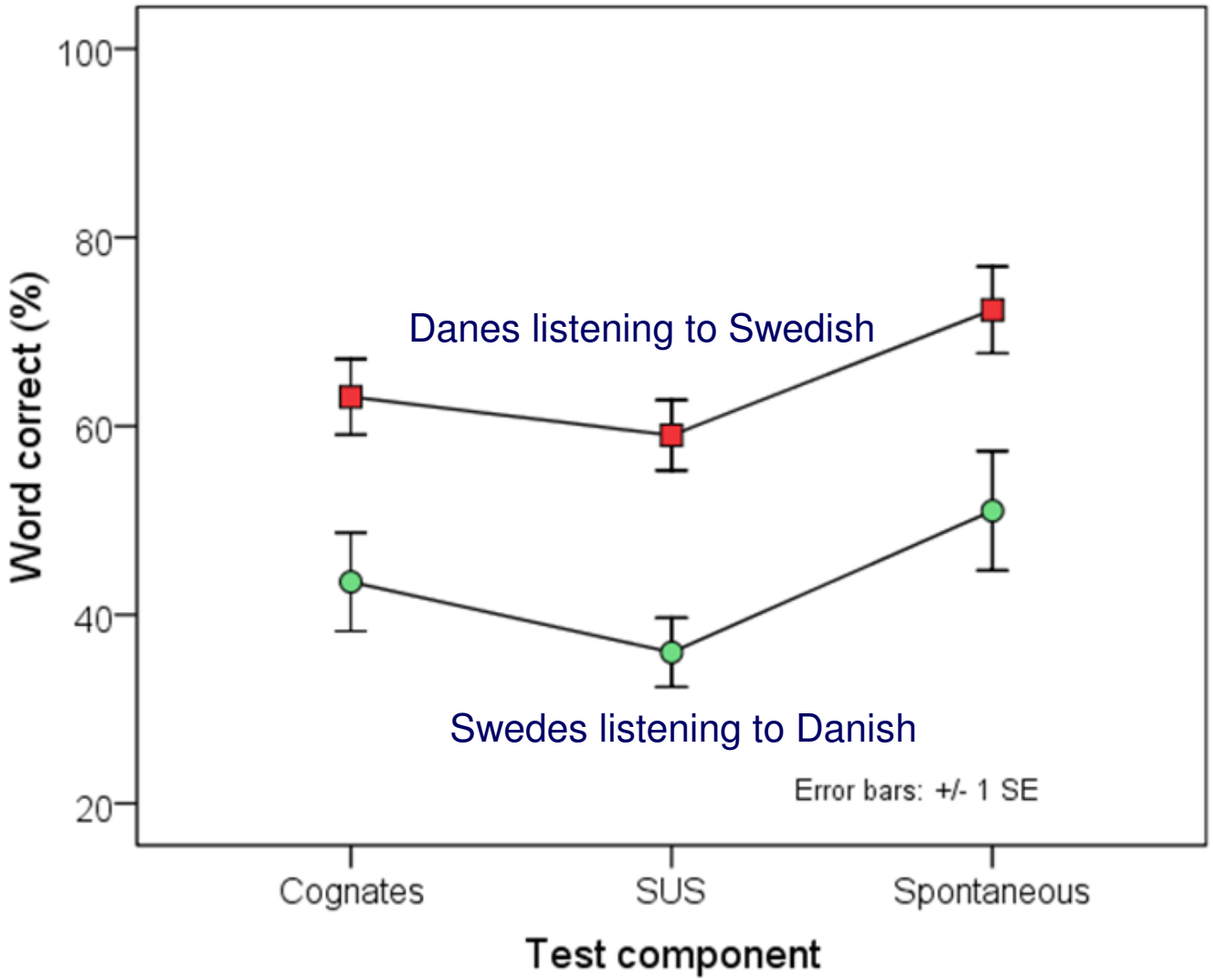


Figure 3. Intelligibility scores (percent correctly recognized words) obtained on three test components by Danish listeners decoding Swedish (squares) and by Swedish listeners decoding Danish (circles). Listener groups (20 Danes, 20 Swedes) were matched with respect to lexical knowledge of and familiarity with the non-native language. Error bars are +/− 1 standard error of the mean.

Introduction: asymmetry

- Asymmetry is often explained extra-linguistically
  - Difference in experience with the other language
    - Often caused by geography (large versus small country)
  - One language is socially dominant
  - Asymmetry in attitude towards the other language

Introduction: asymmetry

- Our question: to what extent can we account for asymmetry in mutual intelligibility from linguistic differences, in the abstraction of extra-linguistic factors?
- Experimental methods required, e.g.
  - using strict monolinguals (young children)
  - using automatic speech recognition = ASR

Today's case

- Mutual intelligibility between Dutch and German
  - Only closely related language pair with commercially available multilingual ASR technology
- Asymmetry
  - Dutch listeners understand German better than Germans understand Dutch (e.g. Haz 2008)
  - Can easily be explained extra-linguistically (geography, dominance, education, media)
- But is the asymmetry also linguistically motivated?
  - No reason to believe it is

Specific question

- How difficult is it for a Dutch listener to recognize German words, and *vice versa*, if they hear the other language for the first time in their life?
  - Cognates only
  - The larger the difference in sounds, the poorer the recognition
  - Relationship may be asymmetrical (due to neutralisation in sound patterns, as found for Chinese dialects)
  - Probably not in the case of Dutch and German

Method: ASR

- Automatic Speech Recognition
  - Training phase:
    - system learns Dutch sounds and sound sequences
      - Simulates a Dutch listener (with no knowledge of German)
    - system learns German sounds and sequences
      - Simulates a German listener (with no knowledge of Dutch)
    - So-called Hidden Markov sound models (HMMs)

Method: ASR

- Automatic Speech Recognition
  - Test phase 1 (after completion of training):
    - D system hears D test materials (high score?)
    - G system hears G materials (high score?)
  - Test phase 2: cross-language
    - D system hears G materials (low score?)
    - G system hears D materials (low score?)
  - Will we find an asymmetry in test phase 2?

Method: ASR

- Practical problem
  - HMM sound models are speaker dependent
  - System has to be trained anew for every different speaker
  - Cross-language test only possible if D and G speaker is the same individual
  - Perfect bilingual speaker needed

Excursion

- The hunt for the perfect bilingual
- Using voice line-up
  - Bilingual speaker must not be identified as deviant in an array of monolingual speakers
    - by Dutch judges
    - by German judges
- Was a major effort…
  - …but we managed

Excursion

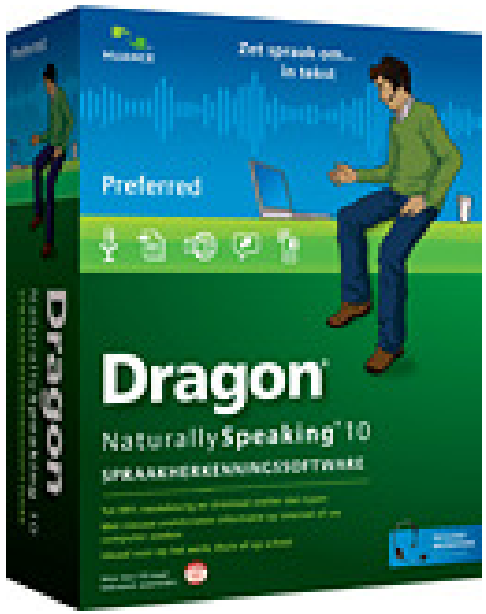
- Our bilingual speaker MM
  - Born 1976 in Switzerland from Dutch parents
  - Dutch at home, Swiss-German at school
  - From 1996 in Netherlands (studying D and G)
  - From 2000 employed in Germany (Berlin, Potsdam, Dortmund) with intermissions
- Was the only candidate that was never identified as deviant in both D and G voice line-ups
  - Demo (reading of D and G training text)

Excursion

- Bilingual speaker MM
  - In Dutch 
  - In German 

Commercial recognizer

- Nuance (formerly Lernout & Hauspie Speech Products) Dragon NaturallySpeaking version 10 for D and for G, ca. € 100 per language module
- Standard version (no specialized vocabulary)



Test materials

- 3000 most frequent nouns
  - Celex Dutch, Celex German
- Cognateship based on etymological dictionary
  - Ca. 750 cognate pairs, exit all others
- Ordered by mean token frequency across pairs
- Spoken as isolated words by bilingual MM
- Presented to ASR as isolated words, each followed by “punt/Punkt” (disabling “language model”)
- Recognition in batch mode (non-interactive)

Results (1)

- Training with pre-recorded materials went smoothly
- Test phase 1 (testing in same language)
  - Dutch: 220 correct out of 768 (29%)
  - Words at the top of list recognized better (higher token frequency matters?)
  - Therefore: further analysis limited to top-200

Results (2)

- Test phase 1 (same language)
- Top-200 words
  - D: 131/200 = 66% correct
  - G: 146/200 = 73% correct
- Check:
  - My own voice trained and tested on top-200
  - D: 128/200 = 64% correct

Intermediate conclusion (1)

- Dragon NaturallySpeaking
  - Does not do a good job on recognizing isolated words
  - Is prevented from making use of context
  - Has not seen enough examplars of initial and final sounds (at edges of words)
    - Sounds at word edges are major source of error

Results (3)

- Test phase 2: cross-language test
  - D (after G-training): 9/200 correct (5%)
  - G (after D-training): 7/200 correct (4%)
- Beyond top-200 stimulus-response cannot be aligned (seemingly random recognition)
- Correct recognition for (near-)identical cognates only

Correctly recognized cognates

	NL	> D		D	> NL
1.	broeder	Bruder	1.	Bruder	broeder
2.	radio	Radio	2.	Radio	radio
3.	loon	Lohn	3.	Lohn	loon
4.	idee	Idee			
5.	artikel	Artikel			
6.	roman	Roman			
7.	ingenieur	Ingenieur			
8.	winter	Winter			
9.	bier	Bier			
			4.	Frau	vrouw
			5.	Werk	werk
			6.	Vater	vader
			7.	Ring	ring

Conclusion (2)

- Results in phase 2 do not support asymmetry in mutual intelligibility between D ~ G
  - 7 vs 193 and 9 vs 191 ratio's do not differ (chi square)

One more attempt

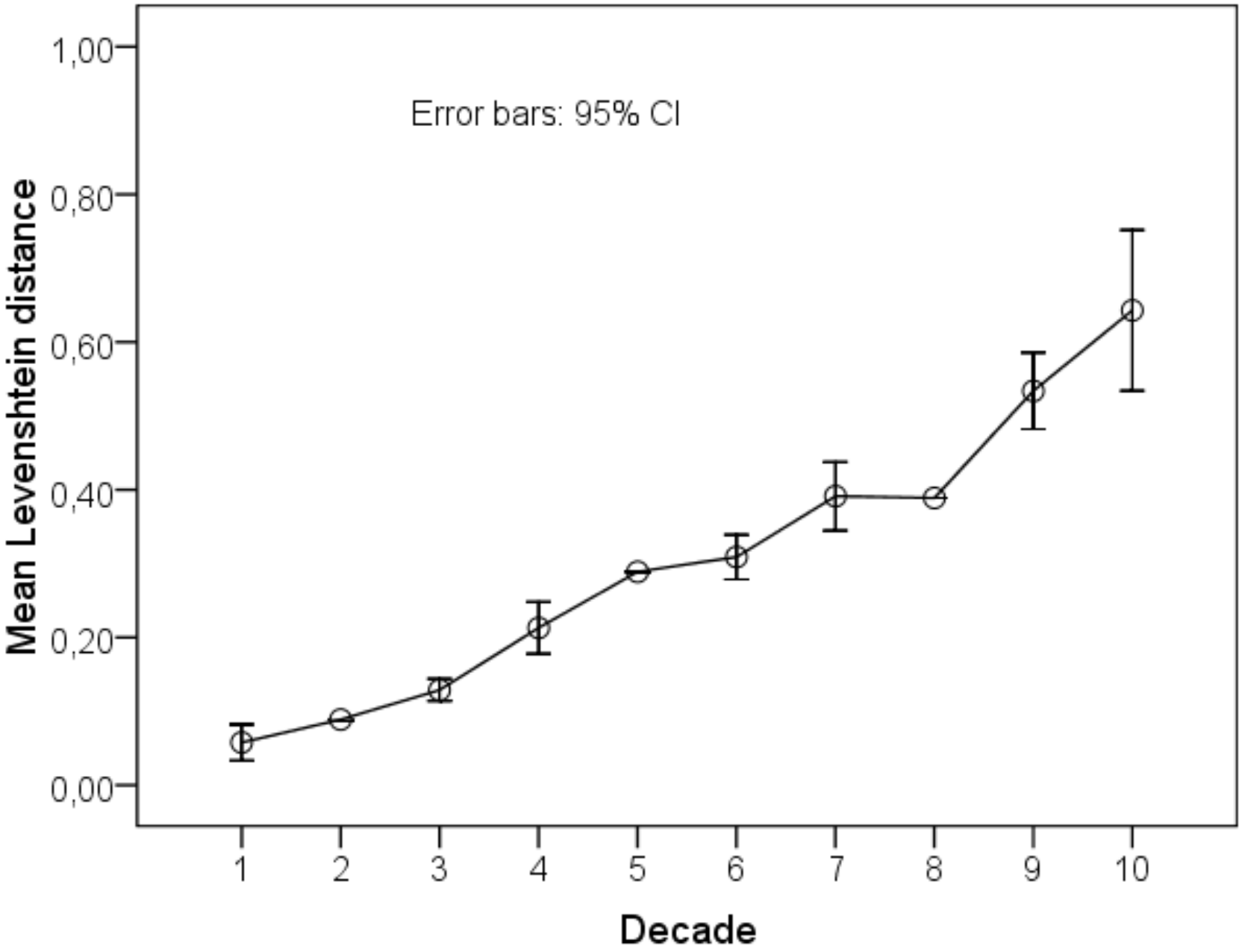
- How to boost the ASR performance?
  - Use test words in minimal context
    - Limit lexical category to nouns only, but excluding information on grammatical gender,
    - Target word onset is in continuous speech, e.g.
    - *ohne X* ('without X')
    - *zonder X*
  - No pre-recorded test materials of this type are available at this time, so I used my own voice

One more attempt

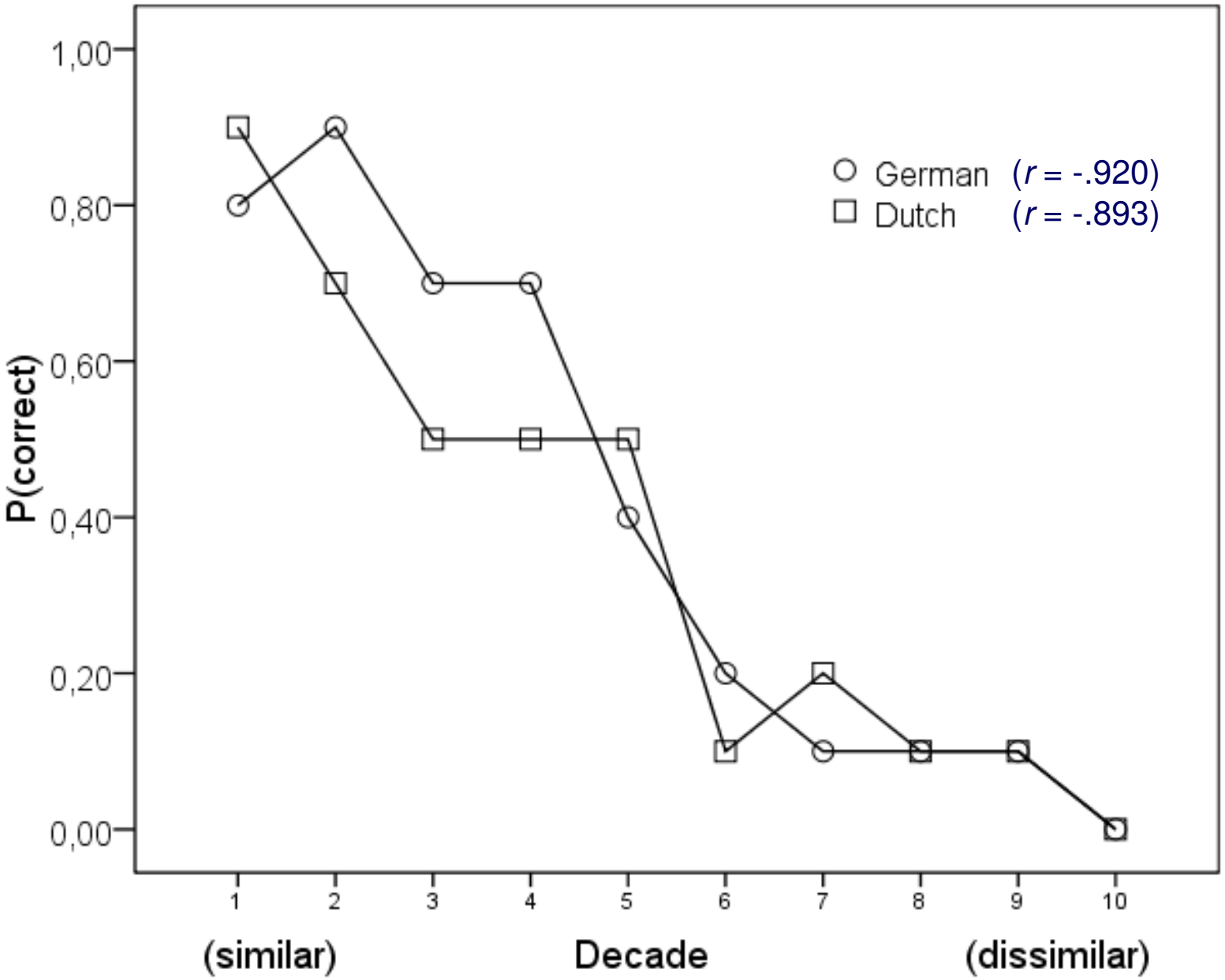
- System was trained with my voice (pre-recorded) in D and in G (but no perfect bilingual)
- Tested interactively on top-100 words
  - Same language D>D: 98% correct
  - G>G: 99% correct
  - Cross-language D>G: 40% correct
  - G>D: 36% correct

One more attempt

- Compute phonetic similarity between cognates
  - Levenshtein string edit distance measure
  - between 0 (no difference)
  - and 1 (completely different)
- Aggregate mean distance for decades (10 words adjacent on distance scale)
- Regress against cross-language recognition scores in Dutch and German



- Phonetic similarity (Levenshtein dist.) per decade



- Correct cross-language recognition of cognate as a function of phonetic similarity

Conclusions

- Addition of minimal context makes ASR a credible model of human performance
  - Near ceiling performance in own language
  - Cross-language recognition more difficult as distance between cognates gets larger
- 36 versus 40% correct cross-language recognition of cognates shows no existence of asymmetry between Dutch and German

Epilogue

- To be continued
  - Repeat pilot with perfect bilingual speaker
  - Validate results against human performance by strictly monolingual children (11 years old)
  - Currently under way
- References

  - Gooskens, C., V.J. van Heuven, R. van Bezooijen & J.J.A. Pacilly (2010). Is spoken Danish less intelligible than Swedish? *Speech Communication*, (in press).
  - Ház, E. (2005). *Deutsche und Niederländer. Untersuchungen zur Möglichkeit einer unmittelbaren Verständigung*. Hamburg: Kovač (Philologia 68).
  - Jensen, J. B. (1989). On the mutual intelligibility of Spanish and Portuguese. *Hispania* 72, 848-852.
  - Tang, C. & V.J. van Heuven (2009). Mutual intelligibility of Chinese dialects experimentally tested. *Lingua* 119, 709-732.