


 rijksuniversiteit
 groningen


22-1-2009 | 1

Conditional entropy

Phonetic distance

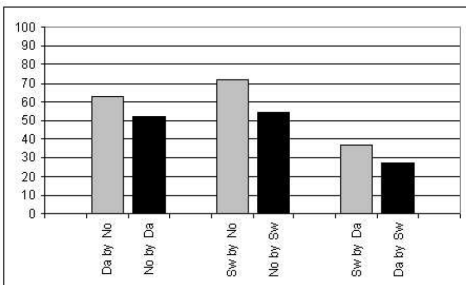
22-1-2009 | 2

- > Levenshtein-afstand is symmetric
- > Problem: Intelligibility scores are often asymmetric



 rijksuniversiteit
 groningen

Results of spoken intelligibility tests

22-1-2009 | 3

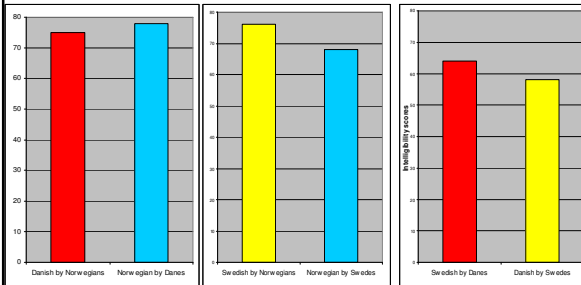


| Direction | Intelligibility Score (%) |
|-----------|---------------------------|
| Da by No | ~60 |
| No by Da | ~50 |
| Sw by No | ~70 |
| No by Sw | ~55 |
| Sw by Da | ~35 |
| Da by Sw | ~25 |



 rijksuniversiteit
 groningen

Results of LRS test

22-1-2009 | 4




| Direction | Intelligibility Score (%) |
|----------------------|---------------------------|
| Danish by Norwegian | ~75 |
| Norwegian by Danish | ~78 |
| Swedish by Norwegian | ~78 |
| Norwegian by Swedish | ~68 |
| Swedish by Danish | ~65 |
| Danish by Swedish | ~60 |


 rijksuniversiteit
 groningen

Method for measuring distances asymmetrically: Conditional entropy

22-1-2009 | 5


- > measure of linguistic distance which is sensitive to asymmetry
- > basis: probability of a sound to match the same sound in a corresponding word in another language
- > entropy measures the level of uncertainty / the degree of surprise based on frequency data
- > conditional entropy measures probabilities based on conditions, e.g. the probability of variable X given variable Y


 rijksuniversiteit
 groningen

Entropy: Example

22-1-2009 | 6

| | | |
|--|---|--|
| Da. <i>kommer</i> /kɔmɛR/ Sw. <i>kommer</i> /kɔm:ɛr/ ,comes' | <i>savner</i> /sɑvnɛR/ <i>saknar</i> /sɑ:knar/ ,misses' | |
| Da. ə $\begin{cases} \epsilon \\ \alpha \end{cases}$ | Sw. | |


 rijksuniversiteit
 groningen

Conditional entropy in the Scandinavian languages

22-1-2009 | 8

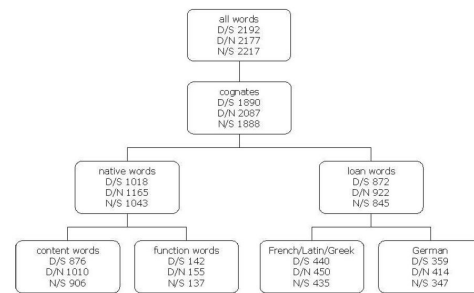
- › Moberg, Gooskens, Nerbonne & Vaillette (2007) use conditional entropy to model the asymmetric mutual intelligibility of the Scandinavian languages
- › basis: database of formal and informal speech
- › divided into lexical groups



rijksuniversiteit
groningen

Database

22-1-2009 | 8



rijksuniversiteit
groningen

Expectations:

22-1-2009 | 9

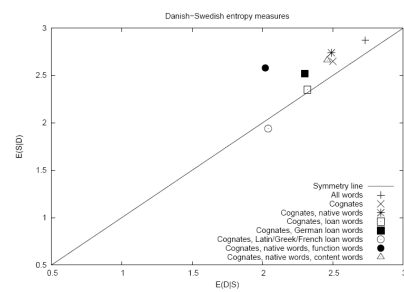
- › Danes understand Swedes better than vice versa, i.e. expected entropy $H(D|S) > H(S|D)$
- › Norwegians understand Swedes better than vice versa, i.e. expected entropy $H(N|S) > H(S|N)$
- › Norwegians and Danes have more or less symmetric intelligibility scores, so we expect the entropies to be symmetric as well



rijksuniversiteit
groningen

Results from Moberg, Gooskens, Nerbonne & Vaillette (2007)

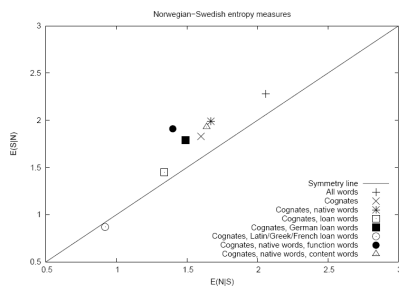
22-1-2009 | 10



rijksuniversiteit
groningen

Results from Moberg, Gooskens, Nerbonne & Vaillette (2007)

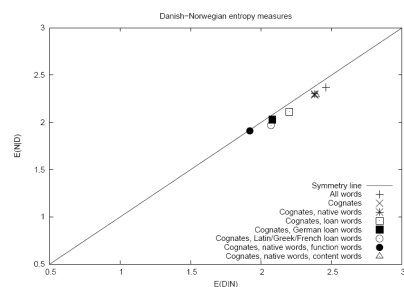
22-1-2009 | 11



rijksuniversiteit
groningen

Results from Moberg, Gooskens, Nerbonne & Vaillette (2007)

22-1-2009 | 12



rijksuniversiteit
groningen

Questions

22-1-2009 | 15

- > What do entropies actually model?
- > Are entropies a good measure for word intelligibility?
- > Do entropies model a first-contact situation?



rijksuniversiteit
groningen