Conditional entropy

Phonetic distance

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

Results of spoken intelligibility tests

Results of LRS test

Method for measuring distances asymmetrically: Conditional entropy

- 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

Entropy: Example
Conditional entropy in the Scandinavian languages

- 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

Database

Expectations:

- 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

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

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