Relating Levenshtein distances to intelligibility scores

Research question

› How well can Levenshtein distances predict intelligibility?

Investigations

Beijering, Gooskens & Heeringa (2009): The intelligibility of 18 Nordic language varieties among Danes

Kürschner, Gooskens & Van Bezooijen (2009): The intelligibility of 384 Swedish words among Danes

Material

› recordings of ‘The North Wind and the Sun’
› 18 Nordic language varieties
› mean 98 words
› phonetic transcriptions of the cognates (historically related words)

Material

The North Wind and the Sun were disputing which was the stronger, when a traveler came along wrapped in a warm cloak. They agreed that the one who first succeeded in making the traveler take off his cloak should be considered stronger than the other. Then the North Wind blew as hard as he could, but the more he blew the more closely did the traveler fold his cloak around him; and at last the North Wind gave up the attempt. Then the Sun shined out warmly, and immediately the traveler took off his cloak. And so the North Wind was obliged to confess that the Sun was the stronger of the two.
**Material**

- 8 Norwegian varieties
- Standard Swedish (Stockholm) + 3 Swedish varieties
- Standard Danish (Lyngby) + 3 Danish varieties
- Finland Swedish (Helsinki)
- Faeroese (Torshavn)

**Intelligibility**

**Stimulus material:**
- 6 sentences in 6 varieties (latin square design)

**Test persons:**
- 18 groups of high school pupils from Copenhagen, aged between 15 and 20 (average 17.6)

**Task:**
- translate word for word into Standard Danish
Correlation

**Dependent variable:**
- % correctly translated words per variety

**Independent variable:**
- Levenshtein distances between Standard Danish and the 18 varieties

Relation between intelligibility and Levenshtein distances

\[ r = -0.86 \]

**Relation between intelligibility and Levenshtein distances**

- It is possible to predict intelligibility of language varieties to a high extent by means of Levenshtein distances

Consonants vs vowels

- Are consonants or vowels more important for the intelligibility?

Consonants vs vowels

- Word recognition in English depends more on correct consonant identification than on the correct identification of vowels (Van Ooijen 1994 and references).

- **Consonants** are more important for the semantic identity of a word than **vowels**; they function as reference points in words.

Consonants vs vowels

**Hypothesis:**

- Deviations in vowels are less damaging for the intelligibility of a closely related language than deviations in the consonants.
The intelligibility of 384 Swedish words among Danes

LRS experiment

Material:
› Recordings of 347 frequent Swedish cognates
› Phonetic transcriptions

Test persons:
› 38 Danish high school pupils aged 16-19

Task:
› Translate words into Standard Danish

Correlation

Independent variable:
› Levenshtein distances between Swedish and Danish words

Dependent variable:
› % correct translations per words

Relation between intelligibility and Levenshtein distances

› $r = -0.33$

› Other factors must be considered in order to predict intelligibility at the word level