On the measurement of morphosyntactic distances

Sebastian Kürschner

General problems

- The effect of morphosyntactic distances on intelligibility has – to our knowledge – never been systematically investigated
- Morphosyntactically driven dialectometry is still in the first steps
- How do you exclude differences on other levels than the morpho-syntactic one (e.g., phonetic differences) in experiments on intelligibility?

Example: Research on morphosyntactic distances

- Szmrecsanyi (submitted) calculates morphosyntactic distances in English dialects
- frequency-based approach based on corpus analysis in FRED (Freiburg English Dialect Corpus)
- definition of 35 morphosyntactic variables

Feature catalogue

- Szmrecsanyi (submitted)
  - material in the corpus (varieties from 38 counties) was coded automatically and semi-automatically with the features
  - feature frequency calculated and extracted into vectors
  - table converted into distance matrix
Example: Syntactic distances based on POS

Nerbonne/Wiersma (2006)

Analysis of a corpus tagged with POS

analysis of 3-grams

distance measure based on permutation tests

Corpus: English speech of two groups of Finnish emigrants to Australia:

- emigrants older than 16 on the date of emigration
- emigrants younger than 17 on the date of emigration

Syntactic dialect atlas variables

Spruit (2006)

Syntactic distance measure based on 507 syntactic variables found in the Syntactic Atlas of the Dutch Dialects:

- variables and their variants were introduced by the editors of the atlas
- frequency-based approach: occurrence of variables is aggregated using dialectometric methods

Example variables

Table 1 Map all 507 variables in NANDV shows the five syntactic variables in the context of short reflexive pronoun as object of kinship reflexive verb.

<table>
<thead>
<tr>
<th>Context</th>
<th>Short reflexive verb</th>
<th>Variables</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>de diens</td>
<td>de diens</td>
<td>de diens</td>
<td>de diens</td>
</tr>
<tr>
<td>z'n z'n</td>
<td>z'n z'n</td>
<td>z'n z'n</td>
<td>z'n z'n</td>
</tr>
<tr>
<td>z'n z'n</td>
<td>z'n z'n</td>
<td>z'n z'n</td>
<td>z'n z'n</td>
</tr>
<tr>
<td>z'n z'n</td>
<td>z'n z'n</td>
<td>z'n z'n</td>
<td>z'n z'n</td>
</tr>
</tbody>
</table>

Table 2 Map all 507 variables in NANDV shows the six syntactic variables in the context of short object relative.

<table>
<thead>
<tr>
<th>Context</th>
<th>Short object relative</th>
<th>Variables</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>de diens</td>
<td>de diens</td>
<td>de diens</td>
<td>de diens</td>
</tr>
<tr>
<td>z'n z'n</td>
<td>z'n z'n</td>
<td>z'n z'n</td>
<td>z'n z'n</td>
</tr>
<tr>
<td>z'n z'n</td>
<td>z'n z'n</td>
<td>z'n z'n</td>
<td>z'n z'n</td>
</tr>
<tr>
<td>z'n z'n</td>
<td>z'n z'n</td>
<td>z'n z'n</td>
<td>z'n z'n</td>
</tr>
</tbody>
</table>

Morphosyntactic distance and intelligibility

- all measures frequency-based
- how to model first-contact situation?
- how to test the influence of morphosyntactic distance?
- how to isolate the morphosyntactic level (or even morphological and syntactic levels separately) from other levels such as lexicon and phonetics?
- how to test the relative contribution of linguistic levels as determinants of intelligibility?