Comparing Annotations of Non-Verbal Vocalisations (NVVs) in Speech Corpora

Jürgen Trouvain & Raphael Werner
Saarland University, Saarbrücken, Germany
trouvain|rwerner@lst.uni-saarland.de

1 Introduction
- Non-verbal vocalisations (NVVs) often investigated with experimental data
- Alternative data source: speech corpora
- Unclear: how are NVVs annotated in corpora?
- Aim: to compare annotations of NVVs in various corpora

2 Corpora
- 11 inspected corpora
- German (de) and English (en)
- Spontaneous and scripted speech
- For references see extended abstract

3 Annotation categories
- Only highly frequent types of NVVs
- Laughs and speech-laughs
- Breath noises
- Coughing and throat clearing
- Clicks and lip smacking

4 Overview

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ICSI</td>
<td>en</td>
<td>spont.</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>n</td>
</tr>
<tr>
<td>AMI</td>
<td>en</td>
<td>spont.</td>
<td>y</td>
<td>n</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>y</td>
</tr>
<tr>
<td>Switchboard</td>
<td>en</td>
<td>spont.</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>Diapix</td>
<td>en</td>
<td>spont.</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>MapTask</td>
<td>en</td>
<td>spont.</td>
<td>y</td>
<td>n</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>y</td>
</tr>
<tr>
<td>Buckeye</td>
<td>en</td>
<td>spont.</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>GECO</td>
<td>de</td>
<td>spont.</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>y</td>
<td>y</td>
<td>n</td>
</tr>
<tr>
<td>Lindenstr</td>
<td>de</td>
<td>spont.</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>KielC.Read</td>
<td>de</td>
<td>scripted</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>IFCASL</td>
<td>de</td>
<td>scripted</td>
<td>n</td>
<td>n</td>
<td>y</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>DIRNDL</td>
<td>de</td>
<td>scripted</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
</tbody>
</table>

y(es) = explicit label found in manual and/or annotation; n(o) = not y

5 Discussion
- NVVs with a 'Cinderella status'
- Only one corpus showing all NVVs
- Breath noises as most frequent NVV not annotated in all corpora
- Further sub-divisions (e.g. breath and laugh) challenging
- Other studies observe many errors of human annotators (incl. filler particles)
- Inconsistencies probably due to a general lack of awareness of NVVs
- Mid-term goal: more transparency or even a common set of annotation guidelines for NVVs
- Widespread annotation would allow to discover (and ideally understand) NVVs as often subtle phenomena
- Automatic and semi-automatic detection would be very helpful