Language learning from an audio description corpus

Eva Schaeffer-Lacroix, Sorbonne Université/Espé de Paris

TaLC13 (Teaching and Language Corpora Conference), Faculty of Education, University of Cambridge
AD produced by future German teachers

Web soap for German learners: *Jojo sucht das Glück* [*Jojo’s pursuit of happiness*]

Episode 12: *It’s coffee time*

AD tool: Youdescribe.org
Audio description, a highly normed text genre

Characteristics of ADs:
• Description of visual events
• Speaking indications

Recommendations for AD authors:
• Be short
• Be precise
• Be objective
• Put yourself in the shoes of the visually impaired

You can see...
No, I can't!
Expert AD corpus

<table>
<thead>
<tr>
<th>be short</th>
<th>compound words</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>present participles</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>be precise</th>
<th>prepositions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>verb particles</td>
</tr>
<tr>
<td></td>
<td>wide range of verbs</td>
</tr>
</tbody>
</table>

AD manuscripts from *Neues aus Buettenwarder*, a series telling rural stories located in northern Germany.

Corpus uploaded to *Sketch Engine* and TXM.

336,723 tokens and 69 text files

Part of speech annotations with TreeTagger & RFtagger.
10:01:46
"Ich bin beim Essen!" ["I am eating!"]
ss Adsche nimmt Griem den Teller weg. [Adsche takes the plate away from Griem.]

Reference points for recordings:
- time indications;
- film script prompts surrounded by quotation marks;
- speech indications highlighted in bold (e.g. "ss" : speak very quickly).
1.19 Aufnahme *<auf dem Essen> auf dem Tisch: (...) 
1.36 Jojo ist verlegen, Lena senkt den Blick. 
1.50 Jojo zeigt *<xxx> einen Teller.

1.19 *<Shot of the meal> on the table: (...) 
1.36 Jojo is embarrassed, Lena lowers her eyes. 
1.50 Jojo *<shows> a plate.
« Be short with compound words. »

Aufnahme auf dem Essen auf dem Tisch

> Groβaufnahme des Essens

[Shot of the meal on the table]
Too long; preposition error
[> Close-up view of the meal]
= compound noun & genitive object
## Compound words in the *Buetttenwarder corpus*

<table>
<thead>
<tr>
<th>Form</th>
<th>Translation</th>
<th>Number of characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>schiffchenförmige</td>
<td>shaped like a ship</td>
<td>17</td>
</tr>
<tr>
<td>schwarzglänzenden</td>
<td>shiny black</td>
<td>17</td>
</tr>
<tr>
<td>selbstgezimmerte</td>
<td>hand timbered</td>
<td>16</td>
</tr>
<tr>
<td>treckerähnlichen</td>
<td>similar to a tractor</td>
<td>16</td>
</tr>
<tr>
<td>baumbewachsenen</td>
<td>grown with trees</td>
<td>15</td>
</tr>
<tr>
<td>blaugestreiften</td>
<td>blue-striped</td>
<td>15</td>
</tr>
<tr>
<td>kartenspielende</td>
<td>card-playing</td>
<td>15</td>
</tr>
</tbody>
</table>
Be precise with prepositions.

Jojo zeigt einen Teller >
Jojo zeigt auf einen Teller
[to show a plat vs.
to point at a plate]
You can show...
- a beer coaster, a "bird", your thumb, a photograph, ...

You can point...
- at/behind/in direction of sth or sb and with sth.
« Be short with present participles. » ✓

Vier lachende Jugendliche

Four smiling young people

kopfschütteln [shaking their head]

grübelnd [brooding]

schmunzelnd [grinning]

lächelnd [smiling]

all present participles

Words per million

0 50 100 150 200 250 300 350 400

Weissensee

Buettenwarde
« Vary your verbs. » ✓
« Don’t use ‘see’. » ✓
Next steps to do

- XML annotations for the expert corpus
  - time indications
  - speech indications
  - compound words
  - …

- Storing the Buettenwarder corpus on Ortolang - CLARIN (rights granted by the owners of the data)
## Action-research with a research group and a control group

<table>
<thead>
<tr>
<th>RG task</th>
<th>CG task</th>
<th>Independent variable</th>
<th>Dependent variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design an audio description scenario with the help of corpus tools</td>
<td>Design an audio description scenario with the help of other tools (e.g. online grammars, text books)</td>
<td>Type of tools</td>
<td>- Type of language descriptions;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- type of created language learning activities.</td>
</tr>
</tbody>
</table>
References


• *YouDescribe*. (2017). Free online tool which can be used to add description to YouTube videos. Developed by The Smith-Kettlewell Eye Research Institute.