Cross-Linguistic Distributional Analyses with Frequent Frames: the Cases of German and Turkish

Hao Wang¹, Barbara Höhle², F. Nihan Ketrez³, Aylin C. Küntay⁴, Toben H. Mintz¹

¹University of Southern California, ²University of Potsdam, ³İstanbul Bilgi University, ⁴Koç University
Word Categorization

NP → D N

• Syntactic categories as basic units of grammar

• Categorizing words is a necessary step for acquiring syntax
Current Approaches

• Phonological and prosodic cues

• Semantic bootstrapping

• Distributional information
Distributional information

• Lexical co-occurrence pattern
  – *the cat is on the mat*

• Originated from structural linguists
  – Bloomfield 1933, Harris 1951

• Advanced into a theory of language acquisition by Maratsos & Chalkley (1980)
Distributional Information in CDS

• Bigrams and other environments (Cartwright & Brent, 1997; Mintz, Newport & Bever, 2002; Redington, Chater & Finch, 1998)
  
  the __

• Frequent frames (Mintz 2003)
  
  – A frame is defined as two jointly occurring words with one word intervening.
  
  you __ the
Procedure for Frequent Frame Analysis

you read the story to Mommy
• you read the
• read the story
• the story to
• would you put the cans back?
• story to Mommy
• you get the nuts.
  you take the chair back.
• you read the story to Mommy.
  you__the  Verb frame
the__to   Noun frame
Categorization Evaluation

\[
\text{Accuracy} = \frac{\text{hits}}{\text{hits} + \text{false alarms}}
\]

- Accuracy is penalized when two words from different grammatical categories are grouped together
- Range: 0 to 1

**Hit**: two words from the same grammatical category grouped together

**False alarm**: two words from different grammatical categories grouped together
**Frequent Frames in English CDS**  
(Experiment 1, Mintz, 2003)

- Six English corpora from CHILDES  
  – Age under 2;6

- Most frequent 45 frames

<table>
<thead>
<tr>
<th>Mean Token Accuracy</th>
<th>Actual</th>
<th>Chance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.98</td>
<td>0.49</td>
</tr>
</tbody>
</table>

Categorization

Categorization
you__it frame (Peter)

433 tokens   93 types

put, see, do, did, want, fix, turned, get, got, turn, throw, closed, think, leave, take, open, find, bring, took, like, knocked, putting, pull, found, make, have, fixed, finish, try, swallow, opened, need, move, hold, give, fixing, drive, close, catch, threw, taking, screw, say, ride, pushing, hit, hiding, had, eat, carry, build, brought, write, wiping, wipe, wind, unzipped, underneath, turning, touching, tore, tie, tear, swallowed, squeeze, showing, show, said, rip, read, reach, pushed, push, play, pick, parking, made, love, left, knock, knew, hid, flush, finished, expected, dropped, drop, draw, covered, closing, call, broke, blow
Frequent frames in English

• Robust cue to word categories
• Potential source for initial bootstrapping
• Good accuracy even for a resource-limited learner (Wang & Mintz, 2008)
Learning problem

• Not only English-speaking children have to learn word categories. Every kid has to!

• Empirical questions
  – Work in typologically different languages?
  – Any modification to the mechanism?
# Cross-linguistic Challenges

<table>
<thead>
<tr>
<th>Language</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>French</td>
<td>Homophony between clitic object pronouns and determiners. Three different definite determiners.</td>
</tr>
<tr>
<td>Spanish</td>
<td>Homophony among function words. Pro-drop.</td>
</tr>
<tr>
<td>Chinese</td>
<td>No inflectional morphology. Pro-drop.</td>
</tr>
</tbody>
</table>
## Cross-linguistic studies

<table>
<thead>
<tr>
<th>Language</th>
<th>Sources</th>
<th>Frames</th>
<th>Accuracy</th>
<th>Corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>French</td>
<td>Chemla, Mintz, Bernal, &amp; Christophe 2009</td>
<td>Most frequent 6 frames (at least 11 types, 18 tokens)</td>
<td><strong>Accuracy = 1.0</strong>, 12% corpus</td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>Weisleder &amp; Waxman 2010</td>
<td></td>
<td><strong>Accuracy = 0.75</strong>, 62% corpus</td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>Cai 2006, Xiao, Cai and Lee 2006</td>
<td>130 frames (at least 15 tokens)</td>
<td><strong>Accuracy = 0.76</strong>, Prominence = 90% (verb &amp; adj.)</td>
<td></td>
</tr>
<tr>
<td>Dutch</td>
<td>Erkelens, 2009</td>
<td></td>
<td><strong>Accuracy = 0.71</strong></td>
<td></td>
</tr>
</tbody>
</table>
German Morphology

• A highly inflected language
  – 4 noun cases
  – 3 genders
  – 6 definite articles

• Many suppletive paradigms
  – der-das-die
  – er-ihn-ihm: he-him-him

• Determiners used pronominally

Ein Journalist interviewte die Sängerin.
A journalist interviewed the singer.

Die wurde auf der ganzen Welt als neuer Opernstar gefeiert.
She was celebrated across the whole world as the new opera star.
German Syntax

• Word order: Variable position of finite verb
  – Either 2nd or final position
  – All kinds of constituents can occur before the verb (subjects, objects, adverbials)

  Er kauft Blumen
  He buys flowers

  weil er Blumen kauft
  because he buys flowers
Turkish Morphology

- **Agglutinative**
  - Inflectional suffixes

<table>
<thead>
<tr>
<th>Word</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ev</td>
<td>&quot;house&quot;</td>
</tr>
<tr>
<td>evler</td>
<td>&quot;houses&quot;</td>
</tr>
<tr>
<td>evlerim</td>
<td>&quot;my houses&quot;</td>
</tr>
<tr>
<td>evlerimiz</td>
<td>&quot;our houses&quot;</td>
</tr>
<tr>
<td>evlerimizde</td>
<td>&quot;in our houses&quot;</td>
</tr>
<tr>
<td>evlerimizdeki</td>
<td>&quot;which is in our houses&quot;</td>
</tr>
</tbody>
</table>
Turkish Syntax

- Free word order
  - SOV (canonical order)
    SVO, OVS, OSV, VSO, VOS

- Absence of function words

- Pro-drop

"Hakan went to school."
Hakan okula gitti.
Ø okula gitti.
Hakan Ø gitti.

"Hakan is reading the book."
Hakan kitabı okuyor.
Hakan book-acc read-IPFV

(Adapted from Underhill, 2006)
## Challenges for Distributional Analysis

<table>
<thead>
<tr>
<th>Language</th>
<th>Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>German</td>
<td>Flexible word order.</td>
</tr>
<tr>
<td></td>
<td>Stem changes in def articles &amp; pronouns.</td>
</tr>
<tr>
<td></td>
<td>Determiners used pronominally.</td>
</tr>
<tr>
<td>Turkish</td>
<td>Free word order.</td>
</tr>
<tr>
<td></td>
<td>Many bound morphemes.</td>
</tr>
<tr>
<td></td>
<td>Absence of function words.</td>
</tr>
<tr>
<td></td>
<td>Pro-drop.</td>
</tr>
</tbody>
</table>
Morpheme-Level Analysis

• Suffixes in languages with rich morphology
  – Many suffixes are restricted to be used with one word class
  – Turkish-learning children start producing inflections very early and they do it correctly

• Distributions at morpheme level
Current Analyses

• German
• Turkish
• Word-level frequent frames
• Morpheme-level frequent frames
Current Analyses – German CDS

• Simone (1;10.22-2;5.19) (Miller 1979)
  – 5685 utterances

• Most frequent 45 frames

• Control condition
  – Words randomly re-arranged between frames

soll ich mal gucken, was die machen ?

soll ich mal guck-en was die mach-en

soll ich mal guck -en was die mach -en
# German Results

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Completeness</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>FF Word</td>
<td>0.86</td>
<td>0.07</td>
<td>884</td>
</tr>
<tr>
<td>FF Morpheme</td>
<td>0.88</td>
<td>0.05</td>
<td>1857</td>
</tr>
</tbody>
</table>
# German Frequent Morpheme Frames

<table>
<thead>
<tr>
<th>Frame</th>
<th>Type</th>
<th>Token</th>
<th>Majority Cat.</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>was__-st</td>
<td>12</td>
<td>122</td>
<td>V</td>
<td>99%</td>
</tr>
<tr>
<td>Maxe__-t</td>
<td>32</td>
<td>107</td>
<td>V</td>
<td>100%</td>
</tr>
<tr>
<td>was__-t</td>
<td>18</td>
<td>91</td>
<td>V</td>
<td>100%</td>
</tr>
<tr>
<td>ge__-t</td>
<td>25</td>
<td>88</td>
<td>V</td>
<td>98%</td>
</tr>
<tr>
<td>-e__-e</td>
<td>32</td>
<td>65</td>
<td>Adj</td>
<td>38%</td>
</tr>
<tr>
<td>du__-st</td>
<td>26</td>
<td>65</td>
<td>V</td>
<td>98%</td>
</tr>
<tr>
<td>wir__-en</td>
<td>22</td>
<td>63</td>
<td>V</td>
<td>100%</td>
</tr>
<tr>
<td>'n__-chen</td>
<td>3</td>
<td>59</td>
<td>Pro</td>
<td>91%</td>
</tr>
<tr>
<td>-e__-en</td>
<td>21</td>
<td>57</td>
<td>V</td>
<td>82%</td>
</tr>
<tr>
<td>pass__auf</td>
<td>2</td>
<td>54</td>
<td>Pt</td>
<td>100%</td>
</tr>
<tr>
<td>-en__mal</td>
<td>11</td>
<td>52</td>
<td>Pro</td>
<td>44%</td>
</tr>
<tr>
<td>das__-t</td>
<td>18</td>
<td>49</td>
<td>V</td>
<td>100%</td>
</tr>
</tbody>
</table>

...
### German Results

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Completeness</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FF Word</strong></td>
<td>0.86</td>
<td>0.07</td>
<td>884</td>
</tr>
<tr>
<td><strong>FF Morpheme</strong></td>
<td>0.88</td>
<td>0.05</td>
<td>1857</td>
</tr>
<tr>
<td><strong>F₁F₂ Word</strong></td>
<td>0.47</td>
<td>0.04</td>
<td>1216</td>
</tr>
<tr>
<td><strong>F₁F₂ Morpheme</strong></td>
<td>0.78</td>
<td>0.10</td>
<td>2742</td>
</tr>
<tr>
<td><strong>F₂₁ Morpheme</strong></td>
<td>0.30</td>
<td>0.07</td>
<td>2672</td>
</tr>
</tbody>
</table>
Frequent Frames in German CDS (Stumper & Lieven, 2009)

- One child (2;0 – 2;6)
  - 28074 utterances
- Most frequent 45 frames

<table>
<thead>
<tr>
<th></th>
<th>Token accuracy</th>
<th></th>
<th>Type accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Analysis</td>
<td>Random</td>
<td>Analysis</td>
</tr>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Standard Labelling</td>
<td>.776&lt;sup&gt;a,e,g&lt;/sup&gt; (.209)</td>
<td>.450&lt;sup&gt;a&lt;/sup&gt; (.001)</td>
<td>.584&lt;sup&gt;b,f,g&lt;/sup&gt; (.284)</td>
</tr>
<tr>
<td>Expanded Labelling</td>
<td>.643&lt;sup&gt;c,e,h&lt;/sup&gt; (.198)</td>
<td>.335&lt;sup&gt;c&lt;/sup&gt; (.001)</td>
<td>.432&lt;sup&gt;d,f,h&lt;/sup&gt; (.213)</td>
</tr>
</tbody>
</table>

<sup>a,b,c,d</sup> Scores differ significantly (Fisher’s Omnibus Test, p < .001).
<sup>e,f,g,h</sup> Means differ significantly (paired t-tests, p < .001).
Current Analyses – Turkish CDS

• **Two children** (Ural, Yuret, Ketrez, Koçbaş & Küntay 2009)
  – Elif (0;9.10-1;9.28), 21741 utterances
  – İrmak (0;9.0-2;0.16), 16024 utterances

• **Control condition**
  – Words randomly re-arranged between frames

```
 sen hep ayaklarını soku[yo(r)]sun .
 sen hep ayak-lar-ın-ı sok-uyo(r)-sun .
 sen hep ayak-PL-POSS&2S-ACC sok-IPFV-2S
 sen hep ayak PL POSS&2S ACC sok IPFV 2S
```
# Turkish Results

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Accuracy</th>
<th>Completeness</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FF Word</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elif</td>
<td>0.54</td>
<td>0.09</td>
<td>1269</td>
</tr>
<tr>
<td>Irmak</td>
<td>0.40</td>
<td>0.11</td>
<td>1656</td>
</tr>
<tr>
<td><strong>FF Morpheme</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elif</td>
<td>0.93</td>
<td>0.06</td>
<td>6102</td>
</tr>
<tr>
<td>Irmak</td>
<td>0.88</td>
<td>0.06</td>
<td>2764</td>
</tr>
</tbody>
</table>
## Turkish Frequent Morpheme Frames (Elif)

<table>
<thead>
<tr>
<th>Frame</th>
<th>Type</th>
<th>Token</th>
<th>Majority</th>
<th>Cat.</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEN__POSS&amp;3S</td>
<td>163</td>
<td>538</td>
<td>N</td>
<td>96%</td>
<td></td>
</tr>
<tr>
<td>ne__IPFV</td>
<td>24</td>
<td>348</td>
<td>V</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>ne__PAST</td>
<td>26</td>
<td>316</td>
<td>V</td>
<td>98%</td>
<td></td>
</tr>
<tr>
<td>QUE__PAST</td>
<td>77</td>
<td>260</td>
<td>V</td>
<td>98%</td>
<td></td>
</tr>
<tr>
<td>DAT__PAST</td>
<td>43</td>
<td>217</td>
<td>V</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>QUE__IPFV</td>
<td>59</td>
<td>215</td>
<td>V</td>
<td>99%</td>
<td></td>
</tr>
<tr>
<td>DAT__IPFV</td>
<td>51</td>
<td>209</td>
<td>V</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>ACC__PAST</td>
<td>71</td>
<td>203</td>
<td>V</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>QUE__FUT</td>
<td>34</td>
<td>165</td>
<td>V</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>LOC__var</td>
<td>52</td>
<td>152</td>
<td>WH</td>
<td>55%</td>
<td></td>
</tr>
<tr>
<td>ACC__IPFV</td>
<td>61</td>
<td>151</td>
<td>V</td>
<td>99%</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Turkish Results

<table>
<thead>
<tr>
<th></th>
<th>Corpus</th>
<th>Accuracy</th>
<th>Completeness</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FF Word</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elif</td>
<td>0.54</td>
<td>0.09</td>
<td></td>
<td>1269</td>
</tr>
<tr>
<td>Irmak</td>
<td>0.40</td>
<td>0.11</td>
<td></td>
<td>1656</td>
</tr>
<tr>
<td><strong>FF Morpheme</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elif</td>
<td>0.93</td>
<td>0.06</td>
<td></td>
<td>6102</td>
</tr>
<tr>
<td>Irmak</td>
<td>0.88</td>
<td>0.06</td>
<td></td>
<td>2764</td>
</tr>
<tr>
<td><strong>Bigram Morpheme</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elif</td>
<td>0.31</td>
<td>0.05</td>
<td></td>
<td>33793</td>
</tr>
<tr>
<td>Irmak</td>
<td>0.39</td>
<td>0.05</td>
<td></td>
<td>16678</td>
</tr>
<tr>
<td>-F_1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elif</td>
<td>0.66</td>
<td>0.10</td>
<td></td>
<td>41540</td>
</tr>
<tr>
<td>Irmak</td>
<td>0.72</td>
<td>0.09</td>
<td></td>
<td>29017</td>
</tr>
</tbody>
</table>
Conclusion

• Frequent morpheme frames are highly accurate categorization contexts in German and Turkish

• Languages with rich morphology and free word order are not a problem for frequent frames, if analyzed at the right level
A plausible mechanism

• Young children have access to morpheme-level information

• Dutch infants can use frequent morpheme frames to categorize nonsense words (Erkelens, 2009)

• Building syntactic representation from morpheme-level distributions
Bootstrapping to word categories

• Target language dependent distributional analysis

• Integration with other sources of information
  – Phonological cues, semantics, language specific distributional cues

• Frequent frames as a potential universal cue for initial bootstrapping
Acknowledgements

• Dilara Koçbaş for transcribing and coding the Turkish data
• Frauke Berger for labeling German data
• USC Psychology Developmental Brown bag

This research was supported by Turkish Academy of Sciences, in the framework of the Young Scientist Award Program granted to Aylin C. Küntay (EA-TÜBA-GEBİP/2001-2-13) and supported in part by a grant from the National Science Foundation (BCS-0721328) to Toben H. Mintz.
## Category Labels in German

<table>
<thead>
<tr>
<th>Category</th>
<th>Labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adj</td>
<td>Adj</td>
</tr>
<tr>
<td>Adv</td>
<td>Adv</td>
</tr>
<tr>
<td>Conj</td>
<td>Conj</td>
</tr>
<tr>
<td>Det</td>
<td>Det</td>
</tr>
<tr>
<td>Interj</td>
<td>Interj</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Prep</td>
<td>Prep</td>
</tr>
<tr>
<td>Pro</td>
<td>Pro</td>
</tr>
<tr>
<td>Pt</td>
<td>Pt</td>
</tr>
<tr>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td>WH</td>
<td>WH</td>
</tr>
</tbody>
</table>
## Category Labels in Turkish

<table>
<thead>
<tr>
<th>Category</th>
<th>Labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJ</td>
<td>ADV</td>
</tr>
<tr>
<td>ART</td>
<td>CO</td>
</tr>
<tr>
<td>CONJ</td>
<td>EXIST</td>
</tr>
<tr>
<td>INTERJ</td>
<td>N</td>
</tr>
<tr>
<td>NEG</td>
<td>NUM</td>
</tr>
<tr>
<td>POST</td>
<td>PRO</td>
</tr>
<tr>
<td>V</td>
<td>WH</td>
</tr>
</tbody>
</table>

35
# Turkish Morpheme Labels

<table>
<thead>
<tr>
<th>Label</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1P</td>
<td>1st person plural</td>
</tr>
<tr>
<td>1S</td>
<td>1st person singular</td>
</tr>
<tr>
<td>2S</td>
<td>2nd person singular</td>
</tr>
<tr>
<td>ABIL</td>
<td>Abilitative mood</td>
</tr>
<tr>
<td>ABL</td>
<td>Ablative case</td>
</tr>
<tr>
<td>ACC</td>
<td>Accusative case</td>
</tr>
<tr>
<td>AOR</td>
<td>Aorist (present tense marker)</td>
</tr>
<tr>
<td>CAUS</td>
<td>causative</td>
</tr>
<tr>
<td>CM</td>
<td>compound marker</td>
</tr>
<tr>
<td>DAT</td>
<td>dative case</td>
</tr>
<tr>
<td>DIM</td>
<td>diminutive</td>
</tr>
<tr>
<td>FUT</td>
<td>future tense</td>
</tr>
<tr>
<td>GEN</td>
<td>genitive case</td>
</tr>
<tr>
<td>INF</td>
<td>infinitive</td>
</tr>
<tr>
<td>IPFV</td>
<td>imperfective (progressive marker)</td>
</tr>
<tr>
<td>LOC</td>
<td>locative case</td>
</tr>
<tr>
<td>NEG</td>
<td>negation</td>
</tr>
<tr>
<td>OPT&amp;1P</td>
<td>optative and 1st person plural fused</td>
</tr>
<tr>
<td>OPT&amp;1S</td>
<td>optative and 1st person singular fused</td>
</tr>
<tr>
<td>OPT&amp;3S</td>
<td>optative and 3rd person singular fused</td>
</tr>
<tr>
<td>PAST</td>
<td>past tense</td>
</tr>
<tr>
<td>PFV</td>
<td>perfective aspect</td>
</tr>
<tr>
<td>PL</td>
<td>plural</td>
</tr>
<tr>
<td>POSS&amp;1S</td>
<td>possessive and 1st person singular fused</td>
</tr>
<tr>
<td>POSS&amp;2S</td>
<td>possessive and 2nd person singular fused</td>
</tr>
<tr>
<td>POSS&amp;3S</td>
<td>possessive and 3rd person singular fused</td>
</tr>
<tr>
<td>QUE</td>
<td>yes-no question particle</td>
</tr>
<tr>
<td>DIM</td>
<td>diminutive</td>
</tr>
<tr>
<td>FUT</td>
<td>future tense</td>
</tr>
<tr>
<td>GEN</td>
<td>genitive case</td>
</tr>
<tr>
<td>INF</td>
<td>infinitive</td>
</tr>
</tbody>
</table>
# German Results

<table>
<thead>
<tr>
<th></th>
<th>Actual</th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy</td>
<td>Completeness</td>
</tr>
<tr>
<td>FF Word</td>
<td>0.86</td>
<td>0.07</td>
</tr>
<tr>
<td>FF Mor</td>
<td>0.88</td>
<td>0.05</td>
</tr>
<tr>
<td>_F₁F₂ Word</td>
<td>0.47</td>
<td>0.04</td>
</tr>
<tr>
<td>_F₁F₂ Mor</td>
<td>0.32</td>
<td>0.05</td>
</tr>
<tr>
<td>_F₁F₂ Mor</td>
<td>0.78</td>
<td>0.10</td>
</tr>
<tr>
<td>F₁F₂_ Word</td>
<td>0.30</td>
<td>0.07</td>
</tr>
<tr>
<td>F₁F₂_ Mor</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Turkish Results

<table>
<thead>
<tr>
<th></th>
<th>Actual</th>
<th></th>
<th>Random</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Corpus</td>
<td>Accuracy</td>
<td>Tokens</td>
<td>Accuracy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Completeness</td>
<td></td>
<td>Completeness</td>
</tr>
<tr>
<td>FF Word</td>
<td>Elif</td>
<td>0.54</td>
<td>0.09</td>
<td>1269</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>Irmak</td>
<td>0.40</td>
<td>0.11</td>
<td>1656</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.27</td>
</tr>
<tr>
<td>FF Mor</td>
<td>Elif</td>
<td>0.93</td>
<td>0.06</td>
<td>6102</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>Irmak</td>
<td>0.88</td>
<td>0.06</td>
<td>2764</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.38</td>
</tr>
<tr>
<td>Bigram Mor</td>
<td>Elif</td>
<td>0.31</td>
<td>0.05</td>
<td>33793</td>
</tr>
<tr>
<td>F₁⁻</td>
<td></td>
<td></td>
<td></td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>Irmak</td>
<td>0.39</td>
<td>0.05</td>
<td>16678</td>
</tr>
<tr>
<td>F₁</td>
<td></td>
<td></td>
<td></td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>Irmak</td>
<td>0.66</td>
<td>0.10</td>
<td>41540</td>
</tr>
<tr>
<td>F₁⁻</td>
<td></td>
<td></td>
<td></td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>Irmak</td>
<td>0.72</td>
<td>0.09</td>
<td>29017</td>
</tr>
<tr>
<td>F₁</td>
<td></td>
<td></td>
<td></td>
<td>0.29</td>
</tr>
</tbody>
</table>
## German Frequent Frames

<table>
<thead>
<tr>
<th>Frame</th>
<th>Type</th>
<th>Token</th>
<th>Majority</th>
<th>Cat.</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>was__denn</td>
<td>19</td>
<td>56</td>
<td>V</td>
<td></td>
<td>100%</td>
</tr>
<tr>
<td>pass__auf</td>
<td>2</td>
<td>54</td>
<td>Pt</td>
<td></td>
<td>100%</td>
</tr>
<tr>
<td>was__'n</td>
<td>10</td>
<td>53</td>
<td>V</td>
<td></td>
<td>100%</td>
</tr>
<tr>
<td>wir__mal</td>
<td>20</td>
<td>37</td>
<td>V</td>
<td></td>
<td>81%</td>
</tr>
<tr>
<td>wo__denn</td>
<td>7</td>
<td>30</td>
<td>V</td>
<td></td>
<td>100%</td>
</tr>
<tr>
<td>was__Mone</td>
<td>6</td>
<td>24</td>
<td>V</td>
<td></td>
<td>87%</td>
</tr>
<tr>
<td>komm__her</td>
<td>1</td>
<td>23</td>
<td>Pt</td>
<td></td>
<td>100%</td>
</tr>
<tr>
<td>was__der</td>
<td>5</td>
<td>22</td>
<td>V</td>
<td></td>
<td>100%</td>
</tr>
<tr>
<td>ist__der</td>
<td>9</td>
<td>22</td>
<td>Pt</td>
<td></td>
<td>50%</td>
</tr>
<tr>
<td>ist__das</td>
<td>3</td>
<td>22</td>
<td>Pt</td>
<td></td>
<td>95%</td>
</tr>
<tr>
<td>Maxe__dir</td>
<td>8</td>
<td>21</td>
<td>V</td>
<td></td>
<td>100%</td>
</tr>
</tbody>
</table>

...
German Morpheme Frame

\textbf{Maxe\textsubscript{—}-t}

107 tokens 32 types

- mach, ha, gib, trink, zeig, zieh, sag, hol, guck, spritz, hilf, komm, brauch, nimm, setz, faell, tu, fang, krieg, sieh, schneide, bring, spiel, schmier, schlaef, iss, stoess, wisch, träg, schieb, hael, blaes
## Turkish Frequent Frames

<table>
<thead>
<tr>
<th>Frame</th>
<th>Type</th>
<th>Token</th>
<th>Majority</th>
<th>Cat.</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>gel__gel</td>
<td>29</td>
<td>69</td>
<td>V</td>
<td>34%</td>
<td></td>
</tr>
<tr>
<td>burada__var</td>
<td>23</td>
<td>68</td>
<td>WH</td>
<td>72%</td>
<td></td>
</tr>
<tr>
<td>ne__Ekin</td>
<td>32</td>
<td>58</td>
<td>V</td>
<td>79%</td>
<td></td>
</tr>
<tr>
<td>orada__var</td>
<td>34</td>
<td>55</td>
<td>N</td>
<td>52%</td>
<td></td>
</tr>
<tr>
<td>ne__orada</td>
<td>7</td>
<td>53</td>
<td>EXIST</td>
<td>83%</td>
<td></td>
</tr>
<tr>
<td>bir__daha</td>
<td>12</td>
<td>47</td>
<td>N</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>çok__bir</td>
<td>24</td>
<td>41</td>
<td>ADJ</td>
<td>80%</td>
<td></td>
</tr>
<tr>
<td>çok__değilmi</td>
<td>21</td>
<td>39</td>
<td>ADJ</td>
<td>69%</td>
<td></td>
</tr>
<tr>
<td>Ekin__Ekin</td>
<td>18</td>
<td>34</td>
<td>WH</td>
<td>47%</td>
<td></td>
</tr>
<tr>
<td>da__var</td>
<td>29</td>
<td>34</td>
<td>N</td>
<td>88%</td>
<td></td>
</tr>
<tr>
<td>bak__bak</td>
<td>23</td>
<td>33</td>
<td>N</td>
<td>42%</td>
<td></td>
</tr>
</tbody>
</table>

...
Turkish Morpheme Frame

\[ \text{ACC\_PAST} \]

203 tokens 71 types

- gör, bul, at, yap, ver, al, unut, ye, çiz, boya, göster, duy, boz, soy, tak, yakala, koy, kandır, yık, çarp, dön, aç, çağrı, beğen, iç, ara, yıka, öğren, bit, kaldırır, dağıt, ısır, acı, yol, ez, korkut, mahvet, üz, giy, kır, dinle, götür, yala, devir, süsle, sok, yut, sula, gıdıkla, kazan, oy, dik, ek, tamamla, sür, topla, çıkar, kes, parçala, as, seyret, kaydet, de, çal, çık, git, yırt, kaşı, söyle, özle, şaşıır
## English Frequent Frames (Eve)

<table>
<thead>
<tr>
<th>Frame</th>
<th>Type</th>
<th>Token</th>
<th>Major Cat.</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>what__you</td>
<td>9</td>
<td>353</td>
<td>aux</td>
<td>99%</td>
</tr>
<tr>
<td>you__to</td>
<td>20</td>
<td>235</td>
<td>v</td>
<td>70%</td>
</tr>
<tr>
<td>you__it</td>
<td>72</td>
<td>207</td>
<td>v</td>
<td>99%</td>
</tr>
<tr>
<td>you__the</td>
<td>46</td>
<td>119</td>
<td>v</td>
<td>95%</td>
</tr>
<tr>
<td>you__a</td>
<td>27</td>
<td>117</td>
<td>v</td>
<td>93%</td>
</tr>
<tr>
<td>are__doing</td>
<td>3</td>
<td>110</td>
<td>pro</td>
<td>100%</td>
</tr>
<tr>
<td>what__that</td>
<td>6</td>
<td>108</td>
<td>cop</td>
<td>93%</td>
</tr>
<tr>
<td>you__me</td>
<td>16</td>
<td>104</td>
<td>v</td>
<td>100%</td>
</tr>
<tr>
<td>would__like</td>
<td>4</td>
<td>96</td>
<td>pro</td>
<td>97%</td>
</tr>
<tr>
<td>to__it</td>
<td>37</td>
<td>88</td>
<td>v</td>
<td>98%</td>
</tr>
<tr>
<td>you__have</td>
<td>16</td>
<td>87</td>
<td>aux</td>
<td>94%</td>
</tr>
<tr>
<td>you__your</td>
<td>31</td>
<td>87</td>
<td>v</td>
<td>95%</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>