Morphology, lexical priming and second language acquisition:

a corpus-study on learner Finnish

Jarmo Harri Jantunen
Sisko Brunni

Phraseology and lexical priming

Lexical priming (Hoey 2004, 2005)
- Habitual co-occurrence of a word with certain linguistic features
- Is part of our (native speakers') intuitive knowledge of words
- Broadly corresponds to Sinclair's (1996, 2004) postulation of the lexical item
  • One single unit of meaning is comprised of five categories of co-selection (collocations, colligations, semantic preferences, semantic prosody and the core)
Phraseology and lexical priming in learner language

- Several studies have provided evidence that language learners grapple with lexical priming
  - Collocations (e.g. Granger 1997, Hasselgren 1994, Nesselhauf 2005, Jantunen 2007)
  - Problems occur on all proficiency levels
- Most corpus studies of phraseology are biased towards Indo-European languages
  - Studies reveal the problems that learners of those languages have in lexical priming
  - Learner language analyses should be complemented by studies that focus on non-Indo-European languages and features of the unit of meaning other than collocations

Finnish and morphological priming

- Finnish is a synthetic language with a rich morphophonological system (e.g. a single verb may have 260 – 1872 different word forms)
- In morphophonologically rich languages
  - e.g. in Finnish, the core varies widely but in less rich languages the core may be more or less fixed
  - the core is often primed to occur in certain word forms (paradigmatic morphological preference, Jantunen 2004)
- Related to Hoey’s previous hypotheses of lexical priming we postulate that: Every word is primed to occur in particular morphophonological forms; this constitutes morphological priming of words
- In language learning morphological priming is as important as other aspects of lexical priming
Data and methodology

- International Corpus of Learner Finnish (ICLFI)
- Three-Phase Comparative Analysis (TPCA, Jantunen 2004)
  - Mono-mother-tongue Corpus (ICLFI-Estonian)
  - Multi-mother-tongue Corpus (ICLFI-Multi)
  - Native Finnish Corpus
- Corpus-driven analysis
  - WordSmith Tools (KeyWords, Scott 2007)
  - both the noun **IHMINEN** ’person, human, people’ and the verb **PITÄÄ** ’to like, to must, to consider, to hold’ are clearly overused in the learner data

Case 1: Noun **IHMINEN** (‘person, human being, pl. people’)

General overview
1) The frequencies of the grammatical forms (nominative, genitive, partitive cases) and the combined frequency of other cases (12) differ in the subcorpora \((x^2 = 260; p < 0.001; \text{crit.val. 22,5})\).
2) TPCA shows that
   - frequencies of the grammatical forms (nominative, genitive, partitive cases) in both learner data differ from the ones in native data
   - ILCFI-Multi differs more than ILCFI-Estonian from the native data.
A detailed analysis of the grammatical cases (nominative, genitive, partitive):

- in ICLFI-Multi, both singular and plural forms deviate from the native data (exc. PART-sg)
- in ICLFI-Estonian, both singular and plural forms equal the native data (exc. PART-sg)

<table>
<thead>
<tr>
<th>Case</th>
<th>ICLFI-ESTONIAN</th>
<th>ICLFI-MULTI</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOM-sg</td>
<td>-</td>
<td>underuse</td>
</tr>
<tr>
<td>NOM-pl</td>
<td>-</td>
<td>overuse</td>
</tr>
<tr>
<td>GEN-sg</td>
<td>-</td>
<td>underuse</td>
</tr>
<tr>
<td>GEN-pl</td>
<td>-</td>
<td>underuse</td>
</tr>
<tr>
<td>PART-sg</td>
<td>overuse</td>
<td>-</td>
</tr>
<tr>
<td>PART-pl</td>
<td>-</td>
<td>overuse</td>
</tr>
</tbody>
</table>

Overused forms:
1) NOM-pl in ICLFI-Multi:

- **ihmiset** is used instead of the passive voice or other generic expression
  
  *Kirkon jälkeen ihmiset menevät / mennään kotiin syömään.* (‘After church service people go home to have a dinner.’) (See Seilonen 2011.)

- **ihmiset** is used in redundant (explicit) structures
  
  *Kaikki ihmiset / Kaikki pitävät juhlista.* (‘Everyone likes parties.’)

**Ihmiset** collocates with **kaikki** (‘everyone’), **monet** (‘many’), **jokut** (‘some’), i.e. it collocates with quantifying pronouns.
2) PART-sg in ICLFI-Estonian:

*ihmistä* is used in existential clauses in phrases that indicate the number of family members

*Perheessäni on # ihmistä.*
('In my family there are # people."

Instead of *Perheessäni on isä, äiti ja 2 siskoa.*
('In my family there are dad, mom and 2 sisters."

*Ihmistä* colligates with existential clauses and seem to have a semantic preference of ‘family’.

3) PART-PL in ICLFI-Multi:

*ihmisiä* is used in existential clauses which include quantification of people

*Parisissa on paljon ihmisiä.*
('In Paris there are a lot people."

*Talossa asuu paljon vanhoja ihmisiä.*
('In the house there live many old people."

*Ihmisiä* colligates with quantifiers and existential clauses.
Case 2. Verb PITÄÄ (‘to like, to must, to consider, to hold’)

General overview

1) The variety of the grammatical forms (correct and incorrect) in both learner subcorpora differ from the native data.
2) The frequencies of the grammatical forms differ in these subcorpora. Both subcorpora also differ from the native data.
3) The frequencies of different senses of verb differ in subcorpora and both of them differ from native data.
4) Some senses are mastered better in certain verb forms.
5) TPCA shows that ILCFI-multi differs less than ILCFI-Estonian from the native data.

Case 2. Form pidän (1. pers. sg.)

- The frequencies in subcorpora:
  - ICFLI-Estonian: 36%
  - ICFLI-multi: 24%
  - ND: 1%

- The senses found in subcorpora:

<table>
<thead>
<tr>
<th>Sense</th>
<th>ICFLI-Estonian</th>
<th>ICFLI-Multi</th>
<th>Native</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘to like’</td>
<td>99% overuse</td>
<td>90% overuse</td>
<td>42%</td>
</tr>
<tr>
<td>‘to must’</td>
<td>1%</td>
<td>1%</td>
<td>0%</td>
</tr>
<tr>
<td>‘to consider’</td>
<td>0% underuse</td>
<td>6% underuse</td>
<td>22%</td>
</tr>
<tr>
<td>phrase</td>
<td>0% underuse</td>
<td>2% underuse</td>
<td>36%</td>
</tr>
<tr>
<td>incorrect use</td>
<td>0%</td>
<td>1%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Case 2: Form *pidän* at different proficiency levels

<table>
<thead>
<tr>
<th></th>
<th>IFCLI-Est</th>
<th>IFCLI-multi</th>
<th>Native</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b</td>
<td>i</td>
<td>a</td>
</tr>
<tr>
<td>'to like'</td>
<td>99%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>'to must'</td>
<td>1%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>'to consider'</td>
<td>0%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>phrase</td>
<td>0%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>other</td>
<td>0%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

- Phraseological use:
  
  *pidän*: IFCLI-Multi advanced 3% and Native 36%
  
  *pitää*: IFCLI-Multi advanced 19% and Native 20%

---

Case 2. Examples of *pidän* and *pitää*

<table>
<thead>
<tr>
<th></th>
<th>'to like'</th>
<th>phraseological use</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>pidän</em></td>
<td><em>Pidän</em> suklaasta. 'I like chocolate.'</td>
<td><em>(Pidän hauskaa.)</em> 'I have a good time.'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>pitää</em></td>
<td><em>Hän</em> pitää suklaasta. 'S/he likes chocolate.'</td>
<td>*Hän pitää hauskaa. 'S/he has a good time.'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>pitää</em> hauskaa (basic form) 'to have a good time'</td>
<td></td>
</tr>
</tbody>
</table>
Morphophonological mistakes in word inflection

- **IHMINEN:**
  - ICLFI-Multi 6.1%
  - ICLFI-Estonian 5.9%

- **PITÄÄ:**
  - Mistakes in consonant change
  - \(\text{pitää} \sim \text{pidän} \sim \text{pidetään} \sim \text{pidettiin}\)
  - ICLFI-Multi 6.6%
  - ICLFI-Estonian 6.4%

Conclusions based on the two case studies

- Learners are not familiar with the paradigmatic morphological primings of the cores
  - atypical frequencies of the grammatical forms
  - the variation of different senses of verbs may be narrower; however the form affects the variation of senses
  - mother tongue background does not seem to affect the ability to produce correct word forms
  - speakers of the related language seem to prime the noun **IHMINEN** in a more native-like way than learners in general, but in the case of the verb **PITÄÄ**, the result is the opposite
Possible reasons behind the results

1) Learners produce untypical phraseological units and therefore paradigmatic morphological priming differs from the one the native speakers produce
2) Learners use the avoidance strategy
3) Learners often acquire from among the variety of meanings the ones that are concrete and learnt first
4) Certain verb forms can also activate many alternative senses
5) The mastery of phraseological use is related to verb forms
6) Furthermore, teaching, text books and instructions of the task may affect the results

References


