Variability of languages in time and space

Lecture 4: Comparing word formation across languages

Magda Ševčíková

November 2, 2018
Lecture 4 – November 2, 2018

- Language typology, language type
- Approaches to cross-linguistic study of word formation
  1. productivity-based approaches
  2. attestedness of word-formation processes across languages
  3. expression of basic concepts across languages
  4. onomasiological approach
Körtvélyessy (2017:2):  

“Language typology is a system or study that divides languages into smaller groups according to similar properties they have. [...] These smaller groups are called language types.”

- a holistic approach to language typology  
  “The classification of languages into language types attempts to ‘match’ the complete language system with one language type.”

- a partial approach to language typology  
  “the classification is based on the analysis of a selected language construction and/or phenomenon (not the entire language), for example the size of the consonantal inventory, the presence vs. absence of articles in language, the order of words in a sentence etc.”

- language universals = features that are shared by all natural languages in the world (Haspelmath et al. 2001)  
  The Universals Archive https://typo.uni-konstanz.de/archive/intro/index.php

sevcikova@ufal.mff.cuni.cz  |  Variability of languages
Comparing word formation across languages

- Detailed linguistic descriptions of word-formation systems available for esp. Indo-European languages
- Only 1 derivational feature in WALS (reduplication as one of morphological features)
- Cross-linguistic study / linguistic typology of word formation very recent
1. productivity-based approaches – no satisfactory results
2. attestedness of individual word-formation processes across languages
   - 55 languages from 28 families (Štekauer et al. 2012)
   - saturation value for Slavic languages (Körtyélyessy 2016)
3. derivational potential of a sample of underived words in individual languages
   - Monika project (40 European languages)
4. onomasiological approach
   - Dokulil 1962, Štekauer 1998
   - onomasiological types (Štekauer 1998, 2016)
   - comparative semantic concepts (Bagasheva 2017)
1/ Productivity-based approaches

- productivity as “the possibility for language users, by means of a morphological process which underpins a form-meaning correspondence in some words they know, to coin, unintentionally, a number of new formations which is in principle infinite” (Schultink 1961:113)
Baayen’s productivity measures

- **category-conditioned degree of productivity (Baayen 1992):**
  \[ P = \frac{n_1}{N} \]
  - \( n_1 \) number of hapax legomena with the particular suffix (words that occur just once in a corpus)
  - \( N \) token frequency (number of all tokens containing the suffix under analysis)

- **hapax-conditioned degree of productivity (Baayen 1993):**
  \[ P^* = \frac{n_{1,E,t}}{h_t} \]
  - \( n_{1,E,t} \) number of hapax legomena with a certain suffix
  - \( h_t \) total number of hapaxes in the corpus
  - “Denoting the number of hapaxes observed for category E after \( t \) tokens of the corpus have been sampled by \( n_{1,E,t} \), and denoting the total number of hapaxes of arbitrary constituency in these \( t \) observations by \( h_t \), we find that the required conditional probability, say \( P^* \), equals \( \frac{n_{1,E,t}}{h_t} \).”
discussion and objections:

- rejection of the possibility to derive productivity from frequencies (van Marle 1992, Dressler – Ladanyi 2000)
- debatable nature of hapax legomena (Dal 2003)
- impact of the data size
- problems of automatic preprocessing of the data (Evert – Lüdeling 2001)
- limited applicability to low-frequency words (Fernandez-Dominguez et al. 2007)

- variable-corpus approach (Gaeta – Ricca 2006)
- combinations of quantitative and qualitative analysis (Lüdeling – Evert 2005, Plag 1999)
Attestedness of word-formation processes across languages

- Štekauer et al. (2012) studied word formation across **55 languages**
  - from 28 language families and 45 language genera (classification based on WALS)

- similarities and differences among languages evaluated in terms of presence vs. absence of individual word-formation processes
  - in which and in how many languages from the sample, a word-formation process is attested?
some form of **derivation** attested in all but one languages in the sample of 55 languages

- no affixation at all in Vietnamese (isolating language), only prefixation but no suffixation in Yoruba (isolating language)
- the significance of derivation varies across languages (about 300 suffixes in Slovene, 1 genuine prefix in Finnish - negation)

**compounding**

- 91 % of languages in the sample

**reduplication** was found very frequently

- 80 % of languages in the sample

**conversion**

- 62 % of languages in the sample

**stress and tone / pitch** are minor in word formation

- with 7 and 13 % of languages, respectively
saturation value indicates the degree to which a particular word-formation system makes use of all the word-formation options under examination

- for Slavic languages (Körtvélyessy 2016)

which and how many word-formation processes are attested in a language

- Körtvélyessy’s study (2016) based on representative descriptions of particular word-formation systems in Müller et al. (2016)

absence/presence of a word-formation process in a language (in POS terms)

the productivity of a word-formation process not taken into consideration

- cf. prefixation vs. postfixation in Czech
Körtvélyessy (2016:483ff):

<table>
<thead>
<tr>
<th>feature</th>
<th>mkd</th>
<th>bos</th>
<th>slv</th>
<th>hrv</th>
<th>srp</th>
<th>bul</th>
<th>hsb</th>
<th>pol</th>
<th>csb</th>
<th>ces</th>
<th>slk</th>
<th>ukr</th>
<th>bel</th>
<th>rus</th>
<th>SAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>N&gt;N</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>14</td>
</tr>
<tr>
<td>V&gt;V</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>14</td>
</tr>
<tr>
<td>A&gt;A</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>14</td>
</tr>
<tr>
<td>Adv&gt;Adv</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>SAT</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>A&gt;N</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>V&gt;N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Adv&gt;N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>A&gt;V</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>N&gt;V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Adv&gt;V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>N&gt;A</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>V&gt;A</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Adv&gt;A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>N&gt;Adv</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>V&gt;Adv</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>A&gt;Adv</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>SAT</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>total SAT</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

number of lang.: 14
number of features: 17
total saturation value: 59
average saturation value (total sat. value / number of lang.): 4.214
relative saturation value (total sat. value / (number of features * number of lang.)): 24.79 %
3/ Derivational potential of a sample of underived words

- Monica project
  https://www.ugr.es/~svalera/Monika/index.html
  - 40 European languages
  - 30 sample words selected from Swadesh list
    - 10 nouns (bone, eye, fire, water, name ...)
    - 10 verbs (cut, give, hold, drink, think ...)
    - 10 adjectives (bad, new, black, warm, long ...)
  - what are the counterparts of these words in individual languages?
  - which words are derived from these words?
4/ Onomasiological approach

  - the act of naming is followed
  - how is a particular concept expressed in a language? which naming
    strategy is chosen by the speaker?
- Dokulil (1962)
  - onomasiological categories of substance, quality, circumstance, and
    action
  - naming strategies modelled as onomasiological types
  - economy of expression vs. semantic transparency as two contradictory
    tendencies
- Bagasheva (2017)
  - 50+ comparative semantic concepts applicable in cross-linguistic
    research into affixation
<table>
<thead>
<tr>
<th>OT1</th>
<th>DingC</th>
<th>DedC</th>
<th>Base</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example</td>
<td>Instrument</td>
<td>Action</td>
<td>Agent</td>
</tr>
<tr>
<td></td>
<td>guitar</td>
<td>play</td>
<td>er</td>
</tr>
<tr>
<td>OT2</td>
<td>DingC</td>
<td>DedC</td>
<td>Base</td>
</tr>
<tr>
<td>Example</td>
<td>0</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>Instrument</td>
<td>Action</td>
<td>Agent</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>play</td>
<td>er</td>
</tr>
<tr>
<td>OT3</td>
<td>DingC</td>
<td>DedC</td>
<td>Base</td>
</tr>
<tr>
<td>Example</td>
<td>Instrument</td>
<td>Action</td>
<td>Agent</td>
</tr>
<tr>
<td></td>
<td>guitar</td>
<td>0</td>
<td>ist</td>
</tr>
</tbody>
</table>

...
<table>
<thead>
<tr>
<th>Semantic Concept</th>
<th>English Example</th>
<th>Bulgarian Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action</td>
<td>reading</td>
<td>strelba</td>
</tr>
<tr>
<td>Agent</td>
<td>killer</td>
<td>ubiec</td>
</tr>
<tr>
<td>Abstraction</td>
<td>justice</td>
<td>pravda</td>
</tr>
<tr>
<td>Causative</td>
<td>empower</td>
<td>zaliva</td>
</tr>
<tr>
<td>Composition</td>
<td>orehovka</td>
<td></td>
</tr>
<tr>
<td>Diminutive</td>
<td>piglet</td>
<td>pospya</td>
</tr>
<tr>
<td>Hyperonymy</td>
<td>archbishop</td>
<td>nadreden</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


sevcikova@ufal.mff.cuni.cz

Variability of languages

https://unibook.upjs.sk/sk/anglistika-a-amerikanistika/365-essentials-of-language-typology


