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

Language typology, language type

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`

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

Approaches to cross-linguistic study of word formation

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 (Körtvélyessy 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)

1/ Baayen's productivity measures

- category-conditioned degree of productivity (Baayen 1992):
  $\mathbf{P} = n_1/\mathbf{N}$
  - 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):
  $\mathbf{P^*} = n_{1,E,t}/h_t$
  - $n_{1,E,t}$ 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 $n_{1,E,t}/h_t$."

Discussion, alternative approaches

- 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)

2/ 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?

2/ Typological conclusions by Štekauer et al. 2012

- 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

2/ Saturation value

- 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

2/ Saturation value: prefixation in Slavic languages

Körtvélyessy (2016:483ff):

feature	mkd	bos	slv	hrv	srp	bul	hsb	pol	csb	ces	slk	ukr	bel	rus	SAT
N>N	X	X	X	X	X	X	X	X	X	X	X	X	X	X	14
V>V	X	X	X	X	X	X	X	X	X	X	X	X	X	X	14
A>A	X	X	X	X	X	X	X	X	X	X	X	X	X	X	14
Adv>Adv				X	X					X	X	X	X	X	7
SAT	3	3	3	4	4	3	3	3	3	4	4	4	4	4	
A>N				X											1
V>N				X											1
Adv>N															0
A>V										X	X				2
N>V	X														1
Adv>V															0
N>A									X						1
V>A				X						X	X				3
Adv>A															0
N>Adv															0
V>Adv															0
A>Adv							X								1
SAT	1	0	0	3	0	0	1	0	1	2	2	0	0	0	
total SAT	4	3	3	7	4	3	4	3	4	6	6	4	4	4	

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

- Dokulil (1962), Štekauer (1998)
  - 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
- Štekauer (1998, 2016)
  - 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

4/ Onomasiological types (Štekauer 1998, 2016)

OT1	DingC	DedC	Base
	R	R	R
Example	Instrument	Action	Agent
	guitar	*play*	*er*
OT2	DingC	DedC	Base
	0	R	R
Example	Instrument	Action	Agent
	0	*play*	*er*
OT3	DingC	DedC	Base
	R	0	R
Example	Instrument	Action	Agent
	guitar	0	*ist*

...

4/ Semantic concepts in affixation (Bagasheva 2017)

Action	En. *read**ing***, Bul. *strel**ba***
Agent	En. *kill**er***, Bul. *ubi**ec***
Abstraction	En. *just**ice***, Bul. *prav**da***
Causative	En. ***em**power*, Bul. ***za**liva*
Composition	Bul. *oreh**ovka***
Diminutive	En. *pig**let***, Bul. ***po**spya*
Hyperonymy	En. ***arch**bishop*, Bul. ***nad**reden*
...	

References

- Baayen, H. (1992): Quantitative aspects of morphological productivity. In: G. E. Booij – J. van Marle (eds.): *Yearbook of Morphology 1991*. Dordrecht: Kluwer, pp. 109–149.
- Baayen, H. (1993): On frequency, transparency, and productivity. In: G. E. Booij – J. van Marle (eds): *Yearbook of Morphology 1992*. Dordrecht: Kluwer, pp.181–208.
- Bagasheva, A. (2017): Comparative semantic concepts in affixation. In J. Santana-Lario & S. Valera-Hernández (eds.): *Competing Patterns in English Affixation*. Bern – Berlin: Peter Lang, pp. 33–65.
- Dal, G. (2003): Productivité morphologique: définitions et notions connexes. *Langue française*, 140, pp. 3–23.
- Dokulil, M. (1962): *Tvoření slov v češtině 1: Teorie odvozování slov*. Praha: Nakl.ČSAV.
- Dressler, W. U. – Ladányi, M. (2000): Productivity in word formation: a morphological approach. *Acta Linguistica Hungarica*, 47, pp. 103–144.
- Dryer, M. S. – Haspelmath, M. (eds., 2013): *The World Atlas of Language Structures Online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info
- Evert, S. – Lüdeling, A. (2001): Measuring morphological productivity: is automatic preprocessing suffiecient? In P. Rayson et al. (eds.): *Proceedings of the Corpus Linguistics 2001 Conference*. Lancaster: Peter Lang, pp. 167–175.
- Fernández-Domínguez, J. et al. (2007): How is Low Productivity Measured? *Atlantis*, 29, pp. 29–54.
- Gaeta, L. – Ricca, D. (2006): Productivity in Italian word-formation: A variable-corpus approach. *Linguistics*, 44, pp. 57–89.
- Haspelmath, M. et al. (eds.; 2001): *Language Typology and Language Universals*. Mouton de Gruyter.

- Körtvélyessy, L. (2016): Word-formation in Slavic languages. *Poznań Studies in Contemporary Linguistics*, 52, s. 455–501.
- Körtvélyessy, L. (2017): *Essentials of Language Typology*. Košice: UPJŠ. https://unibook.upjs.sk/sk/anglistika-a-amerikanistika/365-essentials-of-language-typology
- Lüdeling, A. – Evert, S. (2005): The emergence of productive non-medical -itis. Corpus evidence and qualitative analysis. In S. Kepser & M. Reis (eds.): *Linguistic Evidence. Empirical, Theoretical and Computational Perspectives*. Berlin – Boston: Mouton De Gruyter, pp. 351–370.
- van Marle, J. (1992): The relationship between morphological productivity and frequency: A comment on Baayen's performance-oriented conception of morphological productivity. In G. E. Booij & J. van Marle (eds.): *Yearbook of morphology 1991*. Dordrecht: Kluwer Academic Publishers, pp. 151–163.
- Müller, P. O. et al. (eds.; 2016): *Word-Formation. An International Handbook of the Languages of Europe*. Volume 4. Berlin: de Gruyter.
- Plag, I. (1999): *Morphological productivity. Structural constraints in English derivation*. Berlin, New York: Mouton de Gruyter.
- Schultink, H. (1961): Produktiviteit als morpfologisch fenomeen. *Forum der Letteren*, 2, pp. 110–125.
- Štekauer, P.(1998): *An Onomasiological Theory of English Word-formation*. Amsterdam – Philadelphia: John Benjamins Publishing Company.
- Štekauer, P. (2016): Compounding from an onomasiological perspective. In P. ten Hacken (ed.): *The Semantics of Compounding*. Cambridge: CUP, pp. 54–68.
- Štekauer, P. et al. (2012): *Word-Formation in the World's Languages*. Cambridge: CUP.