Variability of languages in time and space

Lecture 7:

Linguistic typology: Word-formation

Magda Ševčíková

October 29, 2019

- Word formation
  - Adding bound lexical morphemes (affixation)
  - Combining free lexical morphemes (compounding etc.)
  - Without addition of derivational material (conversion etc.)
- 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

- Word formation
  - Adding bound lexical morphemes (affixation)
  - Combining free lexical morphemes (compounding etc.)
  - Without addition of derivational material (conversion etc.)
- 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

Word formation

Štekauer & Lieber (2005:212)

"Word-formation deals with productive and rule-govenered patterns (word-formation types and rules, and morphological types) used to generate motivated naming units in response to the specific naming needs of a particular speech community by making use of word-formation bases of bilateral naming units and affixes stored in the Lexical Component."

= **onomasiological approach** to word formation

vs. **semasiological approach** that proceeds from the already exsting words to their meanings

Morphemes in word formation

- Words are formed by using both types of lexical morphemes
  - **free lexical morphemes** (content words)
    - can function as words, or be combined with other morphemes as **roots**
  - **bound lexical morphemes** (derivational morphemes)
    - cannot be used separately
    - combined (as **affixes**) with free morphemes

Morphemes around one root

- ex. the morphemic structure of the words:
  - *chair*, *chairs*, *dismissed*
  - Czech *nahořklý* 'slightly bitter', *neuvěřitelný* 'unbelievable'

prefix	root	suffix
	chair	
	chair-	*-s*
dis-	*-miss-*	*-ed*
na-	*-hořk-*	*-lý*
ne- u-	*-věř-*	*-i- -teln- -ý*

Morphemes around more roots

- ex. the morphemic structure of the following compounds:
  - German *Abschlussprüfung* 'final exam'
  - German *Jahresabschluss* 'end of the year'
  - Czech *modrooký* 'blue-eyed'

prefix	root	interfix	prefix	root	suffix
Ab-	*-schluss-*			*-prüf-*	*-ung*
	Jahr-	*-es-*	*-ab-*	*-schluss*	
	modr-	*-o-*		*-ok-*	*-ý*

Word-formation processes

- Štekauer et al. (2012) distinguish three groups of word-formation processes according to which type of morphemes is used:
  1. adding bound lexical morphemes (derivational affixes):
     1. affixation / derivation
  2. combinig free morphemes (roots):
     1. compounding
     2. reduplication
     3. blending
  3. without addition of derivational material:
     1. conversion
     2. stress, tone/pitch

Ad 1.1: Affixation

- Affixation / derivation

  is formation of new lexemes by **adding bound lexical morphemes** to a morpheme or to a word in order

  (1) to **change its part-of-speech category**
  - *bad*.adj > *badly*.adv
  - *špatný* 'bad' > *špatně* 'badly'

  (2) to **modify or add a non-grammatical meaning to it**
  - *child*.noun > *childhood*.noun
  - *učitel* 'teacher' > *učitelka* 'female teacher'

  (3) to do **both**
  - *child*.noun > *childish*.adj
  - *dítě* 'child' > *dětský* 'childish'

Ad 1.1: Base word and derivative, motivation and foundation

- The word that enters the derivation is called a **base word**. The word that results from derivation is a **derivative**.
- The base word and the derivative are related both formally and semantically (Dokulil 1962):
  - the meaning of the derivative based on the meaning of the base word = **motivation**
  - the form of the derivative based on the form of the base word = **foundation**

Ad 1.1: Direction of derivation

- The direction of derivation is determined by applying the following assumptions:
  - the base word is expected to have a **simpler morphemic structure** than the derivative
  - the base word is expected to have a **broader meaning** than the derivative
- Additional, empirically observed features can be employed, e.g:
  - the base word is often **more frequent** than the derivative
- Examples:
  - *child* (47,629) > *child*<u>*hood*</u> (642) "state/period of being a child"
  - *large* (26,212) > *to* <u>*en*</u>*large* (503) "to make larger"

– absolute frequency (in parentheses) based on the English section of the InterCorp corpus v10 (Klégr et al. 2017)

Ad 1.1: Types of derivation

- prefixation
  - a bound morpheme (prefix) is attached to the front of a word or of a free morpheme
- suffixation
  - a bound morpheme (suffix) is attached to the end of a word or of a free morpheme
- circumfixation
  - prefix and a suffix are added in one step
    - neither the prefix and the root nor the suffix and the root are attested alone
- infixation
  - a bound morpheme (infix) inserted into a free morpheme

Ad 1.1: Prefixation

- in English (Bauer 1983)
  - majority of prefixes of Latin and Greek origin
    - $moral > \underline{a}moral$
    - $act > \underline{inter}act$
  - native prefixes from prepositions
    - $line > \underline{under}line$
    - $load > \underline{over}load$
  - a continuum between prefixes (*prefixoids*) and first parts of compounds (*neoclassical formations*)
    - $psycho\text{-},\ eco\text{-},\ techno\text{-}$

Ad 1.1: Prefixation in Slavic languages

- mostly adding a semantic feature without changing the part-of-speech category
  - class-maintaining process
    - *veliký*.adj 'big' > *převeliký*.adj 'very big'
    - *psát*.verb 'write' > *zapsat*.verb 'write down'
- highly productive with verbs
  - e.g. Czech:
    - *psát* 'write' > *dopsat* 'finish writing'
    - *psát* 'write' > *připsat* 'add by writing'
    - *psát* 'write' > *vypsat* 'excerpt'
    - *psát* 'write' > *podepsat* 'sign'
    - *psát* 'write' > *nadepsat* 'entitle'
    - *psát* 'write' > *upsat (se)* 'subscribe'
    - *psát* 'write' > *vepsat* 'insert by writing'

Ad 1.1: Suffixation

- either as an addition of the suffix, or replacement of a suffix for another one
  - Czech *učitel* 'teacher' > *učitel<u>ka</u>* 'female teacher'
  - Czech *tanečník* 'dancer' > *tanečn<u>ice</u>* 'female dancer'
- both class-maintaing and class-changing process
  - German *Tänzer*.noun 'dancer' > *Tänzer<u>in</u>*.noun 'female dancer'
  - *work*.verb > *work<u>er</u>*.noun

Ad 1.1: Multiple prefixation and suffixation

- words can be derived through a sequence of prefixation or suffixation steps applied successively
  - prefixation and suffixation
    - $taste > tasteful > tastefully > distastefully$
      or cf. an alternative analysis:
      $taste > tasteful > distasteful > distastefully$
  - multiple prefixation
    - Czech *skočit* 'jump' > *vyskočit* 'jump up' > *povyskočit* 'jump up a little'
  - multiple suffixation
    - Czech *strom* 'tree' > *stromek* 'small tree' > *stromeček* 'very small tree'

Ad 1.1: Circumfixation

- derivation of collective nouns in Tagalog (Štekauer et al. 2012)
  - *Intsik* 'Chinese person' > *ka-intsik-an* 'the Chinese'
  - *pulo* 'island' > *ka-pulu-an* 'archipelago'
- derivation of adjectives of small portion of quality
  - *drzý* 'impudent' > *při-drzlý* 'slightly impudent'
  - neither **drzlý* nor **přidrzý* attested in Czech

  – must be distinguished from subsequent affixation:

    cf. suffixation followed by prefixation
    - Czech *otrávit*.verb 'poison' > *při-otrávit*.verb 'poison partially' > *přiotráv-ený*.adj 'partially poisoned'

Ad 1.1: Infixation

- an infix inserted before the last syllable to derive a negative in Hua (Štekauer et al. 2012)
  - *zgavo* 'embrace' > *zga-'a-vo* 'not embrace'
  - *harupo* 'slip' > *haru-'a-po* 'not slip'

Ad 2.1: Compounding

- Two (or more) free morphemes are combined to form a new lexeme
  - a compound prototypically consists of two parts
    – two root morphemes
      - first / left-hand part vs. second / right-hand part
    – with or without a linking element
- attested across languages, but delimited differently
- borders to other areas are not clear-cut
  - to derivation
    - cf. elements *eco-*, *techno-*, *agro-* interpreted either as prefixes or as first parts of compounds
  - to syntax
    - cf. *flower pot*, *flower-pot*, *flowerpot* (Lieber – Štekauer 2009)

Ad 2.1: Delimiting compounds in English

- Lieber (2005) discusses criteria used for delimitation of compounds in English – most of them are problematic:
  - stress (on the first part)
    - *trúck driver*, *ápple cake* (but *apple píe*)
  - spelling
    - varies a lot: *daisy wheel*, *daisy-wheel*, *daisywheel*
  - lexicalized meaning
    - not applicable to new compounds
  - unavailability of the first part to inflection, anaphora and coordination
    - but *children's hour*, *medical and life insurance*
  - inseparability of the first and second part
    - *truck driver* – **truck fast driver*

Ad 2.2: Reduplication

- A free morpheme is repeated to form a new word.
  - attested both in derivation and in inflection
  - more frequent in derivation
  - different functions:
    - Italian *neri neri* 'really black'
    - Czech *šir-o-šir-ý* 'extremely vast'
    - Spanish *Es un coche-coche* (is-a-car-car) 'It is a very good car'
    - Indonesian *buah-buah-an* (fruit-fruit) 'various sorts of fruit'

Ad 2.3: Blending

- Two free morphemes are reduced and joined to form a new word
  - En. $\underline{smo}ke + \underline{fog} > smog$
  - En. $\underline{br}eakfast + \underline{lunch} > brunch$

    – the base morphemes often overlap in one ore more phonemes/graphemes
  - French $\underline{photocopy} + \underline{pillage} > photocopillage$ 'illegal photocopying'
  - Italian $\underline{cantante} + \underline{autore} > cantautore$ 'singer-songwriter'

Ad 3.1: Conversion

- A new word is coined simply by the change of the part-of-speech category
  - run.verb $>$ run.noun
  - in languages with inflectional morphology, the change of the part-of-speech category can be seen as the change of the set of inflectional features (change of inflectional paradigm) = **transflexion**
    - Czech $zlý$.adj 'evil' $>$ zlo.noun 'evil'
    - German $schlafen$.verb 'sleep' $>$ $Schlaf$.noun 'sleep'

Ad 3.2: Stress and tone / pitch

- Rarely, the replacement of stress is used to form new words
  - e.g. in Vietnamese, or
  - cf. En. $recórd$.verb $>$ $récord$.noun
    - rather classified as conversion

Language typology of word-formation?
Comparing word-formation across languages

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**."

- 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 $n_{1,E,t}/h_t$."

1/ 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.