Variability of languages in time and space

Word formation across languages
Approaches to cross-linguistic study of word formation

Magda Ševčíková

October 23, 2025

Word formation

Štekauer & Lieber (2005:212)

"Word-formation deals with productive and rule-governed 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."

Word formation vs. formation of word forms

BNC via KonText `[word="treat.*"]`

1	treatment	11,609
2	treated	6,914
3	treat	3,527
4	treaty	2,957
5	treating	1,260
6	treatments	839
7	treaties	570
8	treats	534
9	treatise	160
10	treatises	58
11	treatable	43

SYN2015 via KonText `[word="lé[kč].*"]`

1	lékař	4,758
2	lékaři	4,400
3	lékaře	4,095
4	léky	3,320
5	lékařů	3,005
6	lékařské	1,988
7	léčby	1,918
8	léků	1,840
9	léčbu	1,814
10	léčbě	1,600
11	lék	1,562

BNC via KonText `[lemma="treat.*"]`

1	treatment	12,985
2	treat	12,312
3	treaty	5,626
4	treatise	283
5	treated	164
6	treatable	43
7	treaty-making	20
8	treatment-room	10
9	treatment-resistant	3
10	treating	3
11	treaty-based	3

SYN2015 via KonText `[lemma="lé[kč].*"]`

1	lékař	12	léčebna
2	lék	13	lékárník
3	léčba	14	lékařství
4	lékařský	15	léčitel
5	léčit	16	léčený
6	lékárna	17	lékárnička
7	léčebný	18	léčka
8	léčivý	19	léčitelství
9	léčení	20	lékový
10	léčivo	21	lékopis
11	lékařka	22	lékařův

- Morphemes in word formation
- Word-formation processes
    1. Adding bound lexical morphemes (affixation)
    2. Combining free lexical morphemes (compounding etc.)
    3. Without addition of derivational material (conversion etc.)
- Approaches to cross-linguistic study of word formation
    - Productivity-based approaches
    - Attestedness of word-formation processes across languages
    - Derivational potential of a sample of underived words

Types of morphemes

two oppositions combined:

- **grammatical** vs **lexical** morphemes
  - grammatical morphemes change inflection
  - lexical morphemes have (more or less general) lexical meanings on their own
- **bound** vs **free** morphemes
  - a bound morpheme cannot stand alone
  - a free morpheme can stay as a single word

Grammatical morphemes: free vs bound

- **bound grammatical morphemes**
  - = "inflectional morphemes" (endings etc.)
  - add inflectional features without changing lexical meaning: used to create word forms of a given lexeme with the same lexical meaning but different inflections
  - often more than one inflectional meaning (portmanteaus)
  - occur outside derivational morphemes

  e.g. *play-s*, *play-ed*, *play-ing*; *play-er-s*, *book-s*, *dis-lik-ed*
- **free grammatical morphemes**
  - = "function words"

  e.g. *in a book*, *but*, *that*, *them*

Lexical morphemes: free vs bound

lexical morphemes have a lexical meaning by themselves

- **free lexical morphemes**
  - = "content words" (roots and stems)
  - e.g. *book*, *book-s*, *play*, *play-er-s*
- **bound lexical morphemes**
  - = "derivational morphemes" (derivational prefixes, suffixes etc.)
  - cannot be used separately
  - combined (as **affixes**) with free morphemes to form a new word
  - change the meaning and/or the part-of-speech category of words
  - have specialized meanings, added in succession
  - derivational suffixes occur before inflectional morphemes
  - e.g. *book-ish*, *play-er-s*, *dis-lik-ed*; Cz. *uči-tel-k-a*

Morphemes around the root(s)

- En. *chair*, *chairs*, *dismissed*; Cz. *nahořklý* 'slightly bitter', *neuvěřitelný* 'unbelievable'

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

- Ger. *Abschlussprüfung* 'final exam', *Jahresabschluss* 'end of the year'; Cz. *modrooký* 'blue-eyed'

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

- Morphemes in word formation
- Word-formation processes
  1. Adding bound lexical morphemes (affixation)
  2. Combining free lexical morphemes (compounding etc.)
  3. Without addition of derivational material (conversion etc.)
- Approaches to cross-linguistic study of word formation
  - Productivity-based approaches
  - Attestedness of word-formation processes across languages
  - Derivational potential of a sample of underived words

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)
     = affixation / derivation
     1.1 prefixation
     1.2 suffixation
     1.3 circumfixation
     1.4 infixation
  2. combinig free morphemes (roots):
     2.1 compounding
     2.2 reduplication
     2.3 blending
  3. without addition of derivational material:
     3.1 conversion
     3.2 stress, tone/pitch

1. Affixation / derivation

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

(a) to **change its part-of-speech category**

bad.adj $>$ $bad\underline{ly}$.adv
$špatný$ 'bad' $>$ $špatn\underline{ě}$ 'badly'

(b) to **modify or add a non-grammatical meaning to it**

$child$.noun $>$ $child\underline{hood}$.noun
$učitel$ 'teacher' $>$ $učitel\underline{ka}$ 'female teacher'

(c) to do **both**

$child$.noun $>$ $child\underline{ish}$.adj
$dítě$ 'child' $>$ $dět\underline{ský}$ 'childish'

Direction in derivation

base word = the input of derivation *vs* **derivative** = the output of derivation

the derivative is based both formally and semantically on the base word
= **motivation**

- the base word expected to have a **simpler morphemic structure** than the derivative
- the base word expected to have a **broader meaning** than the derivative
- plus other features be employed, e.g. corpus frequency
  - the base word is often **more frequent** than the derivative
    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 freq from the InterCorp corpus v10; Klégr et al. 2017)

1.1 Prefixation

= a bound morpheme (prefix) is attached to the front of a word or of a free morpheme

- 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 and first parts of compounds (*neoclassical formations*): $psycho$-, eco-, $techno$-
- in Slavic languages
  - mostly without changing the part-of-speech category

    $veliký$.adj 'big' $> \underline{pře}veliký$.adj 'very big'
    $psát$.verb 'write' $> \underline{za}psat$.verb 'write down'
  - highly productive with verbs

    Cz. $psát$ 'write' $> \underline{do}psat$ 'finish writing' | $\underline{při}psat$ 'add by writing' | $\underline{vy}psat$ 'excerpt' | $\underline{pode}psat$ 'sign' | $\underline{nade}psat$ 'entitle' | $\underline{u}psat\ (se)$ 'subscribe' | $\underline{ve}psat$ 'insert by writing'

1.2 Suffixation

= a bound morpheme (suffix) is attached to the end of a word or of a free morpheme
  - Cz. *učitel* 'teacher' > *učitel*$\underline{ka}$ 'female teacher'

- both as a class-maintaining or a class-changing process
  - Ger. *Tänzer*.noun 'dancer' > *Tänzer*$\underline{in}$.noun 'female dancer'
  - En. *work*.verb > *work*$\underline{er}$.noun

Multiple prefixation and suffixation

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

1.3 Circumfixation

= prefix and a suffix are added in one step

but neither the prefix and the root nor the suffix and the root are attested alone

- derivation of collective nouns in Tagalog (Štekauer et al. 2012):
  - *Intsik* 'Chinese person' > *kaintsikan* 'the Chinese'
  - *pulo* 'island' > *kapuluan* 'archipelago'
- derivation of adjectives of small portion of quality
  - Cz. *drzý* 'impudent' > *přidrzlý* 'slightly impudent', but neither **drzlý* nor **přidrzý* exist
  - must be distinguished from subsequent affixation:

    cf. suffixation followed by prefixation in Cz.
    otrávit.verb 'poison' > *přiotrávit*.verb 'poison partially' > *přiotrávený*.adj 'partially poisoned'

1.4 Infixation

= a bound morpheme (infix) inserted into a free morpheme

- 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'

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)

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*

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:
  - It. *neri neri* 'really black'
  - Cz. *šir-o-šir-ý* 'extremely vast'
  - Sp. *Es un coche-coche* (is-a-car-car) 'It is a very good car'
  - Indonesian *buah-buah-an* (fruit-fruit) 'various sorts of fruit'

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
  - Fr. $\underline{photocopy} + \underline{pillage} > photocopillage$ 'illegal photocopying'
  - It. $\underline{cantan}te + \underline{autore} > cantautore$ 'singer-songwriter'

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**

  – Cz. $zlý$.adj 'evil' $>$ zlo.noun 'evil'

  – Ger. $schlafen$.verb 'sleep' $>$ $Schlaf$.noun 'sleep'

3.2 Stress and tone / pitch

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

- Morphemes in word formation
- Word-formation processes
  - 1. Adding bound lexical morphemes (affixation)
  - 2. Combining free lexical morphemes (compounding etc.)
  - 3. Without addition of derivational material (conversion etc.)
- Approaches to cross-linguistic study of word formation
  - Productivity-based approaches
  - Attestedness of word-formation processes across languages
  - Derivational potential of a sample of underived words

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

i. productivity-based approaches
ii. attestedness of individual word-formation processes across languages
    - 55 languages from 28 families (Štekauer et al. 2012)
    - saturation value (Körtvélyessy 2016, Körtvélyessy et al. 2020)
iii. derivational potential of a sample of underived words in individual languages
    - 40 European languages (Körtvélyessy et al. 2020)

i. Productivity-based approaches

Productivity (Schultink 1961:113)

"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"

- category-conditioned degree of productivity P $= n_1/\mathrm{N}$ (Baayen 1992)
  - 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 P* $= n_{1,E,t}/h_t$ (Baayen 1993)
  - $n_{1,E,t}$ 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 $n_{1,E,t}/h_t$."

ii. 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?

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 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)
  - for 40 European langs (Körtvélyessy et al. 2020)
- 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

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: 16
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 %

iii. Derivational potential of a sample of underived words

derivational networks in 40 European languages (Körtvélyessy et al. 2020)

- composed of an unmotivated word and all its direct and indirect derivatives
- unmotivated words selected from Swadesh list
  - 10 nouns: *bone, eye, tooth, day, dog, louse, fire, stone, water, name*
  - 10 verbs: *cut, dig, pull, throw, give, hold, sew, burn, drink, know*
  - 10 adjectives: *bad, new, black, straight, warm, old, long, thin, thick, narrow*
- three dimensions of the derivational network:

  1/ derivatives organized into **derivational series** (= a set of words directly motivated by the same base but not mutually motivating one another) ... horizontal dimension of the network
  2/ derivatives organized into **derivational paradigms** (= a set of words that share a common root and each of them motivates the item that immediately follows it) ... vertical dimension of the network
  3/ semantic category added through the affix ... semantic dimension

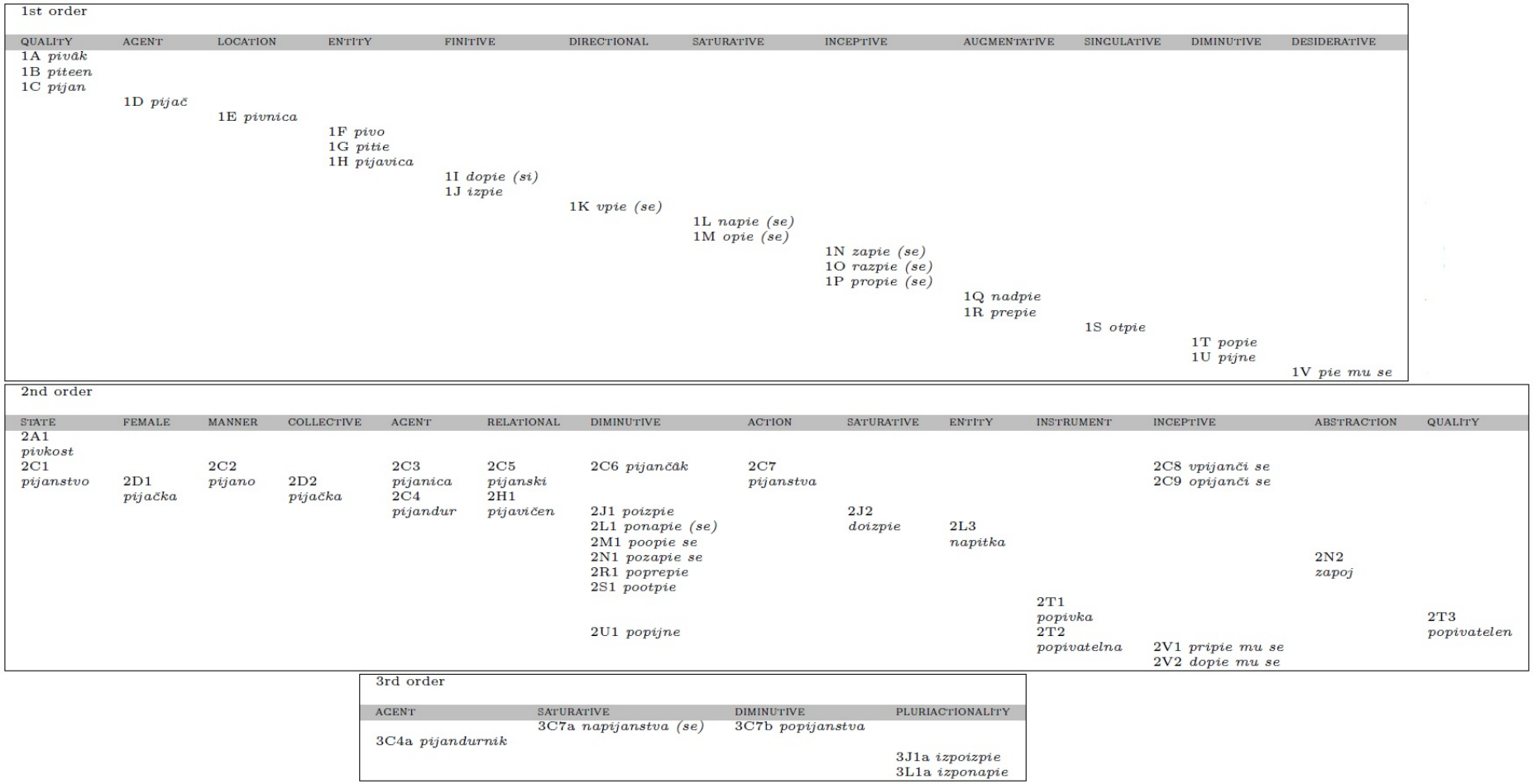
Semantic concepts in affixation

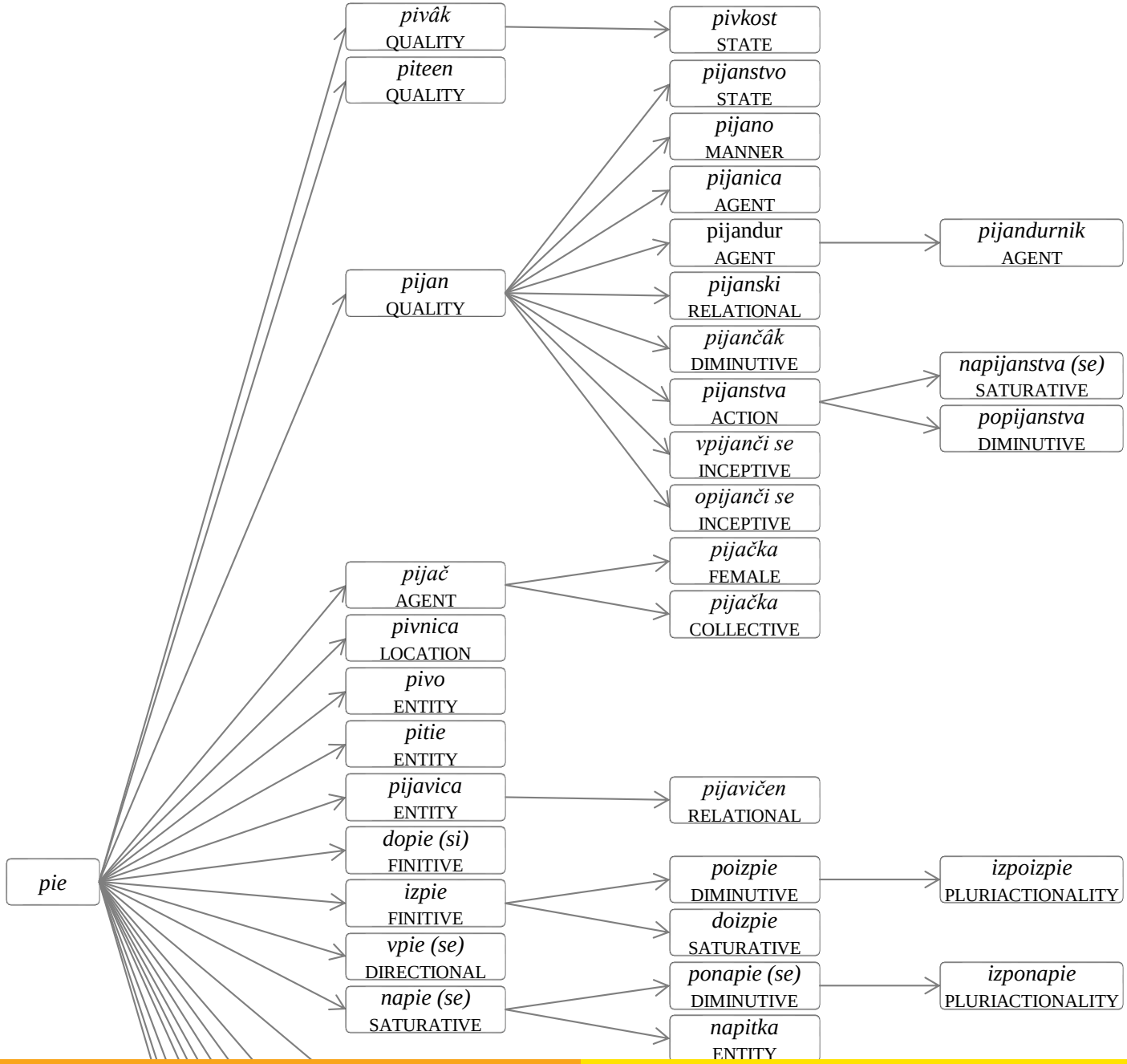
- 50+ comparative semantic categories applicable in cross-linguistic research into affixation (Bagasheva 2017)
  - what meaning is added by attaching the affix to the base word?

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*
...	

The derivational network of the Bulgarian verb *pie* ‘to drink’

(Körtvélyessy et al. 2020:13–16)




pie
pivâk
QUALITY
piteen
QUALITY
pijan
QUALITY
pijač
AGENT
pivnica
LOCATION
pivo
ENTITY
pitie
ENTITY
pijavica
ENTITY
dopie (si)
FINITIVE
izpie
FINITIVE
vpie (se)
DIRECTIONAL
napie (se)
SATURATIVE
pivkost
STATE
pijanstvo
STATE
pijano
MANNER
pijanica
AGENT
pijandur
AGENT
pijandurnik
AGENT
pijanski
RELATIONAL
pijančâk
DIMINUTIVE
pijanstva
ACTION
napijanstva (se)
SATURATIVE
popijanstva
DIMINUTIVE
vpijanči se
INCEPTIVE
opijanči se
INCEPTIVE
pijačka
FEMALE
pijačka
COLLECTIVE
pijavičen
RELATIONAL
poizpie
DIMINUTIVE
izpoizpie
PLURIACTIONALITY
doizpie
SATURATIVE
ponapie (se)
DIMINUTIVE
izponapie
PLURIACTIONALITY
napitka
ENTITY

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.
- Dokulil, M. (1962): *Tvoření slov v češtině 1: Teorie odvozování slov*. Praha: Nakl.ČSAV.
- Dryer, M. S. – Haspelmath, M. (eds., 2013): *The World Atlas of Language Structures Online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info
- 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/filozoficka-fakulta/222-essentials-of-language-typology.html
- Körtvélyessy, L. et al. (2020): *Derivational Networks across Languages*. De Gruyter.
- Müller, P. O. et al. (eds.; 2016): *Word-Formation. An International Handbook of the Languages of Europe*. Volume 4. Berlin: 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. et al. (2012): *Word-Formation in the World's Languages*. Cambridge: CUP.