Measuring the degree of transparency of English derivational suffixes

Verginica Barbu Mititelu[i], Gianina Iordăchioaia[ii], Svetlozara Leseva[iii], Ivelina Stoyanova[iii]

Research Institute for Artificial Intelligence – Romanian Academy[i],
University of Graz, Austria[ii]
Institute for Bulgarian Language – Bulgarian Academy of Sciences[iii]

Workshop 15: Data-based research in word formation
Biennial of Czech Linguistics 2024, Charles University, Prague, Sept. 19, 2024

Interpretation of derived words

- Derivational suffixes (overt or zero) exhibit a high degree of polysemy
- Difficult to model theoretically and problematic for L1/L2 acquisition:

Nominalizing suffixes	**Verbalizing suffixes** (Plag 1999: 125-142)
Event: destruction, building **Result state**: destruction, annihilation **Product**: construction, building, carving **Instrument**: adornment, protection **Location**: parking, residence **Cause**: amusement **Agent**: administration	**Locative** 'put (in)to X': hospitalize, containerize **Ornative** 'provide with X': acidize, nuclearize **Causative** 'make (more) X': randomize, nuclearize **Resultative** 'make into X': peasantize, anglicize **Inchoative** 'become X': aerosolize, grammaticalize **Performative** 'perform X': anthropologize **Similative** 'act like/imitate X': stalinize, marxize

(see Grimshaw 1990, Lieber 2004, 2016, Bierwisch 2009, Melloni 2011, Bauer et al. 2013, Iordachioaia & Melloni 2023, Kawaletz 2023, Valera 2023)

Goal of our study

- Get insights on the semantic transparency of overt and zero derivational suffixes (N → V and V → N) in terms of the morphosemantic relation(s) they establish between base and derivative
- Lexical resources: Princeton WordNet (PWN) (+ directionality of zero derivation from Oxford English Dictionary)
  - N-V sense pairs found in a direct derivational relation
  - N and V semantic classes/primes (n.event, n.artifact, v.change, v.motion etc)
  - Morphosemantic relations (Event, Agent, Instrument, State etc)
- Train and test a machine learning classifier to predict the morphosemantic relation(s) from the suffix and the semantic classes of the base and the derivative

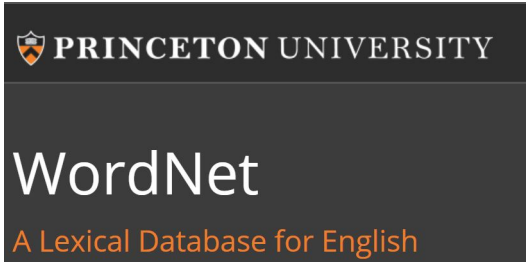
Roadmap

1. Resources
2. Data & methodology
3. Results & discussion
4. Conclusions and future work

Roadmap

1. **Resources**
2. Data & methodology
3. Results & discussion
4. Conclusions and future work

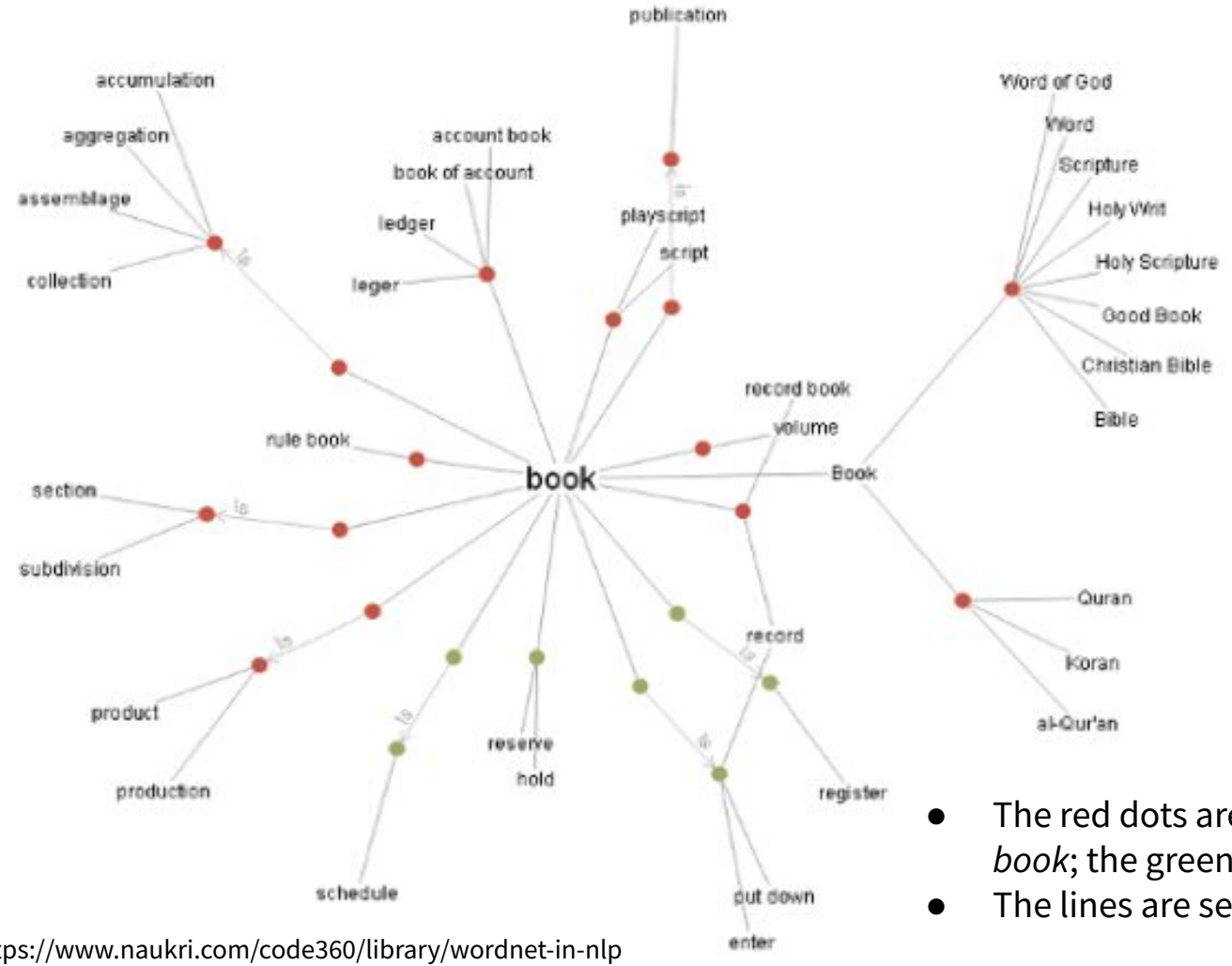
Princeton WordNet



- Princeton WordNet (Fellbaum 1998) is a lexical database of 117,000+ synonym sets (synsets)
- Entries (synsets) correspond to distinct concepts (not lexemes), and contain cognitive synonyms.
- N and V synsets are organized in multiple hierarchies by means of semantic relations: hyponymy, troponymy, meronymy, etc.
- Additionally, N and V synsets are classified into distinct semantic fields by being assigned semantic primes (Miller et al. 1990).

Nouns: 25 semantic primes (~ noun classes), e.g. noun.person, noun.act, noun.artifact, etc.

Verbs: 15 semantic primes (~ verb classes), e.g. verb.motion, verb.change, verb.contact, etc.



- The red dots are synsets for the N *book*; the green dots for the V;
- The lines are semantic relations.

Source: https://www.naukri.com/code360/library/wordnet-in-nlp

25 Semantic primes of WordNet nouns

noun.act: acts or actions
noun.animal: animals
noun.artifact: man-made objects
noun.attribute: attributes of people/objects
noun.body: body parts
noun.cognition: cognitive processes and contents
noun.communication: communicative processes and contents
noun.event: natural events
noun.feeling: feelings and emotions
noun.food: foods and drinks
noun.group: groupings of people or objects
noun.location: spatial position
noun.motive: goals
noun.object: natural objects (not man-made)
noun.person: people
noun.phenomenon: natural phenomena
noun.plant: plants
noun.possession: (transfer of) possession
noun.process: natural processes
noun.quantity: quantities and units of measure
noun.relation: relations b/n people/things/ideas
noun.shape: two and three dimensional shapes
noun.state: stable states of affairs
noun.substance: substances
noun.time: time and temporal relations

15 Semantic primes of WordNet verbs

verb.body: verbs of grooming, dressing and bodily care

verb.change: verbs of size, temperature change, intensifying, etc.

verb.cognition: verbs of thinking, judging, analyzing, doubting

verb.communication: verbs of telling, asking, ordering, singing

verb.competition: verbs of fighting, athletic activities

verb.consumption: verbs of eating and drinking

verb.contact: verbs of touching, hitting, tying, digging

verb.creation: verbs of sewing, baking, painting, performing

verb.emotion: verbs of feeling

verb.motion: verbs of walking, flying, swimming

verb.perception: verbs of seeing, hearing, feeling

verb.possession: verbs of buying, selling, owning

verb.social: verbs of political and social activities and events

verb.stative: verbs of being, having, spatial relations

verb.weather: verbs of raining, snowing, thawing, thundering

Morphosemantic relations in WordNet

- Occur among derivationally related pairs of nouns and verbs.
- Capture the semantics of derivational relations.

net - v	201365945	instrument	**net – n**	103819994	catch with a net; "net a fish"	a trap made of netting to catch fish or ...
net - v	202291548	result	**net – n**	113258362	yield as a net profit; "This sale netted...	the excess of revenues over outlays in a...
net - v	201672168	result	**net – n**	103819595	construct or form a web, as if by weavin...	an open fabric of string or rope or wire...

(Fellbaum, C., Osherson, A., Clark, P.E. (2009). Putting Semantics into WordNet's "Morphosemantic" Links. In: Vetulani, Z., Uszkoreit, H. (eds) Human Language Technology. Challenges of the Information Society. LTC 2007. Lecture Notes in Computer Science(), vol 5603. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04235-5_30)

14 Morphosemantic relations in WordNet

Relation	Description (Koeva et al. 2016)	Example
Agent	a person, a social entity, such as organisations, an animal or a plant that is capable of acting so as to bring about a result	appraise - appraiser
Instrument	either a concrete, usually man-made object, or something abstract, acting under the volition of an Agent	stem - stemmer
Body-part	an inalienable part of the body of an Agent expressed by Ns with the prime N.body (rarely N.animal or N.plant)	flip - flipper
Material	a type of inanimate cause (substances that may bring about a certain effect)	inhibit - inhibitor
Vehicle	means of transportation; as opposed to Instruments their semantic and syntactic behaviour is more similar to Agents	cruise - cruiser
By-means-of	a kind of inanimate cause OR a less causative semantics, rather enabling or facilitating	certify - certificate
Event	processual nominalization involving Ns such as N.act, N.event, N.phenomenon, N.process	approve - approval

14 Morphosemantic relations in WordNet

Relation	Description (Koeva et al. 2016)	Example
State	abstract entities: feelings (N.feeling), cognitive (N.cognition) and other non-dynamic state-of-affairs, such as synsets with the prime N.state	confuse - confusion
Undergoer	entities affected by the situation described and roughly corresponds to the thematic role of Patient/Theme	address - addressee
Result	entities that are produced or come into existence as a result of the situation described by the V	esterify - ester
Property	various attributes and qualities. This relation involves primarily Ns with the prime N.attribute and more rarely N.location	bitter - bitter
Location	a concrete (natural or man-made) or an abstract location where an event takes place	barrel - barrel
Destination	associated with the primes N.person, N.location and N.artifact, corresponding to two distinct interpretations in terms of the thematic role theory – as a Recipient (N.person) or as a Goal (N.artifact, N.location)	classify - class
Uses	a function or purpose of an entity; especially with Vs of putting, the entity is directly involved as the Theme of the V	lipstick - lipstick

Roadmap

1. Resources
2. **Data & methodology**
3. Results & discussion
4. Conclusions and future work

The Core Dataset

- ➔ The Core Dataset used in the experiments consists of **17,634** V – N literal pairs.
- ➔ It was obtained from a larger dataset through applying filtering procedures so as to exclude: spelling (American/British) doublets, derivational pairs with unidentified direction of the relation, non-direct derivations.
- ➔ We collected **11 V → N suffixes** and **5 N → V suffixes**.

The suffixes: V → N

Total: **13,821 pairs**
11 different suffixes
(allomorphs are clustered together)

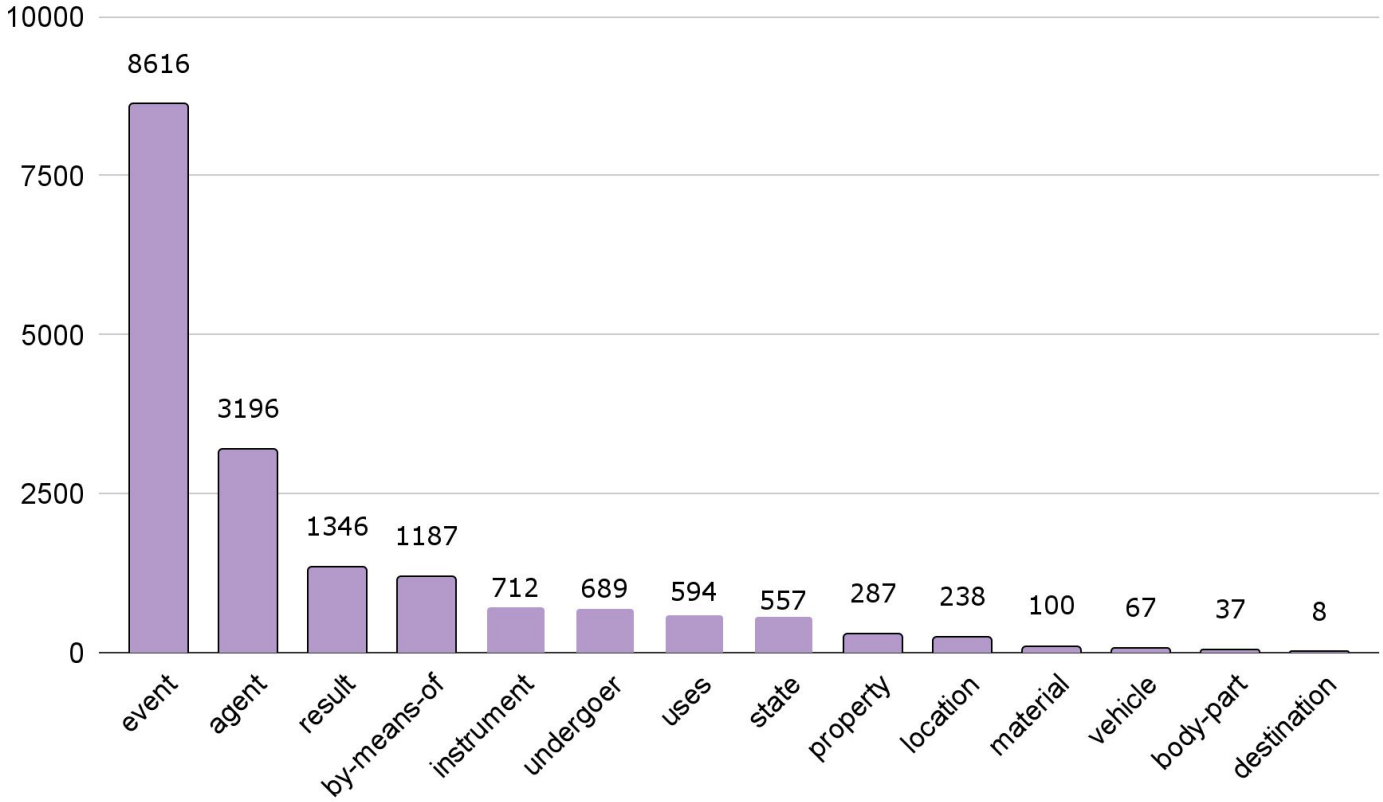
ion	(ation, tion) *absorb - absorption*	4330
er	(or) *roll - roller*	3442
ZeroN	*glide* 'fly in or as if in a glider plane' - *glide* 'the activity of flying a glider'	2366
ing	*play - playing*	1987
ment	*replace - replacement*	699
ance	(ence) *occur - occurrence*	367
ant	(ent) *pollute - pollutant*	159
age	*parent - parentage*	145
al	*dispose - disposal*	135
ure	*press - pressure*	108
ee	*train - trainee*	83

The suffixes: N → V

Total: **3,813 pairs**
5 different suffixes

ZeroV	*fake* 'something that is a counterfeit; not what it seems to be' - *fake* 'make a copy of with the intent to deceive'	2964
ise	*agony - agonise*	463
ate	*acetyl - acetylate*	219
ify	*acetum - acetify*	151
en	*threat - threaten*	16

The morphosemantic relations



Experiments setup: Objective

- To examine the **transparency of the meaning of suffixes** by testing the potential to predict the morphosemantic relation they establish between V and N through basic machine learning algorithms using as features:
  - the semantic class/prime of the verb,
  - the semantic class/prime of the noun, and
  - the suffix.
- We compare the transparency of:
  - Overt vs. zero affixes;
  - Nominal vs. verbal affixes;
  - Individual affixes.

Experiments setup

We perform a set of experiments on the dataset:

- The dataset is presented in **.arff data format**.
- We use **Weka Machine Learning library** to run the experiments.
- We apply two different machine learning algorithms – **OneR and RandomTree**.
- Evaluation is performed using **10-fold cross validation**.

Roadmap

1. Resources
2. Data & methodology
3. **Results & discussion**
4. Conclusions and future work

Experiments results

Test	Setup	ML Classifier	
		OneR	RandomTree
Baseline	Unified affixes (allomorphs clustered together); ZeroN / ZeroV	68.61	73.41
Affixal only	Unified affixes	71.67	74.94
Zero only	ZeroN and ZeroV	66.23	69.47

Experiments results

Test	Setup	ML Classifier	
		OneR	RandomTree
Baseline	Unified affixes; both verbal and nominal; ZeroN / ZeroV	68.61	73.41
Verbal suffixes only	*ZeroV, ise, ate, ify, en*	45.15	53.87
Verbal suffixes – Zero	~~*ZeroV*~~*, ise, ate, ify, en*	34.88	45.2
Nominal suffixes only	*ion, er, ZeroN, ing, ment, ance, ant, age, al, ure, ee*	75.37	77.93
Nominal suffixes – Zero	*ion, er,* ~~*ZeroN*~~*, ing, ment, ance, ant, age, al, ure, ee*	75.44	77.02

Experiments results

Test	Setup	ML Classifier	
		OneR	RandomTree
Baseline	Unified affixes; both verbal and nominal; ZeroN / ZeroV	68.61	73.41
Verbal suffixes only	*ZeroV, ise, ate, ify, en*	45.15	53.87
Verbal suffixes – Zero	~~*ZeroV*~~*, ise, ate, ify, en*	34.88	45.2
Nominal suffixes only	*ion, er, ZeroN, ing, ment, ance, ant, age, al, ure, ee*	75.37	77.93
Nominal suffixes – Zero	*ion, er,* ~~*ZeroN*~~*, ing, ment, ance, ant, age, al, ure, ee*	75.44	77.02

For verbal suffixes, the zero suffix is more transparent.
For only overt suffixes, the method performs worse.

Experiments results

Test	Setup	ML Classifier	
		OneR	RandomTree
Baseline	Unified affixes; both verbal and nominal; ZeroN / ZeroV	68.61	73.41
Verbal suffixes only	*ZeroV, ise, ate, ify, en*	45.15	53.87
Verbal suffixes – Zero	~~*ZeroV*~~*, ise, ate, ify, en*	34.88	45.2
Nominal suffixes only	*ion, er, ZeroN, ing, ment, ance, ant, age, al, ure, ee*	75.37	77.93
Nominal suffixes – Zero	*ion, er,* ~~*ZeroN*~~*, ing, ment, ance, ant, age, al, ure, ee*	75.44	77.02

For nominal suffixes, there is no sensible difference when including / excluding the zero suffix.

Experiments results

Verbal Suffix	ML Classifier	
	OneR	RandomTree
ise	44.34	48.43
ate	46.34	52.44
ify	53.10	61.38
en	38.46	38.46
ZeroV	51.86	56.61

Nominal Suffix	ML Classifier	
	OneR	RandomTree
ion	**81.86**	**81.65**
er	**91.19**	**91.10**
ing	**81.96**	**83.31**
ment	77.63	76.90
ance	74.58	74.58
ant	69.75	73.11
age	72.84	67.90
al	**87.88**	**86.36**
ure	*51.09*	*55.43*
ee	*60.24*	*60.24*
ZeroN	78.57	81.36

Error analysis

Most frequent errors

Correct relation	Assigned relation	Number of errors
by-means-of	event	346
result	event	314
undergoer	event	162
by-means-of	instrument	155
state	event	104
uses	event	98
by-means-of	result	98
undergoer	result	86
event	result	79
event	state	74
uses	instrument	74
uses	result	66
property	event	63

Error analysis: examples

(1) **by-means-of**: -ment, argue (verb.communication) 'give evidence of' – argument (noun.communication) 'a fact or assertion offered as evidence that something is true'
Wrong prediction: **event**

(2) **by-means-of**: -ise, allegorise (verb.change) 'make into an allegory' – allegory (noun.communication) 'a visible symbol representing an abstract idea'
Wrong prediction: **result**

(3) **by-means-of**: -ing, bind (verb.contact) 'provide with a binding' – binding (noun.artifact) 'the protective covering on the front, back, and spine of a book'
Wrong prediction: **uses**

(4) **by-means-of**: ZeroN, clinch (verb.contact) 'secure or fasten by flattening the ends of nails or bolts' – clinch (noun.artifact) 'a small slip noose made with seizing'
Wrong prediction: **instrument**

Error analysis: examples

(5) **result**: ZeroV, bundle (verb.contact) 'gather or cause to gather into a cluster' – bundle (noun.artifact) 'a package of several things tied together for carrying or storing'
Wrong prediction: **instrument**

(6) **result**: -ion, conclude (verb.stative) 'come to a close' – conclusion (noun.communication) 'the last section of a communication'
Wrong prediction: **event**

(7) **result**: -ing, cross (verb.motion) 'meet at a point' – crossing (noun.artifact) 'a junction where one street or road crosses another'
Wrong prediction: **location**

(8) **result**: -ify, classify (verb.cognition) 'arrange or order by classes or categories' – class (noun.group) 'a league ranked by quality'
Wrong prediction: **undergoer**

Roadmap

1. Resources
2. Data & methodology
3. Results & discussion
4. **Conclusions and future work**

Conclusion and future work

- The dataset shows rich polysemy and the overall prediction is not very high: 68-73%;
  - Overt suffixes together show slightly better prediction: 71-74%;
  - Zero suffixes alone (ZeroN + ZeroV) yield slightly worse results: 66-69%;
- Verbalizing suffixes are much less transparent (44-53%), and ZeroV seems to be more transparent than the overt ones; in the absence of ZeroV suffix: only 34-45%;
- Nominalizing suffixes are the most transparent (75-78%), and ZeroN behaves similarly to the overt ones; in its absence: 75-77%;
  - This confirms previous observations that ZeroN is semantically more transparent/compositional than ZeroV (Kisselew et al. 2016, Barbu Mititelu et al. 2023);
  - ZeroN seems to be similar to the overt nominalizing suffixes (Iordachioaia & Melloni 2023; contra Grimshaw 1990, Borer 2013).

Future work

- Check if the number of entries per suffix has an impact (fewer entries, worse prediction) => try to balance the numbers.
- Further analyze the individual suffixes with lower predictability as to whether they are indeed more polysemous and get an insight into the meanings of different suffixes.
- Use machine learning to collect information on which relations often co-occur to possibly cluster them together for the future (e.g. By-means-of and Uses).

References

- Barbu Mititelu, Verginica, Gianina Iordachioaia, Svetlozara Leseva & Ivelina Stoxanova. 2023. The meaning of zero nouns and zero verbs. In S. Kotowski & I. Plag (eds.), *The Semantics of Derivational Morphology*. De Gruyter, Berlin, 63-102.
- Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013. Oxford reference guide to English morphology. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780198747062.001.0001.
- Bierwisch, Manfred. 2009. Nominalization – lexical and syntactic aspects. In Anastasia Giannakidou & Monika Rathert (eds.), Quantification, definiteness, and nominalization, 281–320. Oxford: Oxford University Press.
- Borer, Hagit. 2013. Taking form. Oxford: Oxford University Press.
- Fellbaum, Christiane (ed.). 1998. WordNet: An electronic lexical database. Cambridge, MA: MIT Press.
- Fellbaum, C., Osherson, A., Clark, P.E. (2009). Putting Semantics into WordNet's "Morphosemantic" Links. In: Vetulani, Z., Uszkoreit, H. (eds) Human Language Technology. Challenges of the Information Society. LTC 2007. Lecture Notes in Computer Science(), vol 5603. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04235-5_30
- Grimshaw, Jane. 1990. Argument Structure. Cambridge, MA: MIT Press.
- Iordachioaia, Gianina & Chiara Melloni. 2023. The zero suffix in English and Italian deverbal nouns. Zeitschrift für Sprachwissenschaft 42(1): 109–132. DOI: https://doi.org/10.1515/zfs-2022-2014.
- Kawaletz, Lea. 2023. The semantics of English -ment nominalizations (Empirically Oriented Theoretical Morphology and Syntax 12). Berlin: Language Science Press.

References

- Kisselew, Max, Laura Rimell, Alexis Palmer & Sebastian Padó. 2016. Predicting the direction of derivation in English conversion. In Proceedings of the ACL SIG-MORPHON workshop, 93–98. Berlin: Association for Computational Linguistics. https://aclanthology.org/W16-2015.
- Koeva, Svetla, Svetlozara Leseva, Ivelina Stoyanova, Tsvetana Dimitrova & Maria Todorova. 2016. Automatic prediction of morphosemantic relations. In Proceedings of the 8th global WordNet conference (gwc), 169–177. Bucharest, Romania: Global Wordnet Association. https://aclanthology.org/2016.gwc-1.26.
- Lieber, Rochelle. 2004. Morphology and lexical semantics. Cambridge: Cambridge University Press.
- Lieber, Rochelle. 2016. English nouns. The ecology of nominalization. Cambridge: Cambridge University Press.
- Melloni, Chiara. 2011. Event and result nominals. A morpho-semantic approach. Bern: Peter Lang.
- Miller, George A., Richard Beckwith, Christiane Fellbaum, Derek Gross & Katherine Miller. 1990. Introduction to WordNet: An online lexical database. International Journal of Lexicography 3(4). 235–244.
- Valera, Salvador. 2023. The semantics of noun-to-verb zero-derivation in English and Spanish. Zeitschrift für Sprachwissenschaft 42(1): 153-180. DOI: https://doi.org/10.1515/zfs-2022-2016

Thank you!

vergi@racai.ro

gianina.iordachioaia@uni-graz.at

zarka@dcl.bas.bg

iva@dcl.bas.bg