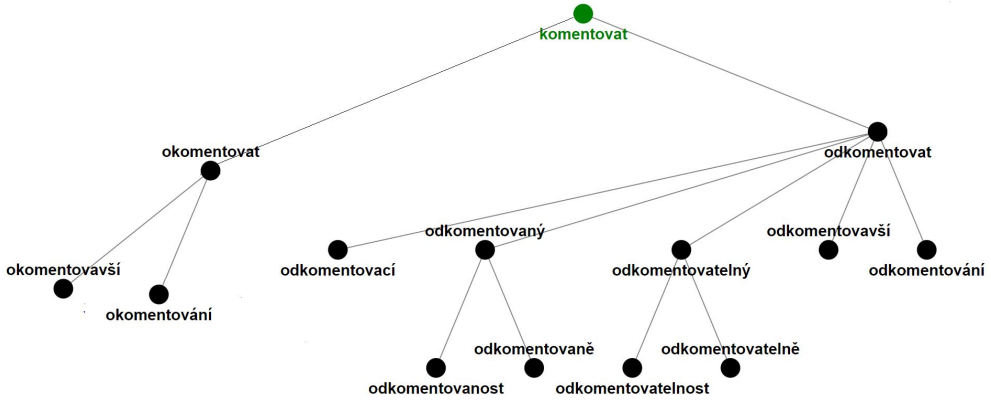
Understanding borrowing through derivational morphology

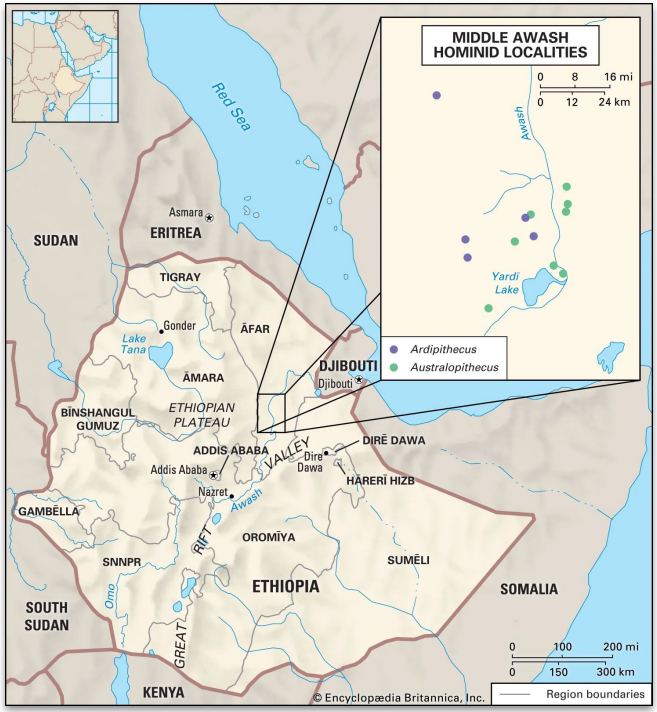
A case study of Czech verbs

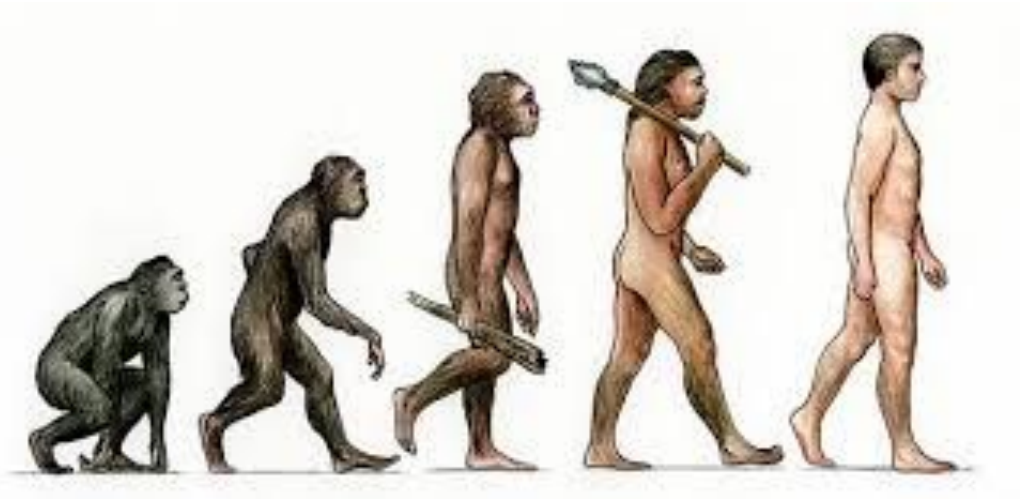


Abishek Stephen and Zdeněk Žabokrtský

Charles University
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics

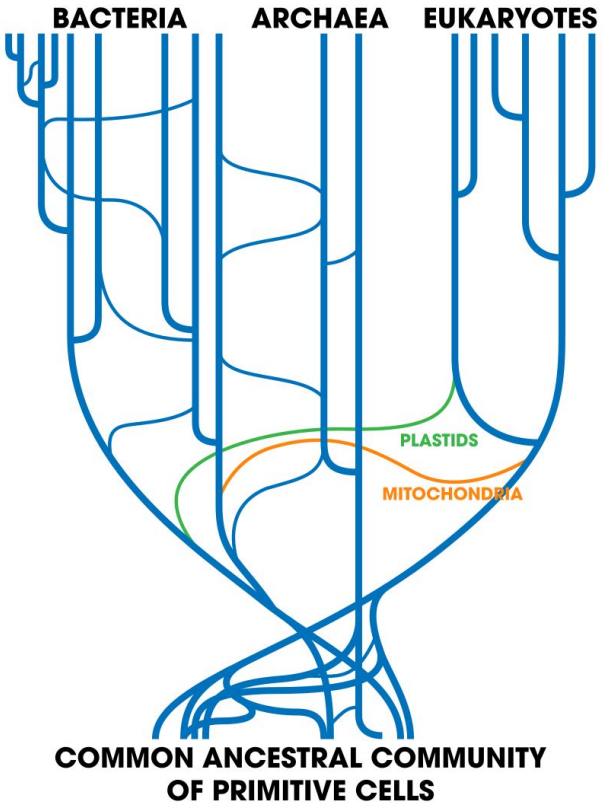
Aftermath of The Great Rift Valley



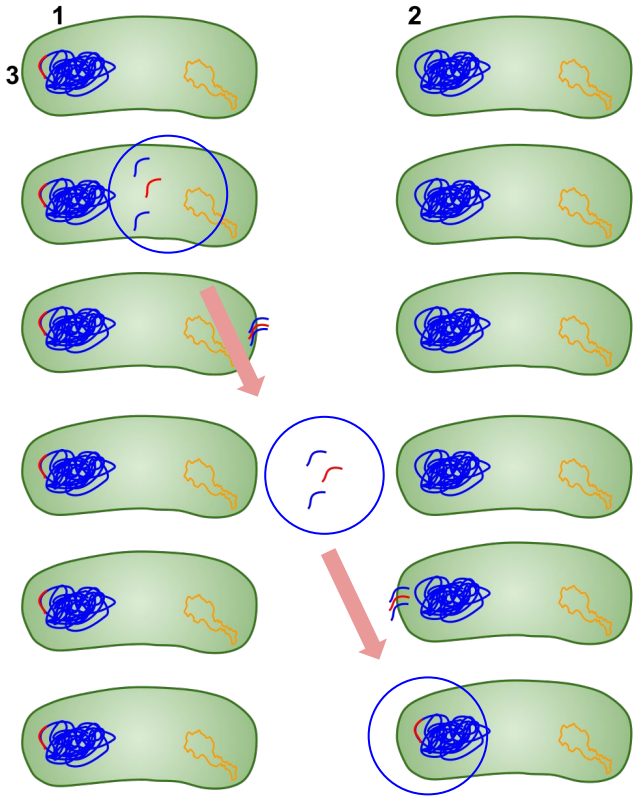


https://www.britannica.com/science/human-evolution

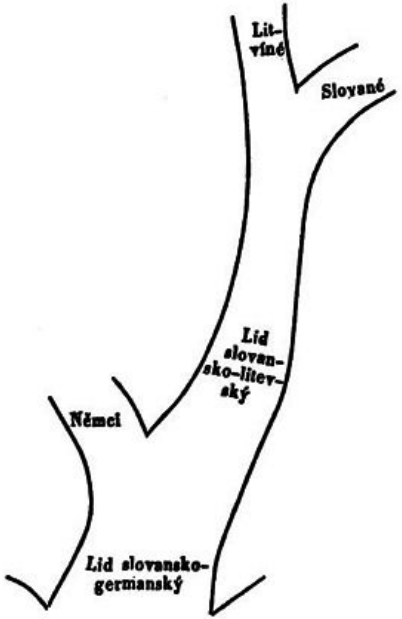
Lateral gene transfer



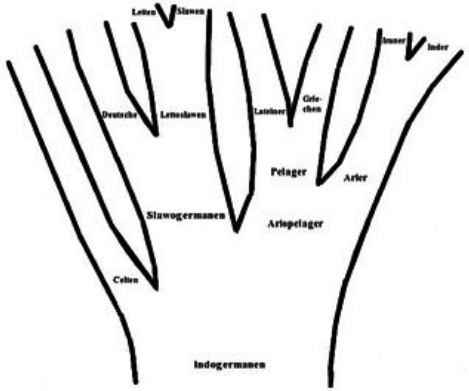
https://en.wikipedia.org/wiki/Horizontal_gene_transfer



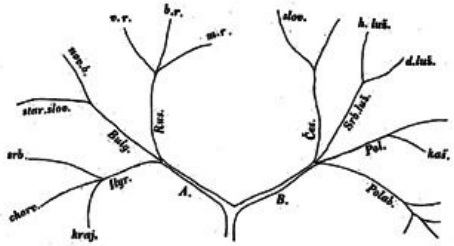
The Tree and The Wave



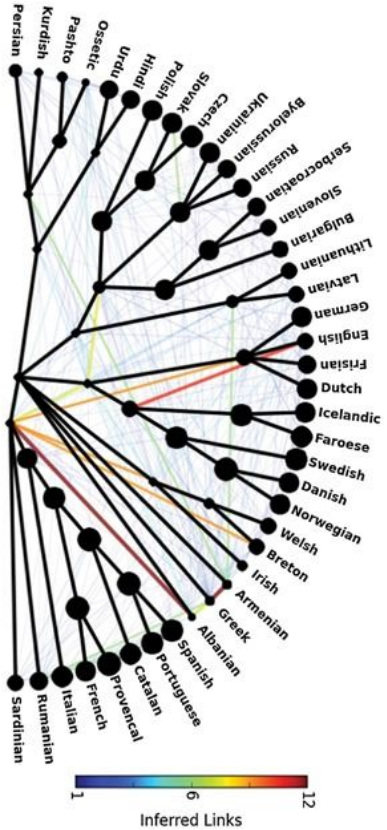
A) August Schleicher, 1853 [21]



B) August Schleicher 1853 [22]

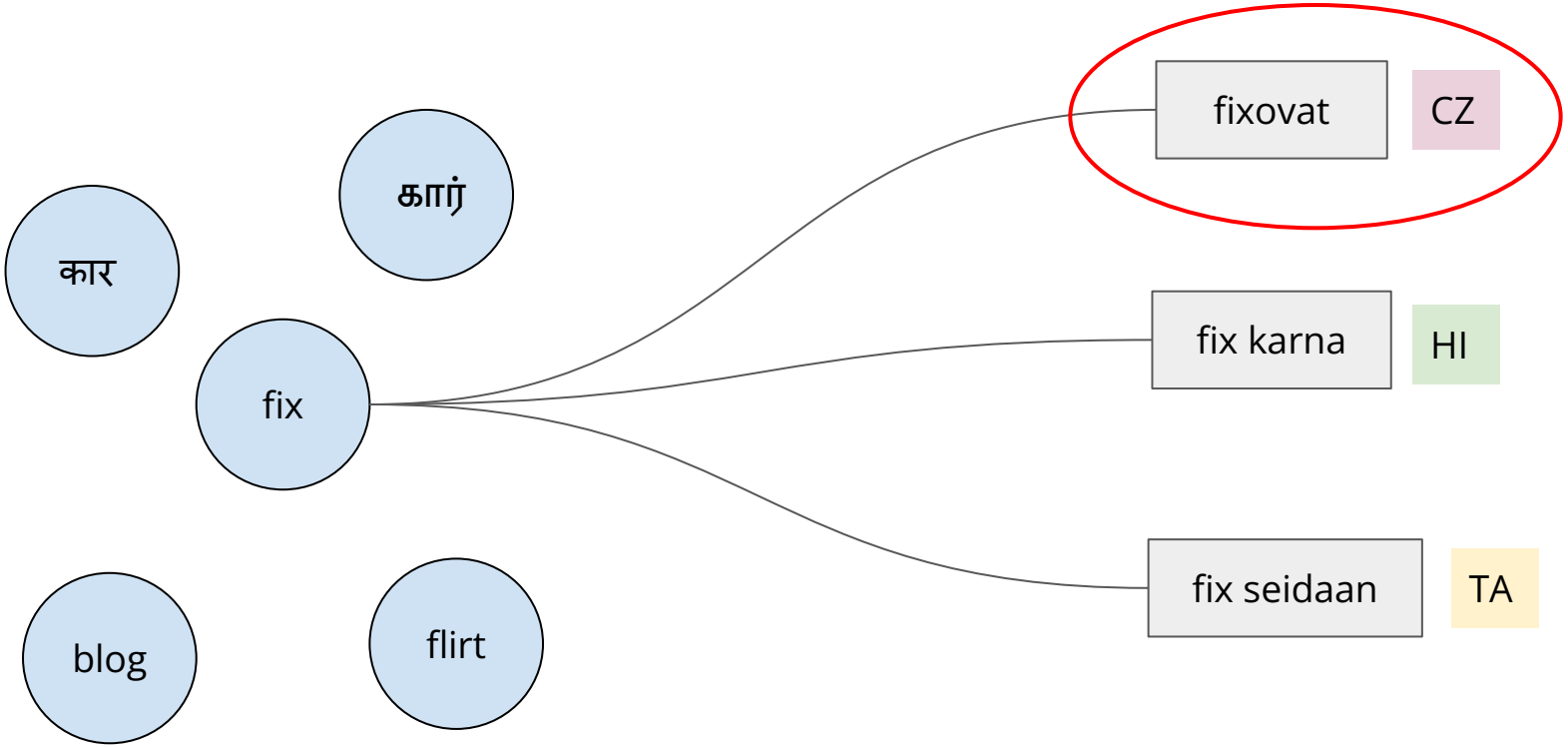


C) František Ladislav Čelakovský 1853 [24]

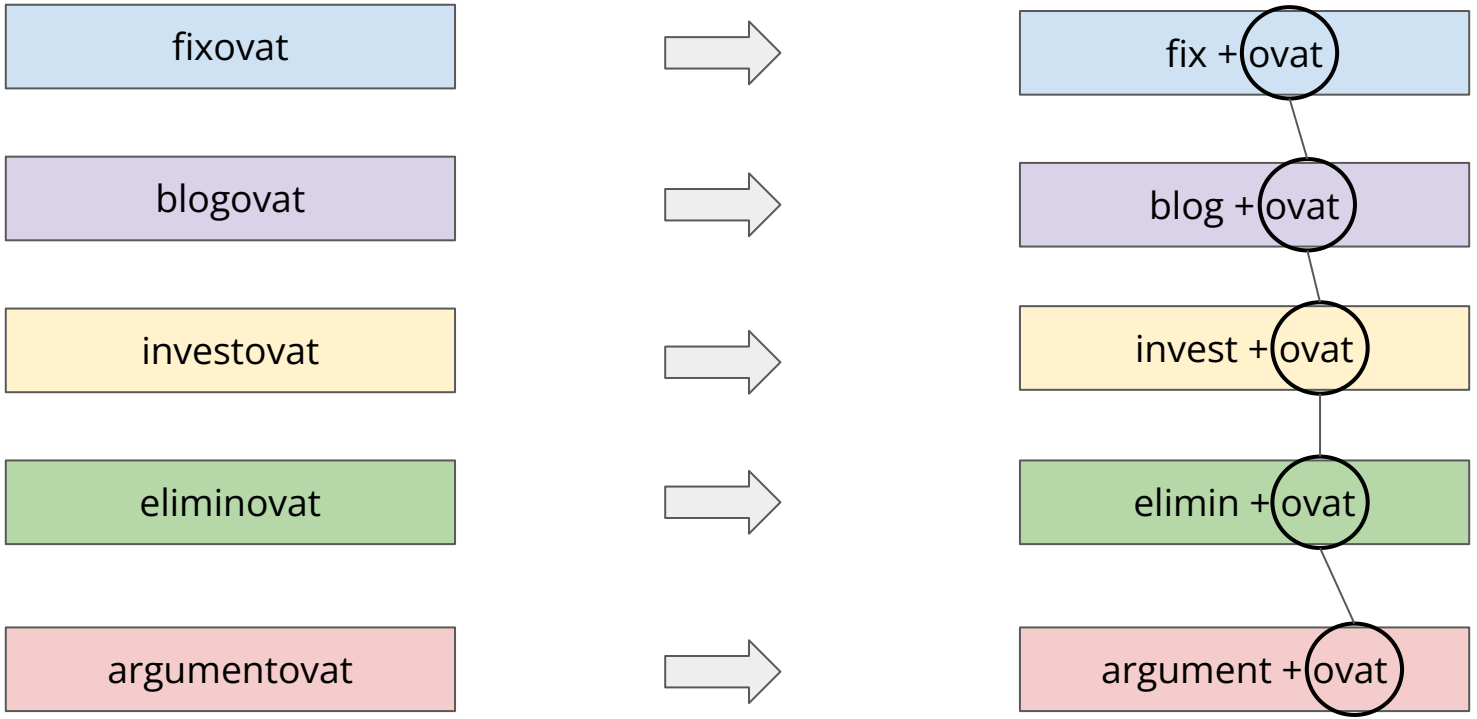


List, Johann-Mattis & Nelson-Sathi, Shijulal & Geisler, Hans & Martin, William. (2014). Networks of lexical borrowing and lateral gene transfer in language and genome evolution. BioEssays : news and reviews in molecular, cellular and developmental biology. 36. 10.1002/bies.201300096.

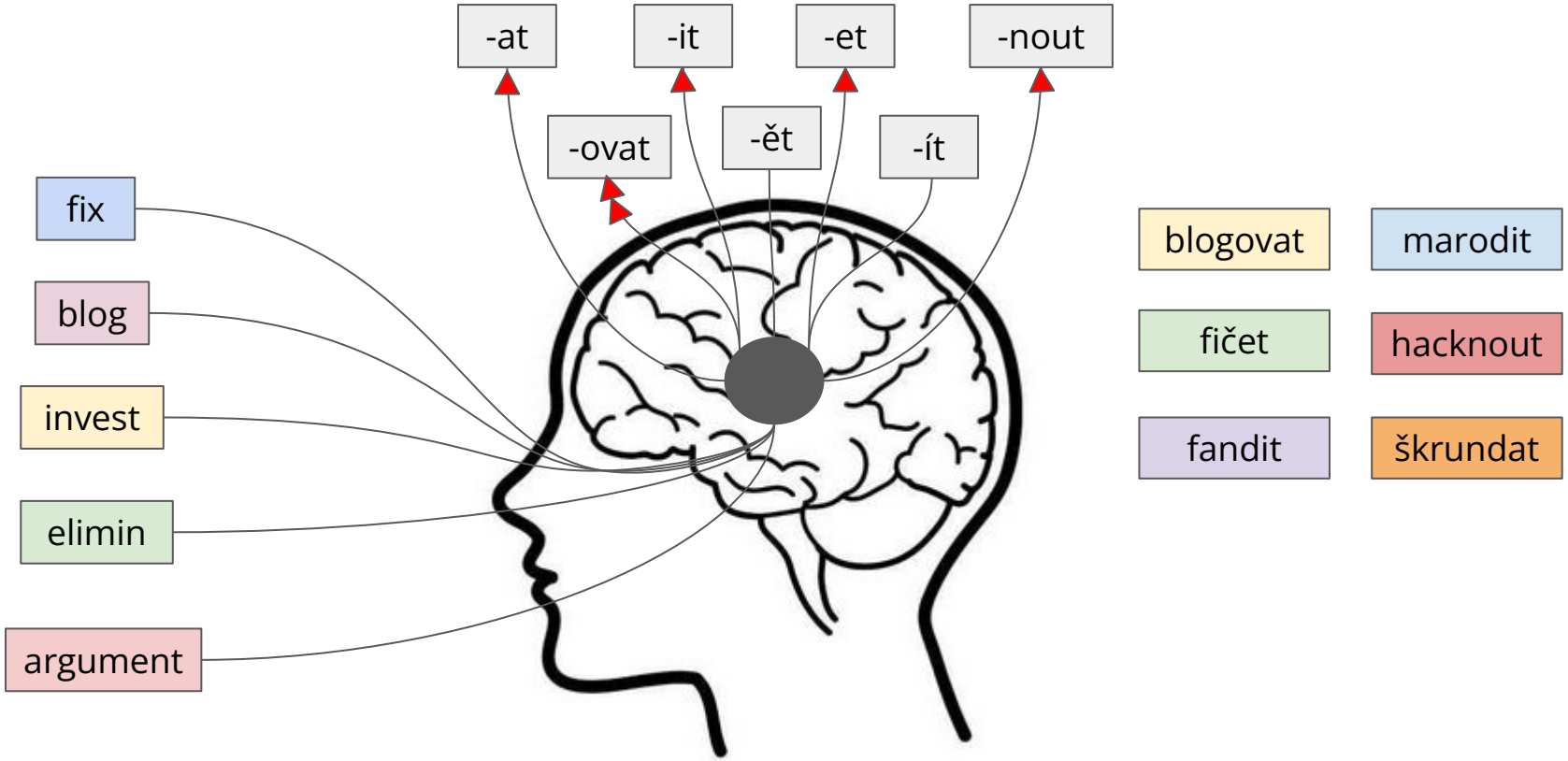
The Morph



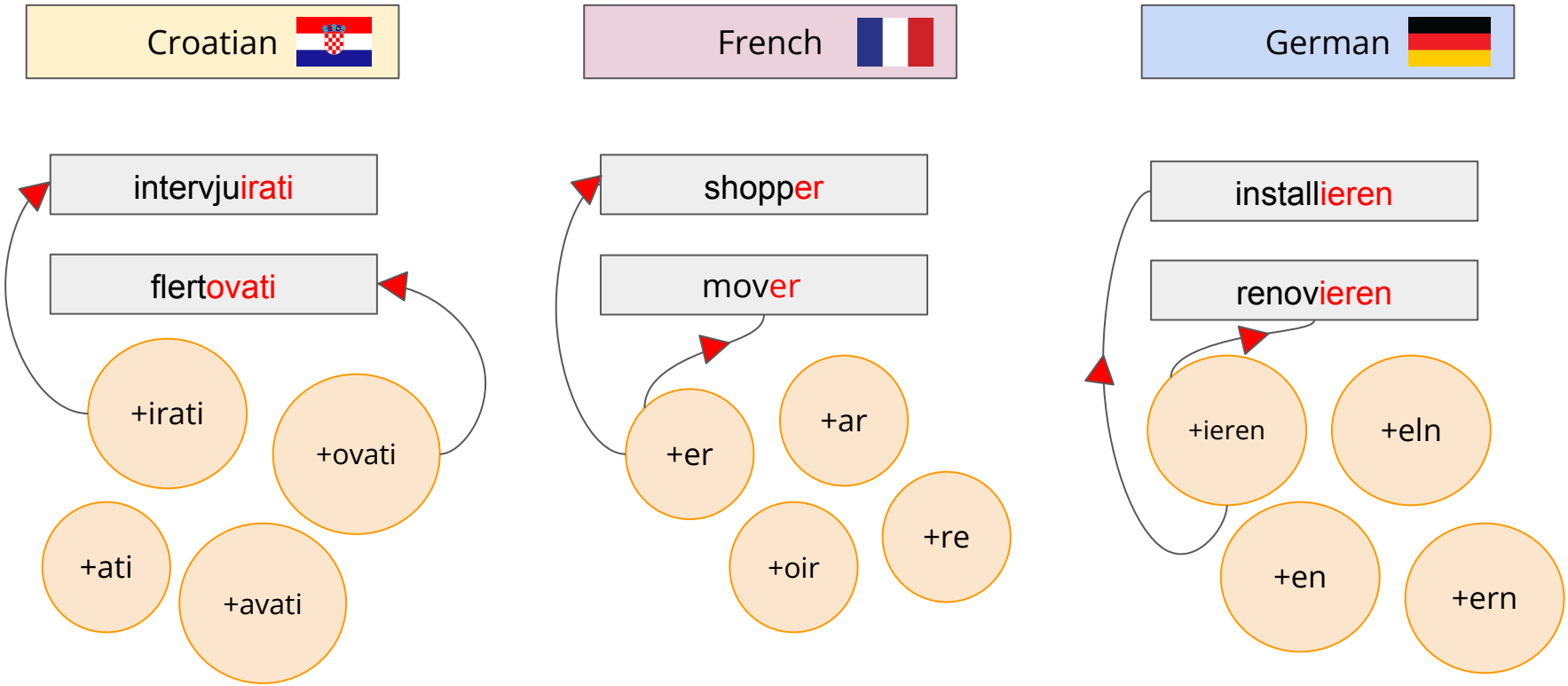
Verbs with borrowed roots



THE NEURAL Network

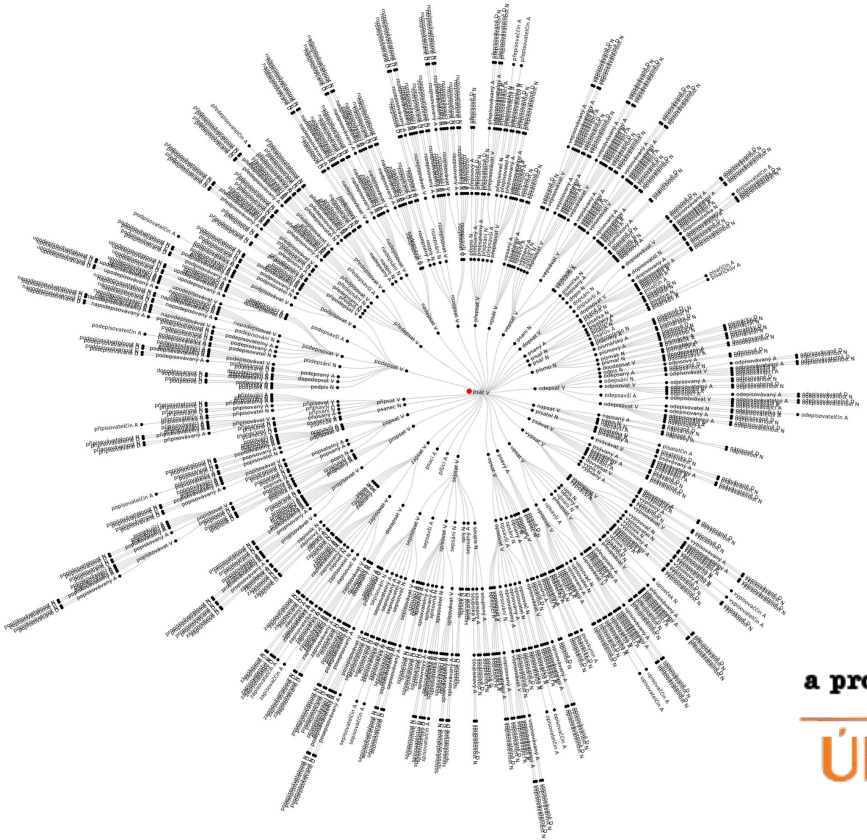


Integration Strategies



Approach

DeriNet



DeriNet 2.1, contains over 1 million lexemes

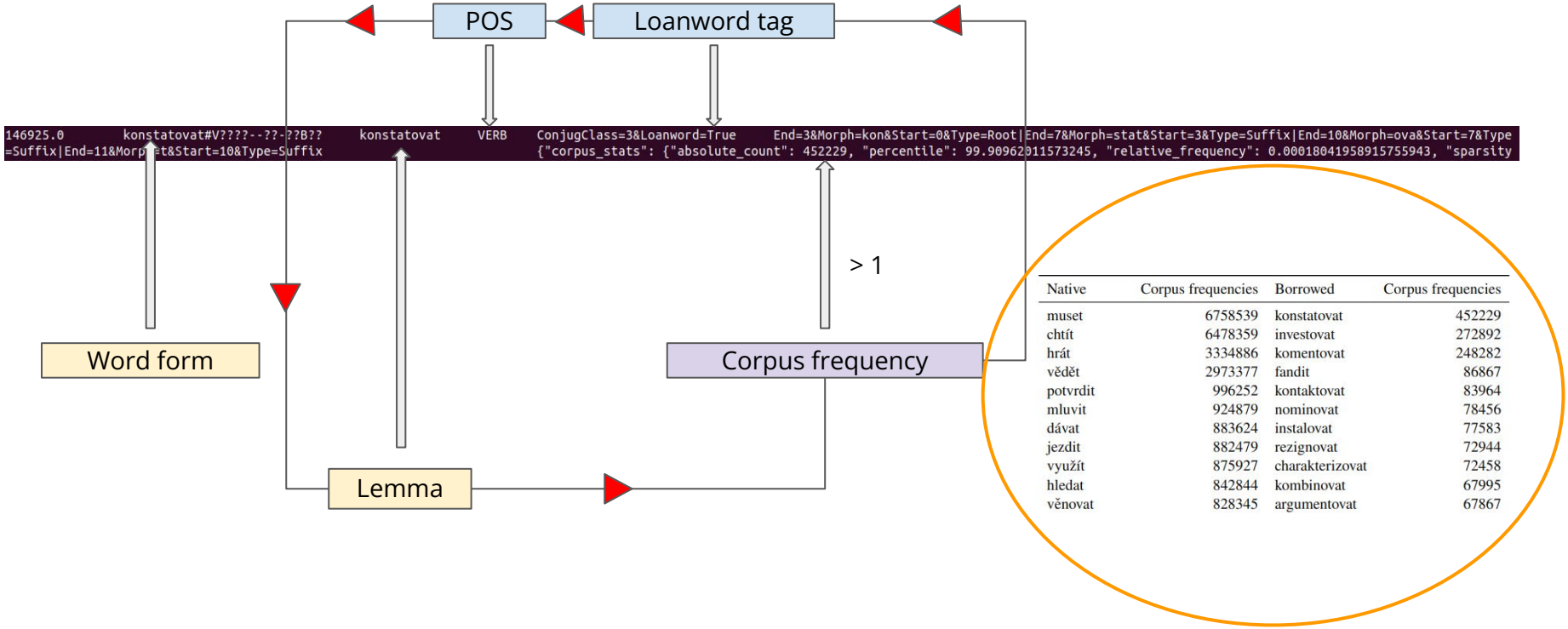
DeriNet is a lexical network which models word-formation relations in the lexicon of Czech.

autogenerated full morphological segmentations of all lemmas

a project by ÚFAL

LINDAT CLARIAH-CZ

DeriNet



Native	Corpus frequencies	Borrowed	Corpus frequencies
muset	6758539	konstatovat	452229
chtít	6478359	investovat	272892
hrát	3334886	komentovat	248282
vědět	2973377	fandit	86867
potvrdit	996252	kontaktovat	83964
mluvit	924879	nominovat	78456
dávat	883624	instalovat	77583
jezdit	882479	rezignovat	72944
využít	875927	charakterizovat	72458
hledat	842844	kombinovat	67995
věnovat	828345	argumentovat	67867

Evaluation and Results

Frequencies

Type	Total verbs	Corpus attested
Native root	42930	19854
Borrowed root	13378	3972

Table 1: Frequencies of verbs in DeriNet

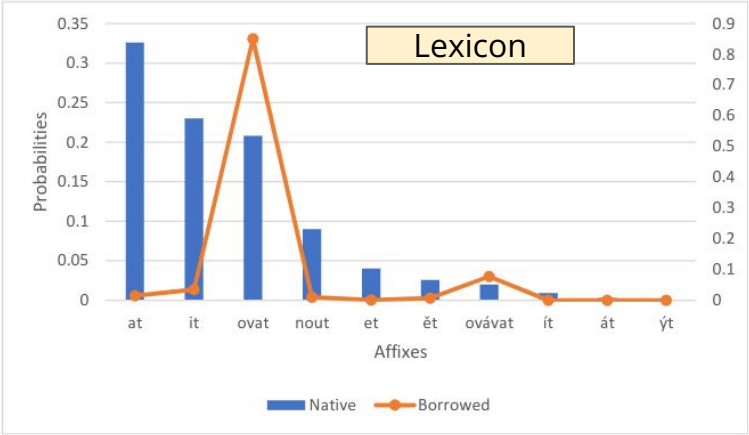
Affixes	Native root	Tokens	Borrowed root	Tokens
-at	6481	19225633	62	10724
-it	4492	25400609	141	130751
-ovat	4132	10791009	3377	3749598
-nout	1780	6024365	41	22712
-ovávat	413	128614	307	3307
-et	778	12355416	3	347
-ět	519	6481870	28	860
-át	67	5676169	0	0
-ít	195	11439639	0	0
-ýt	28	228658	0	0

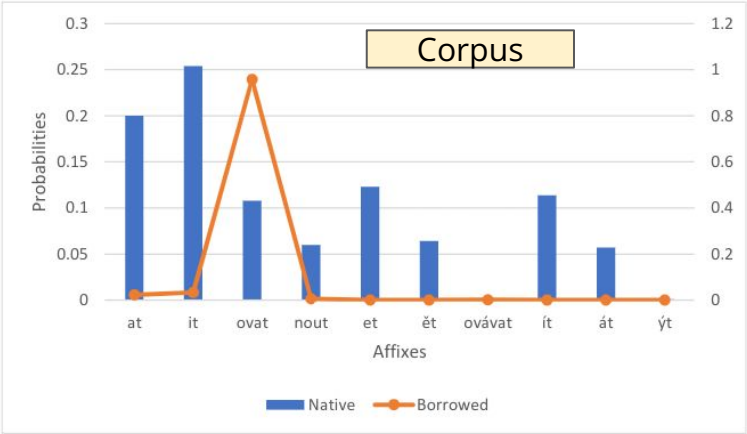
Table 2: Frequencies of derivational affixes

Entropy

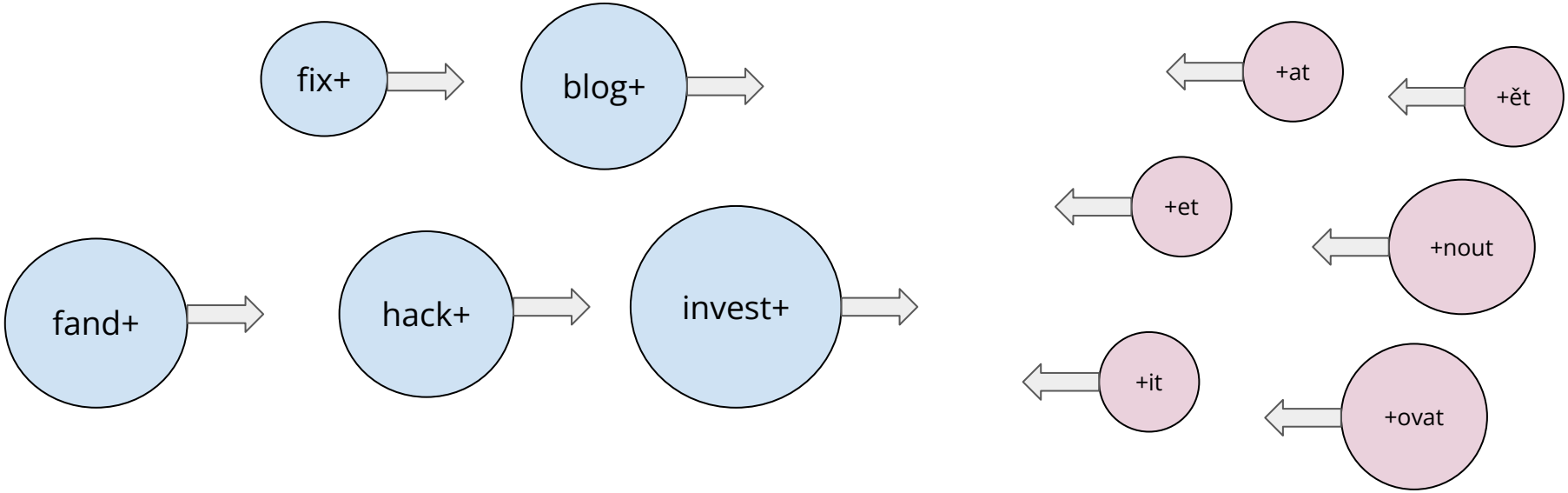
Origin of root	Affix	Lexicon frequencies		Corpus frequencies	
		P(Suffix|Origin)	Entropy	P(Suffix|Origin)	Entropy
Native	*-at*	0.326	2.330	0.200	2.803
	-it	0.230		0.254	
	-ovat	0.208		0.108	
	-nout	0.090		0.060	
	-ovávat	0.020		0.001	
	-et	0.040		0.123	
	-ět	0.026		0.064	
	-át	0.003		0.057	
	-ít	0.009		0.114	
	-ýt	0.001		0.002	
Borrowed	*-at*	0.015	0.873	0.023	0.401
	-it	0.035		0.033	
	-ovat	0.850		0.958	
	-nout	0.010		0.006	
	-ovávat	0.078		0.001	
	-et	0.001		0.000	
	-ět	0.007		0.000	
	-át	0.000		0.000	
	-ít	0.000		0.000	
	-ýt	0.000		0.000	

Table 3: Probabilities and entropies of derivational affixes





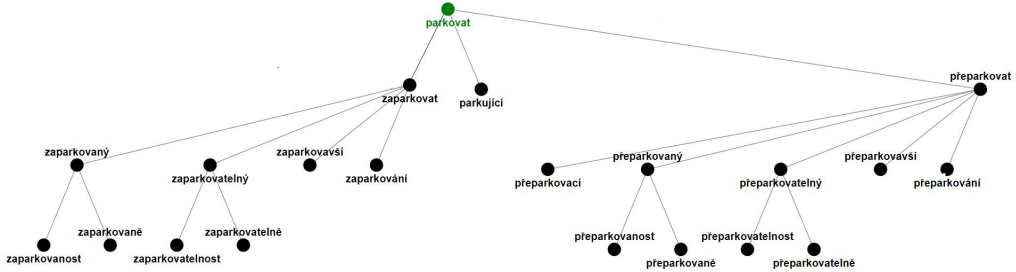
Dice coefficient

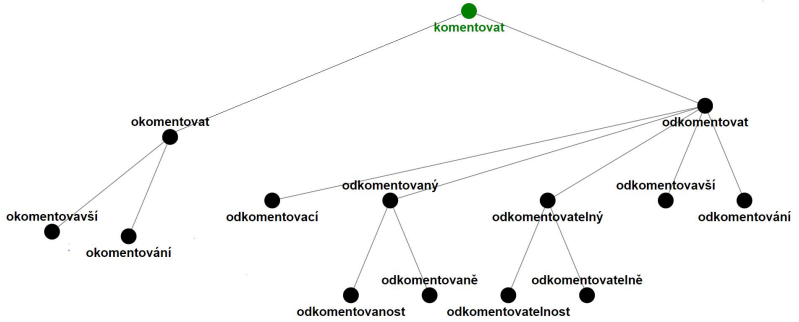


	-at	*-it*	*-ovat*	*-nout*	*-ovávat*	*-et*	*-ět*	*-át*	*-ít*	*-ýt*
Native root	0.322	0.404	0.188	0.114	0.002	0.220	0.122	0.107	0.205	0.005
Borrowed root	0.001	0.009	0.406	0.005	0.001	0	0	0	0	0

Table 4: Dice coefficient based on corpus frequencies

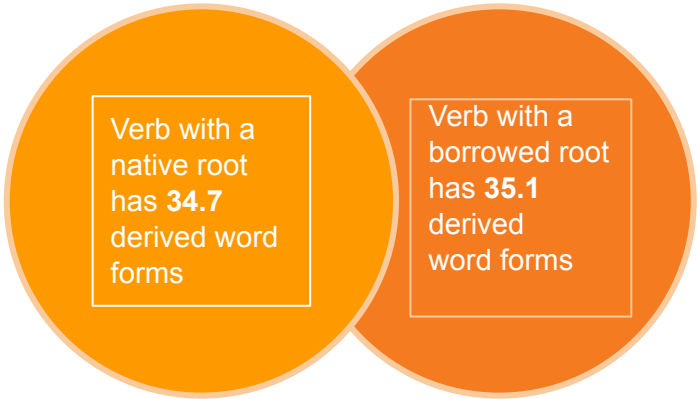
Children






```
if word.is_loanword != "None" and word.absolute_count > 1:
   if word.is_loanword == "True":
           count_loan++
           count_children_loan += word.get_all_children()
           if word.pos == "VERB":
               count_verb_loan++
               count_children_verb_loan += word.get_all_children()
   else:
           count_native++
           count_children_native += word.get_all_children()
           if word.pos == "VERB":
               count_verb_native++
               count_children_verb_native += word.get_all_children()

avg_children_verb_loan: float = count_children_verb_loan / count_verb_loan
avg_children_verb_native: float = count_children_verb_native / count_verb_native
```



Prefixes

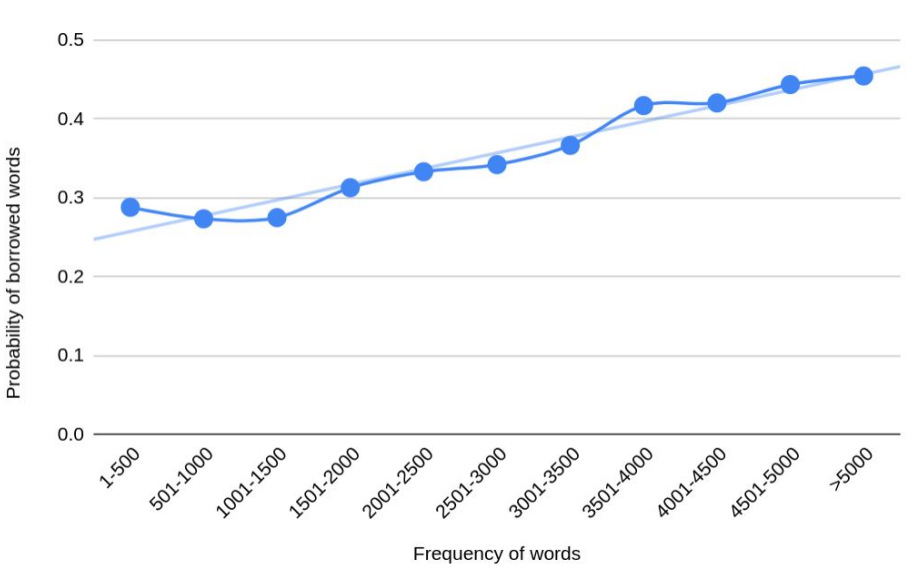
	do-	*roz-*	*o-*	*po-*	*pod-*	*od-*	*u-*	*v-*	*vy-*	*z-*	*za-*	*pře-*	*před-*
Native root	616	649	568	661	134	506	652	623	1228	1263	1226	454	117
Borrowed root	108	38	70	44	7	104	23	47	187	338	214	140	14

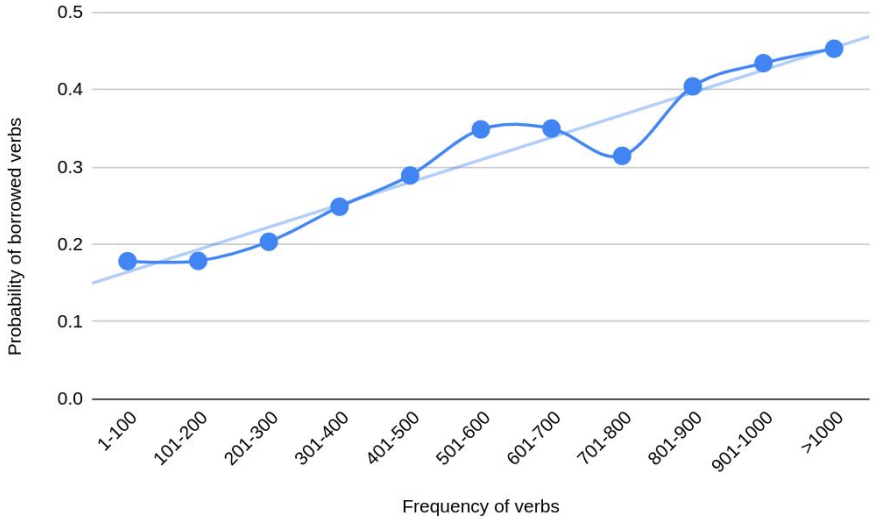
Table 5: Lexicon frequencies of the prefixed verbs with native and borrowed roots

	do-	*roz-*	*o-*	*po-*	*pod-*	*od-*	*u-*	*v-*	*vy-*	*z-*	*za-*	*pře-*	*před-*
Native root	112	123	141	129	30	126	129	106	306	270	293	119	26
Borrowed root	69	23	51	37	4	66	25	23	142	291	160	91	11

Table 6: Lexicon frequencies of the prefixed verbs with native and borrowed roots with affix *-ovat*

Frequency of loanverbs





Central tendency

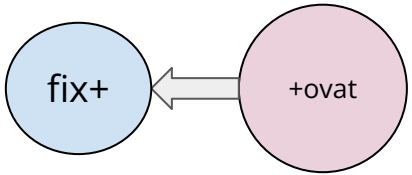
Affixes	Native root	Tokens	Borrowed root	Tokens
-at	6481	19225633	62	10724
-it	4492	25400609	141	130751
-ovat	4132	10791009	3377	3749598
-nout	1780	6024365	41	22712
-ovávat	413	128614	307	3307
-et	778	12355416	3	347
-ět	519	6481870	28	860
-át	67	5676169	0	0
-ít	195	11439639	0	0
-ýt	28	228658	0	0

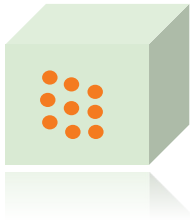
Table 2: Frequencies of derivational affixes

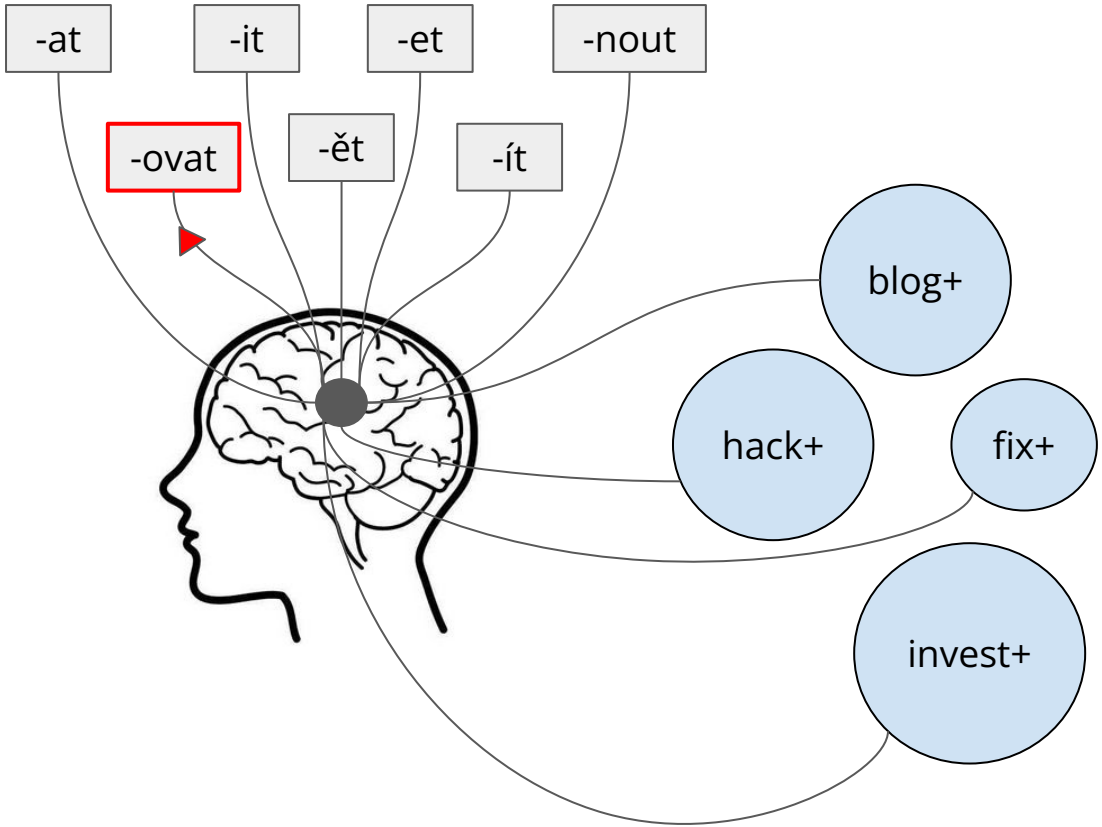
8.6 million

Summary

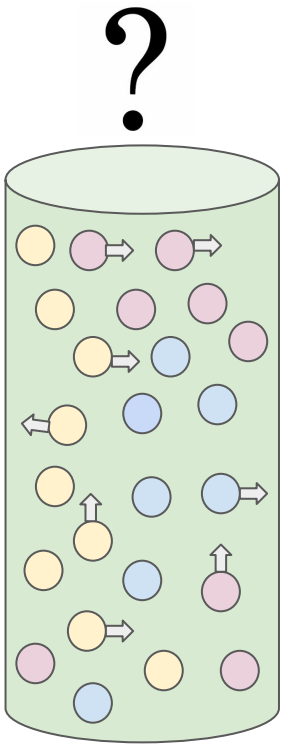
Summary







Summary



To keep the morphological system out of chaos that can be caused due to the incoming borrowed words, the *central* derivational process is extended towards handling the morphology of loanwords in the presence of multiple verb conjugation classes.

These underlying mechanisms act as a positive pressure for accepting borrowings and thus contribute to the evolution of language in terms of its vocabulary range and morphological specializations to name a few among the various other modifications.

Thus, the verb integration strategies or in this study the derivational processes led by the conjugation classes play a vital role in language change and evolution over time.

References

Anthony T. Arlotto. 1972. Introduction to Historical Linguistics. Houghton Mifflin, Boston.

Alan D. Baddeley, Neil Thomson, and Mary Buchanan. 1975. Word length and the structure of short-term memory. Journal of Verbal Learning and Verbal Behavior 14(6):575–589.

Ondrej Blaha. 2022. Dynamics of Conjugation Pattern “Kupuje” in Contemporary Czech (On Material of Journalistic Texts,1990-2019). Bohemica Olomucensia 14(2):40–54.

Lyle Campbell. 2020. Historical linguistics : an introduction / Lyle Campbell. Edinburgh University Press, Edinburgh, fourth edition.

Rudolf Filipovic. 1980. Transmorphemization: Substitution on the Morphological Level Reinterpreted, Revue publiee par les Sections romane, italienne et anglaise de la Faculte des Lettres de l’Universite de Zagreb, volume 25 (-), pages 1–8.

Rudolf Filipovic. 1981. Morphological Categories in Linguistic Borrowing, Revue publi
ee par les Sections romane, italienne et anglaise de la Faculte des Lettres de l’Universite de Zagreb, volume 26 (-), pages 197–207.

Rebecca J. Hall, Fiona J. Whelan, James O. McInerney, Yaqing Ou, and Maria Rosa Domingo-Sananes. 2020. Horizontal Gene Transfer as a Source of Conflict and Cooperation in Prokaryotes.Frontiers in Microbiology 11.

Martin Haspelmath. 2009. II. Lexical borrowing: Concepts and issues, De Gruyter Mouton, Berlin, New York, pages 35–54.

Olga Kolesnikova. 2016. Survey of Word Co-occurrence Measures for Collocation Detection. Computacion y Sistemas 20:327–344.

References

Johann-Mattis List, Shijulal Nelson-Sathi, Hans Geisler, and William Martin. 2014. Networks of lexical borrowing and lateral gene transfer in language and genome evolution. BioEssays 36(2):141–150.

Yaron Matras and Evangelia Adamou. 2020. Borrowing, Routledge, pages 237–251.

Yaron Matras and Janette Sakel. 2007. Investigating the mechanisms of pattern replication in language convergence. Studies in Language 31(4):829–865.

Padraic Monaghan and Sean G. Roberts. 2019. Cognitive influences in language evolution: Psycholinguistic predictors of loan word borrowing. Cognition 186:147–158.

Salikoko S. Mufwene. 2001. Language contact, evolution, and death: how ecology rolls the dice, Cambridge University Press, page 145–166. Cambridge Approaches to Language Contact.

Shijulal Nelson-Sathi and Johann-Mattis List. 2011. Networks uncover hidden lexical borrowing in Indo-European language evolution. Proceedings of the Royal Society pages 1794–1803.

Mark Pagel, Quentin Atkinson, and Andrew Meade. 2007. Frequency of Word-Use Predicts Rates of Lexical Evolution throughout Indo-European History. Nature 449:717–20.

Shana Poplack. 2018. Borrowing in the speech community. In Borrowing: Loanwords in the Speech Community and in the Grammar Oxford University Press.

References

Frank Seifart. 2019. 2. Contact-induced change, De Gruyter Mouton, Berlin, Boston, pages 13–23.

Elizabeth Skippington and Mark A. Ragan. 2013. Lateral Genetic Transfer and Cellular Networks, Springer New York, New York, NY, pages 123–135.

S.G. Thomason. 2001. Language Contact. Edinburgh University Press.

Thank you!