Derivational Morphemes as Markers of Borrowed Words in Czech

Abishek Stephen, Vojtěch John

September 5, 2025 at University of Fribourg

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

Outline

Outline - Introduction

Morphology of Borrowed Words

- When a foreign root or stem is borrowed from a language, the morphological integration within the linguistic subsystem of the recipient language is completed by derivational and inflectional morphemes (Poplack et al., 1988).

Morphology of Borrowed Words

- When a foreign root or stem is borrowed from a language, the morphological integration within the linguistic subsystem of the recipient language is completed by derivational and inflectional morphemes (Poplack et al., 1988).

- Bláha (2022) shows that verbal conjugation with the affix *-ova* is highly productive for borrowed roots (nominal), which can be seen as a mechanism for *marking* loanwords in Czech (Stephen and Žabokrtský, 2023, cf.). This is also the case for Russian, where the borrowed stems are productively attached to the verbalizer *-ova* (Wohlgemuth, 2009).

Morphology of Borrowed Words

- When a foreign root or stem is borrowed from a language, the morphological integration within the linguistic subsystem of the recipient language is completed by derivational and inflectional morphemes (Poplack et al., 1988).

- Bláha (2022) shows that verbal conjugation with the affix *-ova* is highly productive for borrowed roots (nominal), which can be seen as a mechanism for *marking* loanwords in Czech (Stephen and Žabokrtský, 2023, cf.). This is also the case for Russian, where the borrowed stems are productively attached to the verbalizer *-ova* (Wohlgemuth, 2009).

- The additional piece of evidence comes from the languages spoken in the Balkans where the formative affix *-s-* is productively used as a loanverb marker (Gardani et al., 2015).

Morphological Pathways of Borrowings

- The path through which an affix enters a language can also influence the loanword integration mechanisms. Affixes can be transferred as direct (borrowing of the affix directly from a donor language) or indirect borrowings (through multimorphemic words) (Seifart, 2015).

Morphological Pathways of Borrowings

- The path through which an affix enters a language can also influence the loanword integration mechanisms. Affixes can be transferred as direct (borrowing of the affix directly from a donor language) or indirect borrowings (through multimorphemic words) (Seifart, 2015).

- In Czech, we find instances which can be interpreted to mimic these two pathways in the absence of sufficient analysis. For example, the Czech noun *dysfunkční* ('dysfunctional'), is composed of a Latin origin derivational morpheme *dys*- and root morpheme *funk(c)*, and *houslista* ('violinist') with a native stem *housl* but a borrowed affix *-ist*.

Morphological Pathways of Borrowings

- The path through which an affix enters a language can also influence the loanword integration mechanisms. Affixes can be transferred as direct (borrowing of the affix directly from a donor language) or indirect borrowings (through multimorphemic words) (Seifart, 2015).

- In Czech, we find instances which can be interpreted to mimic these two pathways in the absence of sufficient analysis. For example, the Czech noun *dysfunkční* ('dysfunctional'), is composed of a Latin origin derivational morpheme *dys*- and root morpheme *funk(c)*, and *houslista* ('violinist') with a native stem *housl* but a borrowed affix *-ist*.

- But these cases cannot be very clearly distinguished until the individual morphemes are labeled based on their source types.

Partial Borrowing

- It remains unclear which component of a word, its root, the derivational affix, or the inflectional affix serves as the primary indicator of its status as a loanword.

Partial Borrowing

- It remains unclear which component of a word, its root, the derivational affix, or the inflectional affix serves as the primary indicator of its status as a loanword.

- The borrowed affixoids can sometimes function as roots themselves, for example, *supr-* in *suprový* ('super') highlighting that morpheme lexicality exists on a continuum rather than as a binary distinction.

Partial Borrowing

- It remains unclear which component of a word, its root, the derivational affix, or the inflectional affix serves as the primary indicator of its status as a loanword.
- The borrowed affixoids can sometimes function as roots themselves, for example, *supr-* in *suprový* ('super') highlighting that morpheme lexicality exists on a continuum rather than as a binary distinction.
- Therefore, a nuanced approach to extracting borrowing signals should involve analyzing the loan status of individual morphemes rather than entire words, focusing on various types of morphemes or affixes.

Outline - Workflow

Approach

- Training data is curated from DeriNet (Olbrich et al., 2025). All our training data and codes are made publicly available[1].

[1] `https://github.com/abishekjs/MorphAttention`

Approach

- Training data is curated from DeriNet (Olbrich et al., 2025). All our training data and codes are made publicly available[1].

- In our experiments, we use the task of labeling loanwords to extract *borrowedness* scores for individual morphemes that tell us to what extent the morphemes encompass the loan status.

[1] `https://github.com/abishekjs/MorphAttention`

Approach

- Training data is curated from DeriNet (Olbrich et al., 2025). All our training data and codes are made publicly available[1].

- In our experiments, we use the task of labeling loanwords to extract *borrowedness* scores for individual morphemes that tell us to what extent the morphemes encompass the loan status.

- We train a multinomial Naive Bayes classifier and extract the feature probabilities of the morphemes.

[1] `https://github.com/abishekjs/MorphAttention`

Approach

- Training data is curated from DeriNet (Olbrich et al., 2025). All our training data and codes are made publicly available[1].

- In our experiments, we use the task of labeling loanwords to extract *borrowedness* scores for individual morphemes that tell us to what extent the morphemes encompass the loan status.

- We train a multinomial Naive Bayes classifier and extract the feature probabilities of the morphemes.

- We compare the results using the attention mechanism (Vaswani et al., 2017), using an LSTM-based binary classifier. Additionally, we rope in a pre-trained RobeCzech (Straka et al., 2021) model for vizualising the attentions on the subword level.

[1] `https://github.com/abishekjs/MorphAttention`

Outline - Methodology

Data

- For our experiments, we use DeriNet v2.3, a lexical network that models the relations between words in the Czech lexicon. There are tags within DeriNet for native words and loanwords along with morphological segmentations.

Data

- For our experiments, we use DeriNet v2.3, a lexical network that models the relations between words in the Czech lexicon. There are tags within DeriNet for native words and loanwords along with morphological segmentations.

- We use the morph classifier by John (2024), to classify the morphemes into three categories- roots, derivational morphemes, and inflectional morphemes (Table 1).

Data

- For our experiments, we use DeriNet v2.3, a lexical network that models the relations between words in the Czech lexicon. There are tags within DeriNet for native words and loanwords along with morphological segmentations.
- We use the morph classifier by John (2024), to classify the morphemes into three categories- roots, derivational morphemes, and inflectional morphemes (Table 1).

POS	Borrowed	Native	Roots	Derivational Affixes	Inflectional Affixes
Noun	100079	197132	39850	4460	56
Verb	13378	42930	6714	719	2
Adjective	85320	199802	26927	3150	14
Adverb	45874	109465	20454	2318	21
Total	244651	549329	46246	5533	66

Table 1: Data overview with the frequency counts of native and borrowed words, along with the counts of affixes across those words conditioned by the POS categories.

Experiment: Naive Bayes Classifier I

- In the first experiment, we used the multinomial Naive Bayes classifier.

Experiment: Naive Bayes Classifier I

- In the first experiment, we used the multinomial Naive Bayes classifier.

- All morphemes in a native word are labeled as native, all morphemes in a borrowed word are labeled as borrowed, and as features we take unique id of the morpheme and its role in a given word (root vs. derivational vs. inflectional); thus, we take into account that the same morpheme might be derivational on some occasions and root on others. For example, *před-* in *předpokládaný* 'assumed' vs. in *přednosta* 'principal').

Experiment: Naive Bayes Classifier I

- In the first experiment, we used the multinomial Naive Bayes classifier.

- All morphemes in a native word are labeled as native, all morphemes in a borrowed word are labeled as borrowed, and as features we take unique id of the morpheme and its role in a given word (root vs. derivational vs. inflectional); thus, we take into account that the same morpheme might be derivational on some occasions and root on others. For example, *před-* in *předpokládaný* 'assumed' vs. in *přednosta* 'principal').

$$P(c \mid f_1, f_2) \propto P(c) \prod_{i=1}^{n} P(f_i \mid c)$$

Experiment: Naive Bayes Classifier I

- In the first experiment, we used the multinomial Naive Bayes classifier.
- All morphemes in a native word are labeled as native, all morphemes in a borrowed word are labeled as borrowed, and as features we take unique id of the morpheme and its role in a given word (root vs. derivational vs. inflectional); thus, we take into account that the same morpheme might be derivational on some occasions and root on others. For example, *před-* in *předpokládaný* 'assumed' vs. in *přednosta* 'principal').

$$P(c \mid f_1, f_2) \propto P(c) \prod_{i=1}^{n} P(f_i \mid c)$$

$$\hat{y} = \operatorname*{arg\,max}_{c \in C} P(c) \prod_{i=1}^{n} P(f_i \mid c)$$

Experiment: Naive Bayes Classifier II

- We extract predicted probabilities or the *borrowedness* score, which measures how strongly a given morpheme, in a specific role (e.g. *-ova* as a derivational suffix), is a marker of the borrowed status.

Experiment: Naive Bayes Classifier II

- We extract predicted probabilities or the *borrowedness* score, which measures how strongly a given morpheme, in a specific role (e.g. *-ova* as a derivational suffix), is a marker of the borrowed status.

- For example, borrowed roots are expected to appear mostly or exclusively in words marked as borrowed in DeriNet. As a result, their borrowedness score should be close to 1, since it is the only borrowed element in the word.

Experiment: Naive Bayes Classifier II

- We extract predicted probabilities or the *borrowedness* score, which measures how strongly a given morpheme, in a specific role (e.g. *-ova* as a derivational suffix), is a marker of the borrowed status.

- For example, borrowed roots are expected to appear mostly or exclusively in words marked as borrowed in DeriNet. As a result, their borrowedness score should be close to 1, since it is the only borrowed element in the word.

- In contrast, inflectional affixes tend to attach to both native and borrowed stems without strong preference. Therefore, their borrowedness score should be roughly proportional to the overall ratio of borrowed to native words in the dataset.

Experiment: Naive Bayes Classifier II

- We extract predicted probabilities or the *borrowedness* score, which measures how strongly a given morpheme, in a specific role (e.g. *-ova* as a derivational suffix), is a marker of the borrowed status.

- For example, borrowed roots are expected to appear mostly or exclusively in words marked as borrowed in DeriNet. As a result, their borrowedness score should be close to 1, since it is the only borrowed element in the word.

- In contrast, inflectional affixes tend to attach to both native and borrowed stems without strong preference. Therefore, their borrowedness score should be roughly proportional to the overall ratio of borrowed to native words in the dataset.

- We set a frequency threshold of ≥ 20, so that the classifier relies on only such morphemes that appear frequently enough to provide robust predictions.

Results: Naive Bayes Classifier I

- The mean (Table 2) of the feature likelihoods or predicted probabilities of the classifier correspond to the distribution of borrowed and native words in the data.

Statistics	Roots	Derivational Affixes	Inflectional Affixes
Mean	0.38	0.44	0.38
Variance	0.18	0.15	0.08

Table 2: The mean and variance of the feature likelihoods of the data.

Results: Naive Bayes Classifier I

- The mean (Table 2) of the feature likelihoods or predicted probabilities of the classifier correspond to the distribution of borrowed and native words in the data.

Statistics	Roots	Derivational Affixes	Inflectional Affixes
Mean	0.38	0.44	0.38
Variance	0.18	0.15	0.08

Table 2: The mean and variance of the feature likelihoods of the data.

- The derivational affixes show a higher score than roots and inflectional suffixes, which might be because of affixoids, which appear quite frequently in the data.

Results: Naive Bayes Classifier I

- The mean (Table 2) of the feature likelihoods or predicted probabilities of the classifier correspond to the distribution of borrowed and native words in the data.

Statistics	Roots	Derivational Affixes	Inflectional Affixes
Mean	0.38	0.44	0.38
Variance	0.18	0.15	0.08

Table 2: The mean and variance of the feature likelihoods of the data.

- The derivational affixes show a higher score than roots and inflectional suffixes, which might be because of affixoids, which appear quite frequently in the data.

- The variances (Table 2) are interesting, showing that the derivational affixes are almost as salient as the roots with regard to the origin of the words.

Results: Naive Bayes Classifier II

- We present the relative borrowedness scores for the derivational morphemes in Table 3.

Derivational Affixes	Frequency	RB Scores
fon	29	0.557
kilo	635	0.557
trans	718	0.557
ova	118329	0.003
sk	18101	0.005
iv	7515	0.007
haz	50	-0.440
řík	32	-0.440
přah	27	-0.440

Table 3: The relative borrowedness (RB) scores for the top, middle and the least ranking derivational affixes.

- The scores indicate how much more or less predictive a morpheme is compared to the mean borrowedness score of derivational affixes across all the POS categories.

Results: Naive Bayes Classifier III

- It should be noted, however, that quite often, roots got misidentified as derivational affixes and this method brought these errors up - generally, the more lexical a morpheme is, the more salient it tends to be for borrowedness identification; firstly, because it might cause the word to be classified as borrowed, secondly, because the borrowed derivational affixes, even if reanalyzed, do not tend to be much productive.

Results: Naive Bayes Classifier III

- It should be noted, however, that quite often, roots got misidentified as derivational affixes and this method brought these errors up - generally, the more lexical a morpheme is, the more salient it tends to be for borrowedness identification; firstly, because it might cause the word to be classified as borrowed, secondly, because the borrowed derivational affixes, even if reanalyzed, do not tend to be much productive.
- In very specific contexts and it is the more lexical ones (*-fikace*, *-ismus*) and these then cause the whole word to be more plausibly regarded as borrowed.

Results: Naive Bayes Classifier IV

POS	Derivational Affixes											
Noun	bio	ion	ex	ment	ism	multi	auto	anti	kilo	ing	inter	kom
Verb	kom	ment	ion	de	ex	di	iz	is	ur	para	isova	syn
Adjective	ion	ex	ment	multi	bio	inter	kom	kilo	trans	auto	ent	mikro
Adverb	ion	ment	ex	multi	kom	di	para	bio	anti	inter	trans	auto

Table 4: The derivational affixes with highest (left to right) relative borrowedness (RB) score segregated by POS tags.

- In Table 4, we present the derivational affixes with the highest RB scores. Affixes such as *bio*, *auto* have a higher RB score in nouns, adjectives, and adverbs.

Results: Naive Bayes Classifier IV

POS	Derivational Affixes											
Noun	bio	ion	ex	ment	ism	multi	auto	anti	kilo	ing	inter	kom
Verb	kom	ment	ion	de	ex	di	iz	is	ur	para	isova	syn
Adjective	ion	ex	ment	multi	bio	inter	kom	kilo	trans	auto	ent	mikro
Adverb	ion	ment	ex	multi	kom	di	para	bio	anti	inter	trans	auto

Table 4: The derivational affixes with highest (left to right) relative borrowedness (RB) score segregated by POS tags.

- In Table 4, we present the derivational affixes with the highest RB scores. Affixes such as *bio*, *auto* have a higher RB score in nouns, adjectives, and adverbs.
- Although the affixes *para* and *de* mostly render the verbs the loanword status. The important observation is that most of these affixes are affixoids.

Results: Naive Bayes Classifier IV

POS	Derivational Affixes											
Noun	bio	ion	ex	ment	ism	multi	auto	anti	kilo	ing	inter	kom
Verb	kom	ment	ion	de	ex	di	iz	is	ur	para	isova	syn
Adjective	ion	ex	ment	multi	bio	inter	kom	kilo	trans	auto	ent	mikro
Adverb	ion	ment	ex	multi	kom	di	para	bio	anti	inter	trans	auto

Table 4: The derivational affixes with highest (left to right) relative borrowedness (RB) score segregated by POS tags.

- In Table 4, we present the derivational affixes with the highest RB scores. Affixes such as *bio*, *auto* have a higher RB score in nouns, adjectives, and adverbs.
- Although the affixes *para* and *de* mostly render the verbs the loanword status. The important observation is that most of these affixes are affixoids.
- Usually borrowed as part of neoclassical compounds like *antibiotika* ('antibiotic') or *automat* ('dispenser') are productive enough to participate in word formation processes and also saliently preserve their foreign origin.

Results: Naive Bayes Classifier V

- The classifier struggles with compounds, noisy classification, and segmentation, and lacks contextual knowledge about the given morpheme. Additionally, it does not address homomorphy.

Results: Naive Bayes Classifier V

- The classifier struggles with compounds, noisy classification, and segmentation, and lacks contextual knowledge about the given morpheme. Additionally, it does not address homomorphy.
- To overcome these issues, we incorporate the attention mechanism in the classifiers.

Experiment: Attention-based LSTM Architecture

- We implement a light-weight classifier using the LSTM (Long Short-Term Memory) architecture.

Experiment: Attention-based LSTM Architecture

- We implement a light-weight classifier using the LSTM (Long Short-Term Memory) architecture.
- We embed individual morphemes using bidirectional LSTM cell; consequently, we pass these embeddings to four attention heads.

Experiment: Attention-based LSTM Architecture

- We implement a light-weight classifier using the LSTM (Long Short-Term Memory) architecture.
- We embed individual morphemes using bidirectional LSTM cell; consequently, we pass these embeddings to four attention heads.
- Then, we sum the attention-weighted embeddings and pass the resulting vector to one dense layer of size 512, and a sigmoid classification layer.

Experiment: Attention-based LSTM Architecture

- We implement a light-weight classifier using the LSTM (Long Short-Term Memory) architecture.
- We embed individual morphemes using bidirectional LSTM cell; consequently, we pass these embeddings to four attention heads.
- Then, we sum the attention-weighted embeddings and pass the resulting vector to one dense layer of size 512, and a sigmoid classification layer.
- We train the classifier for 8 epochs in batches of size 256 on a subset of training data.

Experiment: Attention-based LSTM Architecture

- We implement a light-weight classifier using the LSTM (Long Short-Term Memory) architecture.
- We embed individual morphemes using bidirectional LSTM cell; consequently, we pass these embeddings to four attention heads.
- Then, we sum the attention-weighted embeddings and pass the resulting vector to one dense layer of size 512, and a sigmoid classification layer.
- We train the classifier for 8 epochs in batches of size 256 on a subset of training data.
- For comparison, we used Naive Bayes and LinearSVM classifiers from Scikit-learn, using character and morpheme level embeddings.

Results: Attention-based LSTM Architecture I

- The results are presented in Table 5. For attention extraction, we used the mean values of the attention heads.

Method	ADJ			ADV			NOUN			VERB		
	Prec.	Rec.	Acc.	Prec.	Rec.	Acc.	Prec.	Rec.	Acc.	Prec.	Rec.	Acc.
C-LSVC	96.5	95.6	97.7	96.1	94.3	97.2	95.6	93.6	96.3	97.5	97.0	98.7
C-NB	87.0	92.6	93.8	86.0	92.8	93.4	87.3	92.2	92.8	86.4	94.6	95.1
M-LSVC	95.3	93.8	96.9	95.5	89.6	95.7	92.9	93.4	95.3	98.6	95.9	98.7
M-NB	92.3	93.8	95.9	90.6	91.8	94.8	91.6	90.6	94.0	95.9	93.6	97.5
Neural	93.3	89.8	87.0	92.0	90.0	82.0	92.1	88.4	88.5	88.5	88.0	60.5

Table 5: Results for the baselines (NB = Naive Bayes, LSVC = linear SVC; M - morpheme-based version, C = character-based version) and the neural classifier - in %

Results: Attention-based LSTM Architecture II

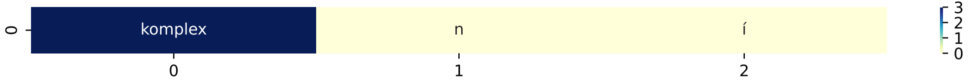
- As for the results (Table 6), the attention really seems to pick at least one of the borrowed morphemes in borrowed words and generally tend to pick roots or derivational affixes.

Classifier	R	D	I
Adjective	159487	119590	6045
Adverb	135210	20101	28
Noun	245616	50622	973
Verb	21984	34222	102

Table 6: The number of words for all the POS categories in which the roots (R), derivational affixes (D), and inflectional affixes (I) get the highest custom classifier attention scores.

Results: Attention-based LSTM Architecture III

(a) dynamicky



(b) komplexní

Results: Attention-based LSTM Architecture IV

(c) motocyklista

(d) telefonovat

Figure 1: The attention weights across morphemes extracted using our custom classifier.

Experiment: RobeCzech I

- To test how the subword embeddings can be leveraged for the binary classification task, we finetune RobeCzech (Straka et al., 2021).

Experiment: RobeCzech II

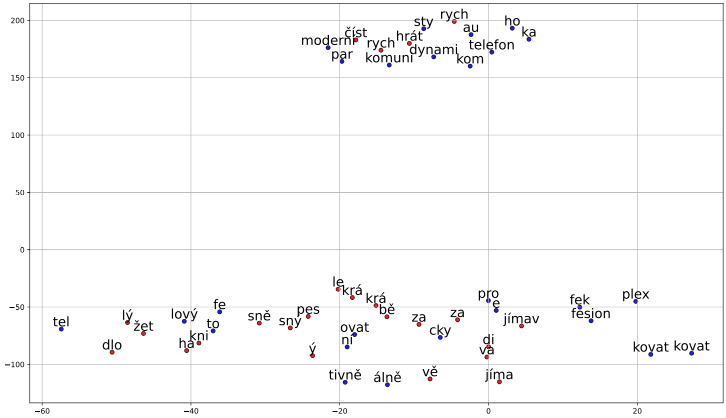


Figure 2: The subword embeddings of words extracted from RobeCzech. The native words and borrowed words are projected for nouns ● verbs ■ adjectives ◆ and adverbs ▲.

Results: RobeCzech I

The attention maps (Figure 3) show that the attention spans more than one subword to perform the binary class predictions. This suggests that a pre-trained language model like RobeCzech leverages the broader lexical context of the subwords to inform its predictions.

(a) dynamicky

(b) komplexní

Results: RobeCzech II

(c) motocyklista

(d) telefonovat

Figure 3: The attention weights across the subword tokens extracted from the RobeCzech based binary classifier.

Outline - Conclusion

Summary

1. We use DeriNet to train a multinomial Naive Bayes classifier and an attention-based binary LSTM classifier to classify Czech words as native or borrowed.

Summary

1. We use DeriNet to train a multinomial Naive Bayes classifier and an attention-based binary LSTM classifier to classify Czech words as native or borrowed.
2. We extract feature probabilities and attention scores from the classifiers for individual morphemes.

Summary

1. We use DeriNet to train a multinomial Naive Bayes classifier and an attention-based binary LSTM classifier to classify Czech words as native or borrowed.
2. We extract feature probabilities and attention scores from the classifiers for individual morphemes.
3. Our results indicate that the derivational morphemes serve as reliable markers of the borrowed status of words, along with root morphemes in Czech.

Summary

1. We use DeriNet to train a multinomial Naive Bayes classifier and an attention-based binary LSTM classifier to classify Czech words as native or borrowed.
2. We extract feature probabilities and attention scores from the classifiers for individual morphemes.
3. Our results indicate that the derivational morphemes serve as reliable markers of the borrowed status of words, along with root morphemes in Czech.
4. The comparison of results is further analyzed on the basis of the part-of-speech (POS) categories of the words.

Summary

1. We use DeriNet to train a multinomial Naive Bayes classifier and an attention-based binary LSTM classifier to classify Czech words as native or borrowed.
2. We extract feature probabilities and attention scores from the classifiers for individual morphemes.
3. Our results indicate that the derivational morphemes serve as reliable markers of the borrowed status of words, along with root morphemes in Czech.
4. The comparison of results is further analyzed on the basis of the part-of-speech (POS) categories of the words.
5. We also provide a glimpse of attention visualization based on the attention weights of RobeCzech based binary classifier.

Thank you!

stephen,john@ufal.mff.cuni.cz

References I

Ondřej Bláha. Dynamics of Conjugation Pattern “Kupuje” in Contemporary Czech (On Material of Journalistic Texts, 1990-2019). *Bohemica Olomucensia*, 14 (2):40–54, 2022. ISSN 0231634X.

Francesco Gardani, Peter Arkadiev, and Nino Amiridze. *Borrowed morphology: an overview*, pages 1–24. De Gruyter Mouton, Berlin, München, Boston, 2015. ISBN 9781614513209. doi: doi:10.1515/9781614513209.1. URL `https://doi.org/10.1515/9781614513209.1`.

Vojtěch John. *Morph classifier*. Univerzita Karlova, Matematicko-fyzikální fakulta, 2024.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach, 2019.

Michal Olbrich, Viktória Brezinová, Šárka Dohnalová, Vojtěch John, Lukáš Kyjánek, Aleš Papáček, Emil Svoboda, Magda Ševčíková, Jonáš Vidra, and Zdeněk Žabokrtský. DeriNet 2.3, 2025. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Shana Poplack, David Sankoff, and Christopher Miller. The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics*, 26: 47–104, 01 1988.

Frank Seifart. Direct and indirect affix borrowing. *Language*, 91(3):511–532, 2015.

Abishek Stephen and Zdeněk Žabokrtský. Understanding Borrowing through Derivational Morphology: A Case Study of Czech Verbs. In *The Fourth International Workshop on Resources and Tools for Derivational Morphology* , pages 49–59, Zagreb, Croatia, 2023. Croatian Language Technology Society. ISBN 978-953-55375-5-7.

Milan Straka, Jakub Náplava, Jana Straková, and David Samuel. *RobeCzech: Czech RoBERTa, a Monolingual Contextualized Language Representation Model*, page 197–209. Springer International Publishing, 2021. ISBN 9783030835279.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Jan Wohlgemuth. *Backmatter*, pages 303–459. De Gruyter Mouton, Berlin, New York, 2009. ISBN 9783110219340. doi: doi:10.1515/9783110219340.bm. URL `https://doi.org/10.1515/9783110219340.bm`.