
Article:
Bermel, N. orcid.org/0000-0002-1663-9322 and Knittl, L. (2023) Trajectories of change in paradigmatic cells in Czech. Naše řeč, 106 (5). pp. 247-274. ISSN 0027-8203

https://doi.org/10.58756/n51062301

Trajectories of change in paradigmatic cells in Czech[1]

Neil BERMEL | School of Languages and Cultures, University of Sheffield
Luděk KNITTL | School of Languages and Cultures, University of Sheffield

We examine a well-known phenomenon in the development of the Czech nominal declension system: the gradual supplanting of the original *o*-stem ending in the locative singular with the *u*-stem ending. We observe that, contrary to expectations from the literature based primarily on studies of English, this shift has been in progress for a millennium and, in the high-frequency nouns for which we have enough data to observe, the opposing trend is also frequently in evidence: the *o*-stem ending is introduced to lexemes where it was not found earlier. In the absence of a single, overriding motivation that could have derailed this shift from following the classic 'S-curve' pattern, we propose re-examining the *retextualization* model as a more fitting one for the complex interaction of factors and forms found in languages with complex inflectional morphology.

Key words: corpus linguistics, declension, diachrony, inflection, morphology

Klíčová slova: diachronie, flexe, korpusová lingvistika, skloňování, tvarosloví

1 Introduction

This article aims to interrogate a common historical approach to morphological change on the basis of Czech data and to examine why those data seem to show a different outcome not accounted for in this approach. Accounts of historical change in language tend to proceed from the vantage of their end points. Working backwards once we know what that end point is, we can trace how a change began and spread, with its exponents eventually becoming the new default selection. As to the way this change plays out, Croft (2000, p. 183) sums up the current consensus by saying that "the time course of the propagation of a language change typically follows an S-curve". By this he means that a change begins in a few places, then undergoes a period of rapid expansion, and then tends to slow down in its final phase, with some stubborn holdouts. Denison (2003, p. 56) describes this in more detail as follows:

> What actually happens much of the time is more like 'slow, slow, quick, quick, slow'. After the phase when the new form gains ascendancy rather rapidly, the process of change slows down again as the last remnants of the older state linger

[1] The authors gratefully acknowledge the support of Leverhulme Trust grant RPG-407 'Acceptability and Forced-choice Judgments in the Study of Linguistic Variation' and AHRC grant AH/T002859/1 'Feast and Famine: Confronting Overabundance and Defectivity in Language'.

> on. […] The whole thing can last hundreds of years altogether, indeed may never be wholly completed, but the bulk of the change is located within a much narrower slice of time where the slope is steeper.

Blythe and Croft (2012) propose that the S-curve is the basic model for change, but acknowledge that not all changes that start along an S-curve necessarily follow it to completion:

> To our knowledge there are no clearly documented cases of a change going toward completion that follows either a simple linear trajectory or an exponential curve (either slow start with a rapid completion and no tapering off, or an immediate rapid increase followed by a slow completion rate). There are, however, examples of variation that does not seem to be going toward completion, at least in the documented time period. These examples appear to exemplify either reasonably stable variation with the variants fluctuating around a mean percentage value, or a rise and fall of a relatively low frequency variant, commonly in competition with an incoming variant that is going to (near) completion on an S-shaped trajectory (Blythe – Croft, 2012, p. 280).

In other words, for unknown reasons, changes may 'stall' partway such that variation is maintained (as per Thornton, 2012), or the curve of change may be interrupted because another change 'derails' the original one. For example, a reanalysis could distribute the two variants in a new, stable configuration, or a new variant could further complicate the system.

An example of a reanalysis change in Slavonic morphology is the well-known development of a 'locative sub-case' in Russian. In this change, described, *inter alia*, by Brown (2007) and Janda (1996), a prepositional case morph from a defunct declensional class is absorbed into a new class, to which the bulk of its lexemes have migrated. According to Brown (2007, p. 63) this 'exapted' morph enjoys a period of prosperity in the 17th c., having expanded beyond its original remit, but by the 18th c. it has been reanalysed as a marker of location (cf. other, non-locational functions of the prepositional case) applied only to a subset of nouns in the class, and comments are found in the scholarly literature to the effect that from the 19th c. onwards fewer nouns occur with it. The increase in usage in the 17th c., which might have looked like the upward part of an S-curve, is followed by a decline in usage, such that today this locative morph is found with under a hundred nouns (Brown (2007, p. 64) cites Zaliznjak's *Grammatičeskij slovar' russkogo jazyka* [Grammatical Dictionary of Russian], which lists 88 such nouns).

Czech nominal declensions are an interesting test bed for 'residual' forms from older declension patterns. In common with all but the southeastern corner of the Slavonic world, Czech has portmanteau morphs that indicate the number, gender and case of the nominal referent. Reconstructions of the pre-written precursor from

which Czech is descended, Proto-Slavonic (PSl), posit a similarly elaborated declensional system based around stem classes. While direct evidence of these elements in the descendant languages is patchy, we can nonetheless detect or posit their presence for many forms of a given lexeme through comparative reconstruction and recourse to the oldest Slavonic texts (Bräuer, 1969; Janda, 1996, pp. 83–91; Meillet, 1965, pp. 379–432; Townsend – Janda, 1996, pp. 145–177).

The 'stem' for which the class is named refers to what Lamprecht et al. (1986, p. 137) call the "stem-formational formant" (*kmenotvorný formant*), i.e. the theme vowel or cluster comprising the initial element of the desinence, to which further formants were attached (see e.g. Janda (1996, p. 84) on the original Proto-Indo-European stem formants, for which she uses the traditional term *theme vowel*). Traditionally (if imprecisely), when a stem form has the propensity to attach the *a*-formant, we say it belongs to the *a*-stem class and so forth. There were over a dozen of these classes in PSl; larger classes contained nouns of various genders.

With substantial changes in the phonology of PSl and its inheritor dialects, the stem class system proved transitory. Modern Czech and other contemporary languages have gender and stem-final palatalization as their main organising principles. The complex evolution is depicted in part in Figure 1.

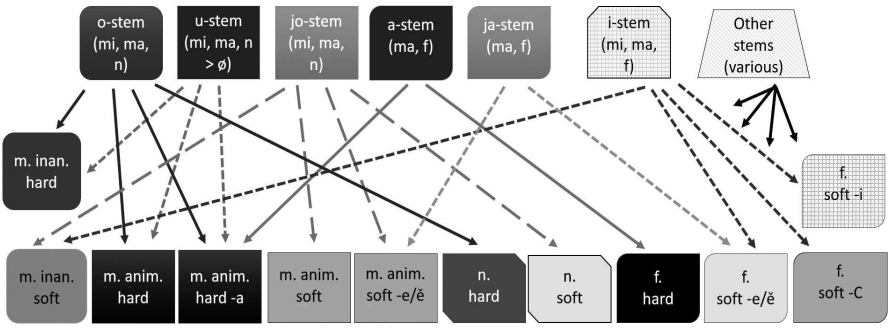

Figure 1: Declension classes in Proto-Slavonic and modern Czech.

The top line of Figure 1 shows the major stem classes reconstructed for PSl by the 9th c. The larger classes are *o*-stem, *u*-stem, *jo*-stem, *a*-stem, *ja*-stem and *i*-stem; there are a further dozen or so smaller classes (*en*-stem, *s*-stem, *ū*-stem, etc.). All the major classes encompassed words of multiple genders. The genders as found from the earliest written period onwards are indicated by the abbreviations *mi* (masculine inanimate), *ma* (masculine animate), *f* (feminine), *n* (neuter).

The bottom line of Figure 1 shows the declension classes of modern Czech (MoCz) by the time it emerges in the mid-19th c.: eleven major patterns, for which gram-

matical gender is the primary determinant and stem-final consonant the secondary determinant.

In between are arrows that show the trajectory of lexemes from PSl to MoCz. Although the trend is from fewer to more classes, we additionally see the break-up of some of the older groupings, with individual lexemes going to different patterns, causing a complex redistribution of material.

The notation of Figure 1 is schematic enough to cause difficulties. We speak of a lexeme 'moving' from one 'paradigm' in one 'system' to a 'different' paradigm in a different system. Adopting a cognitive, associative view of what a 'paradigm' is, we can propose at least three ways of describing it:

(1) for any lexeme, the paradigm is the sum of stored exemplars a speaker acquires in his lifetime;

(2) for any lexeme, the paradigm is the set of forms of all associated lexemes that we can, in the absence of stored knowledge, call upon to produce a form by analogy;

(3) for any group of lexemes, a paradigm is a 'shortcut' in the form of a generalization or higher-order schema that results from the routinization of the actions in (1) and (2).

Perspective (1) imposes a conservative view of the paradigm: that it replicates (with the inevitable slippage of human imprecision) what we have acquired. Perspective (2) introduces other patterns that can exert 'pressure' or influence on our choice of forms, opening the path to innovation. Perspective (3) formalizes this as a higher-order pattern that is explicitly shared, thus resembling much more the traditional view of 'membership' in an abstract category – although we should view it cautiously so as not to reify it. In this view, category membership results from many individuals reaching very similar results through similar exposure, so even in this 'bird's eye view' the potential for differing results – and thus for change – is built in.

The pattern in MoCz that we have studied over the last decade (Bermel, 2010; Bermel – Knittl, 2012a; 2012b; Bermel et al., 2015; 2018) is the left-most box in Figure 1 marked 'mi hard'. As can be seen, this hard masculine inanimate pattern inherits lexemes primarily from two PSl patterns: the *o*-stems and the *u*-stems. The *o*-stems were a large, open class of nouns with both animate and inanimate referents. The *u*-stems were a small class of one or two dozen nouns, some of which displayed high frequency (see e.g. *syn* 'son' above) and tended to be monosyllabic.[2]

[2] After the fall of the super-short *jer* vowels that had previously constituted an extra syllable on the end of the word.

Lamprecht et al. (1986, pp. 141, 150) posit that the 'Proto-Czech' (very late pre-written period) declension classes were as in Table 1.

class	*o*-stem *chlap* 'fellow'			*u*-stem *syn* 'son'		
case	singular	dual	plural	singular	dual	plural
nom.	chlap	chlapa	chlapi	syn	syny	synove
gen.	chlapa	chlapú	chlap	synu	synovu	synóv
dat.	chlapu	chlapoma	chlapóm	synovi	synma	synem
acc.	chlap	chlapa	chlapy	syn	syny	syny
voc.	chlape	chlapa	chlapi	synu	syny	synove
loc.	chlapě	chlapú	chlapiech	synu	synovu	synech
instr.	chlapem	chlapoma	chlapy	synem	synma	synmi

Table 1: Declension classes in 'Proto-Czech' (Lamprecht et al., 1986).

The modern patterns differ, having animacy as their guiding principle, as seen in Table 2.

class	masculine hard animate *chlap* 'fellow'		masculine hard inanimate *komín* 'chimney'	
case	singular	plural	singular	plural
nom.	chlap	chlapi	komín	komíny
gen.	chlapa	chlapů	komína/komínu	komínů
dat.	chlapovi/chlapu	chlapům	komínu	komínům
acc.	chlapa	chlapy	komín	komíny
voc.	chlape	chlapi	komíne	komíny
loc.	chlapovi/chlapu	chlapech	komíně/komínu	komínech
instr.	chlapem	chlapy	komínem	komíny

Table 2: Declension classes in modern Czech.

Lamprecht et al. (1986, p. 144) point out that the {ě} morph of the Proto-Czech locative singular (loc. sg.) case is now a secondary ending found alongside the dominant {u} morph. In an earlier study, we inventoried this class in the SYN2005 balanced corpus and found that only 53 nouns appear in the corpus with {ě} alone, and 392 lexemes appear with both {ě} and {u} morphs in this case.[3] This compares with

[3] The morph {ě} is pronounced [e] in MoCz and is spelled <ě> or <e> in Czech orthography. It is often accompanied by predictable alternations in the character of the preceding consonant or consonant cluster.

6,803 lexemes where {u} is the only morph in the loc. sg. (for details see Bermel – Knittl, 2012a, p. 99).

Modern grammars attempt to explain and motivate the distribution of endings in this case using semantic, phonological, word-formational, etymological and syntactic features (see Bermel, 1993; 2004; Bermel – Knittl, 2012a, pp. 94–95 for an overview). We will examine a few of these in Section 5, but they represent competing and overlapping tendencies, rather than conditioning the appearance of a particular ending in the contemporary language.

In MoCz, then, {u} has become the default loc. sg. ending for inanimate masculine nouns; we will term it the 'expansive' ending for this paradigm. By contrast, {ě} continues to be used with lexemes in the masc. inan. paradigm, and has not disappeared over the thousand-plus years that this change has been underway. We will term it the 'recessive' ending.

Using historical corpora of Czech, we aim to assess how this change is proceeding (as it is not yet complete) using the example of the loc. sg. If the S-curve model of change holds, we would expect to trace a slow first appearance, followed by a rapid period of change where the vast majority of the lexicon shifts to the expansive ending, and should now find ourselves in a long, slow 'tail' of the change. If this is not so, then we may need to posit a different description.

2 Historical corpora of Czech

There are three major corpora of historical Czech, two produced by the Czech National Corpus Institute (<www.korpus.cz>) and one by Czech Language Institute; we have made use of the former two.[4]

Diakorp is the public reference corpus accessed via the KonText interface, supporting *key word in context* (KWIC) view and a range of statistics. It contains 3.5m word forms (4.1m tokens) drawn from a range of time periods. Most of the text in it (48% of tokens) dates from the nineteenth century; other centuries are each represented by 3–18% of the total token count. The corpus has been orthographically normalized to aid in searching, but much variation inherent in texts from widely differing time periods has been left intact. Texts are annotated but the corpus is not morphologically tagged or lemmatized; regular expressions are used to retrieve variants or multiple word-forms (Kučera et al., 2015).

[4] We were made aware of the latter resource after the collection of these data was finished; the publicly available interface for this corpus at the time was not suitable for our purposes. For that reason, this analysis concentrates on developments in the later period (1800–2000), where the CNCI corpora are stronger. Developments in the earlier period can be the subject of further data collection and analysis in the future.

Diakon is a larger corpus (145m tokens) designed for use with the SyD interface. SyD does not display KWIC; instead, for historical data it displays graphs and tables with the frequency of competing features over adjustable time frames. It contains all the Diakorp texts, plus many others (including a selection from the synchronic SYN series of corpora), but has not undergone the thorough editing process used for Diakorp – hence actual concordances from it are not available.

We planned to use Diakon with the SyD interface to look up the historical trajectories of individual forms, getting a simple graphic and quantitative overview of their development without looking at concordances. However, without recourse to searches using tags or lemmas, there were difficulties with this approach. Many morphs in Czech, as can be seen above, are not unique to a single syntactic or paradigmatic slot. There is syncretism within paradigms and material is reused across paradigms. All of the morphs used in our study are polyfunctional, meaning a search on a specific form ending in that morph often draws numerous irrelevant forms into the net; this casts doubt on the validity of any purely quantitative exercise based on regular expressions.[5] The tremendous variation in form observed across premodern texts further problematized the reliability of our results.

For diachronic research questions, corpus size came up repeatedly as an issue. Even a corpus the size of Diakon was, at 145m tokens, problematically small in any given time frame for reliable data on oblique case forms of individual words, and global searches were compromised by the problems described above.

For example, a word like *obal* 'wrapping' as seen in Figure 2 has two loc. sg. exponents *obalu~obale*, and in the SYN2005 synchronic corpus they appear respectively 481 and 18 times. This is a promisingly large figure, but when we query Diakon via SyD, we find few examples in any time period: only seven texts between 1300 and 1800 contain these forms. Even though the evidence for a shift thereafter seems convincing, we are looking at only 10 examples across those five centuries, so we cannot deduce anything about the speed or process of the shift, and without access to KWIC, we do not know whether the {u} exponents in fact represent loc. sg. forms. Furthermore, there is the nagging worry that the two shifts seen in Figure 2 – from {ě} to {u} in 1800 and a reappearance of {ě} after 1980 – could be the result of the dramatically differing text proportions in those three periods. Texts are sparse up until 1800, then they get much longer and more numerous, with a further frequency leap after 1980.

[5] For example, {u} serves as a gen., dat., voc. and loc. sg. exponent within the paradigm we are examining, but also as the acc. sg. of some fem. nouns. Meanwhile, {a} can be the gen. sg. exponent in our paradigm, but can also be the nom. sg. exponent for fem. nouns, the acc. sg. exponent for some masc. anim. nouns and even some masc. inan. nouns, and {ě}, in addition to being a loc. sg. exponent in our paradigm, can be used to form the nom. sg., nom. pl., acc. pl. or gen. sg. of some fem. nouns and the dat. and loc. sg. of other fem. nouns.

With these problems complicating the methodology, we examined Diakon data via the KWIC interface of KonText. As even large and informative data sets from the SYN corpora can result in small and inconclusive data sets in Diakon, we limited our searches to the highest-frequency items from SYN2005: to be included, an item had to (1) display at least some variation in the given case in SYN2005 and (2) have 1000+ tokens in our survey of the SYN2005 corpus (see Bermel and Knittl (2012a) on our methodology for collecting data from SYN2005).[6] We were then able to remove examples from syncretic slots (e.g. dative and vocative forms) manually.[7]

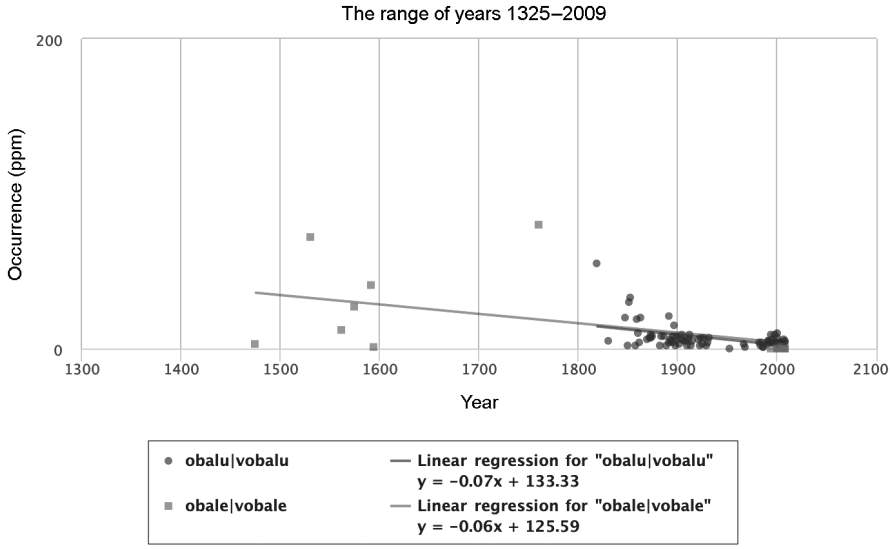

Figure 2: Sparse data on competition in forms from SyD.[8]

[6] Searches took into account as many variant spellings as possible. Many, but not all, of the texts in Diakon have been edited and their spelling normalized according to common working practice, to aid in searching, but some variation remains (for more on normalization, see <http://wiki.korpus.cz/doku.php/cnk:diakorp>).

[7] A left-context sort on the data from Diakon was used to group similar syntactic structures. To avoid replicating work we had already done on synchronic corpora, we only included data from Diakon up through 1950. While removing examples of case syncretism was the main task, we were alert to idiosyncratic issues that occur within individual lexemes, e.g. unstable gender diachronically in e.g. *způsob*/*způsoba* 'manner, fear, kind', etc.

[8] Figure 2 shows the output of our query in the SyD interface. The vertical axis in this graph shows *i.p.m.* (instances per million tokens) in a given year, while the horizontal axis shows the year. The total time span indicated in the header is 1325–2009. Regression lines for both search terms (*obalu*|*vobalu* and *obale*|*vobale*) are shown. The app outputs its graphs in colour, so certain aspects may not be visible here.

Given the uneven composition of the Diakon corpus, raw numbers yield odd-looking data, so for the analysis in Section 4 and following, we grouped our data into 50-year intervals starting in 1300 and finishing in 1950 and looked at *proportions of forms* in those intervals to gauge trajectories in that period, comparing this with data from the SYN2005 corpus.

3 The locative singular data

The data set for the loc. sg. includes 51 nouns with a frequency over 1000 loc. sg. forms in the SYN2005 corpus. The SYN2005 data have been manually verified to ensure that e.g. homonymous gen. sg. and dat. sg. forms are not accidentally included, and one noun – *den* 'day' – was subsequently eliminated due to the fact that it comes from an original consonant-stem declension and its forms are thus not representative of what happened to the original *u*-stem and *o*-stem types. All the remaining nouns show some evidence of variation between the two endings in SYN2005, although in some instances that variation is lopsided, with only a few examples of one exponent. The full list in frequency order is: *případ*, *svět*, *základ*, *život*, *dům*, *byt*, *stůl*, *zápas*, *les*, *stát*, *ostrov*, *proces*, *západ*, *východ*, *provoz*, *dvůr*, *tábor*, *jazyk*, *sál*, *kostel*, *obchod*, *areál*, *způsob*, *dopis*, *závod*, *úřad*, *led*, *okres*, *hrad*, *hlas*, *most*, *parlament*, *vůz*, *pád*, *zákon*, *venkov*, *úvod*, *oběd*, *obraz*, *časopis*, *obvod*, *bod*, *papír*, *kout*, *ústav*, *festival*, *koncert*, *přechod*, *klín*, *pořad*, *hřbitov* (English glosses can be found in Subsections 4.1–4.4 below and in the appendix). The distribution of these lexemes in SYN2005 is unsurprising, as seen in Figure 3: there are a few lexemes with tens of thousands of tokens, and the vast majority have fewer than five thousand tokens.

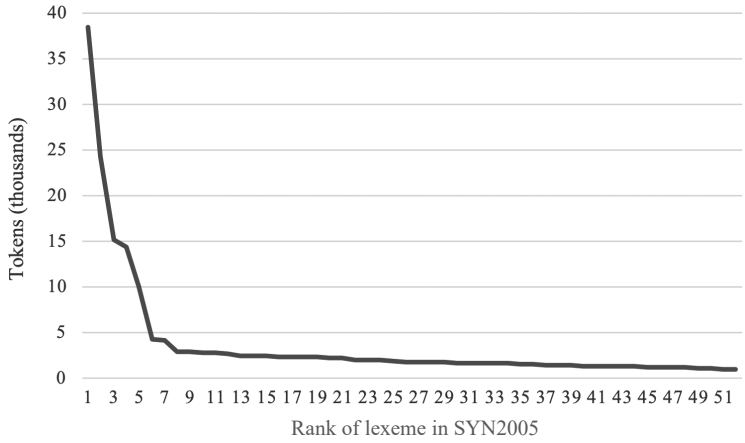

Figure 3: Distribution of top-ranked items in the SYN2005 corpus.

The data from Diakon were also manually sifted to remove homonymous forms, include common spelling variants in pre-modern Czech, etc. They prove to be on a similar curve, although less numerous. The highest-frequency item in the loc. sg. has just under 8,000 tokens, while two lexemes have no attestations in the loc. sg., as can be seen in Figure 4. Altogether the searches generated 38,826 contexts from Diakon, in addition to the existing 191,259 contexts from SYN2005.

Of the 51 lexemes investigated, 32 show a strong preference in SYN2005 – more than 80% of tokens – for the recessive {ě} ending. Twelve show a strong preference for the expansive {u} ending. Seven show a moderate preference for {ě} and one shows a moderate preference for {u}.

This predominance of the recessive ending is not what we would have expected. We proceed by examining some of the obvious possible explanations using data from Diakon, and in the absence of a clear finding, we look for alternatives elsewhere in the literature. The obvious explanations are as follows:

– These lexemes might have historically used the {ě} morph, and so what we see here is then no more than the conservation of an unusual shape in high-frequency forms – a common enough occurrence (see Section 4).

– There could be a confounding factor at work that is skewing the results towards {ě}. Perhaps the factors said to promote the use of {u} are absent for these lexemes or for the contexts in which they are used. We could then say that the problem lies in the distribution of our data (see Section 5).

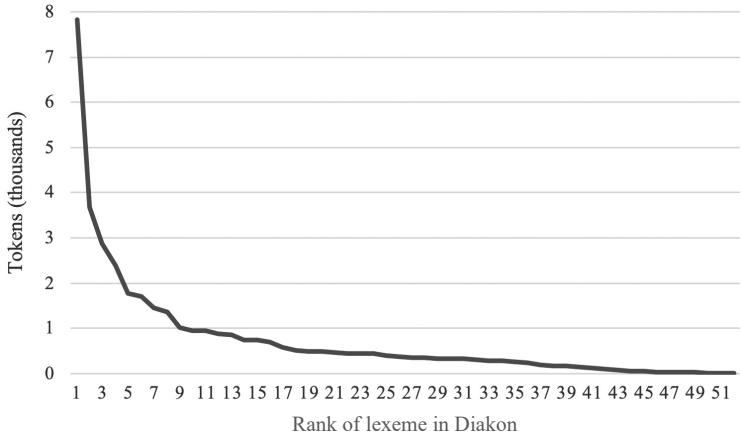

Figure 4: Distribution of items from Figure 3 in the Diakon corpus.

4 Conservation of high-frequency forms

The easiest explanation for our data is if they point to the conservation of morphologically unproductive high-frequency forms. We know that very high-frequency forms behave differently from lower-frequency forms. It is even suggested that they have special status as exceptions, making poor models for analogical extension of patterns.[9] In this view, it would be unsurprising if the highest-frequency forms represented an older state of the language.

When examining the data, a straightforward count of tokens is uninformative, as the data are irregularly distributed through the Diakon corpus and the pre-1800 numbers are dwarfed by the post-1800 totals. One way around this, as seen in the SyD interface in Figure 4, would be to calculate occurrences per million tokens for a moving window of the corpus, but this information is not readily retrievable for manual calculation. We settled on figuring the percentage of tokens in each time frame with {ě}/{u} vis-à-vis totals for both exponents. These can then be mapped onto a graph of the sort seen in Figure 5. The 50-year intervals allow us to judge the direction of change.

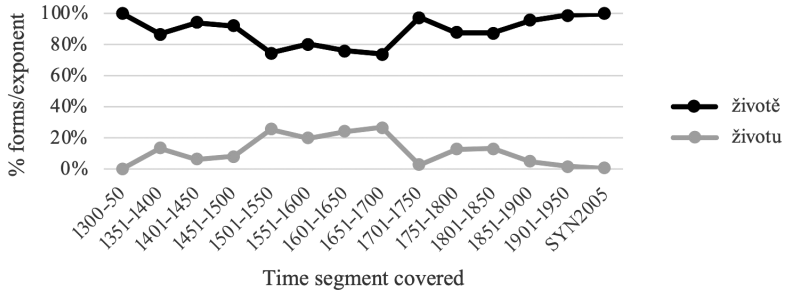

Figure 5: Percentage of loc. sg. forms for život ‘life’ in Diakon and SYN2005.

As can be seen in Figure 5, the {u} exponent for this lexeme is generally found quite rarely; for most periods it occurs in under 20% of the tokens, although it seems to

[9] For example, Bybee (2006) differentiates between what she calls “extreme high frequency” items and “high frequency” items. The former undergo demonstrably different sorts of changes, and she suggests that specifically high-frequency items, rather than extreme high frequency items, make the best sources for extending constructions to new items. She does not specify a single cut-off point, saying: “The impossibility at the moment of specifying ranges for extreme high, medium, and low is only a function of the state of our knowledge. As more empirical studies appear, absolute frequency ranges for each phenomena will eventually be specifiable” (Bybee, 2006, p. 715). Only a few of our loc. sg. cells would stand a chance of inclusion in the “extreme high” category, with a frequency of 10,000 or more in a 100m-token corpus: *případ* ‘case’ (38,506), *svět* ‘world’ (24,214), *základ* ‘basis’ (15,133), *život* ‘life’ (14,341), *dům* ‘house, building’ (10,072).

rise a bit in the period from 1500–1700 before dipping again. The {ě} exponent dominates throughout the period and achieves close to 100% dominance by the modern period.

Many lexemes are sparsely attested in Diakon prior to 1800, making the graphs unilluminating, as seen in Figure 6. There are several reasons for this. Our selection of words is based on frequency counts from SYN2005, whose composition is 1/3 journalistic texts. Therefore, there is a bias towards the journalistic lexicon of the late 1990s, including some recent borrowings and administrative terms. Early texts, on the other hand, are stronger on religious content: lexemes such as *život* 'life' and *zákon* 'law' are frequent, while everyday lexemes such as *led* 'ice' and *oběd* 'lunch' are rare.

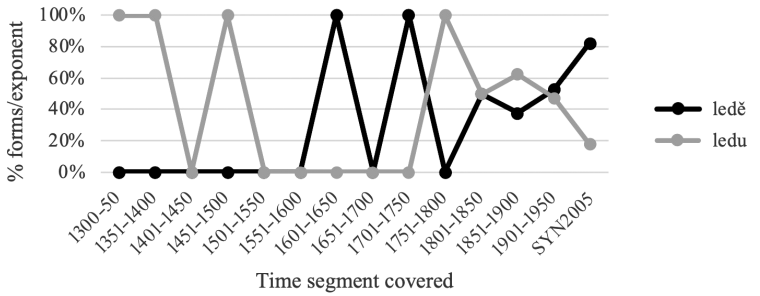

Figure 6: Percentage of loc. sg. forms for *led* 'ice' in Diakon and SYN2005.

A clearer picture appears if we collapse the pre-1800 slots into one, as seen in Figure 7, focusing on developments from the 19th c. onwards.

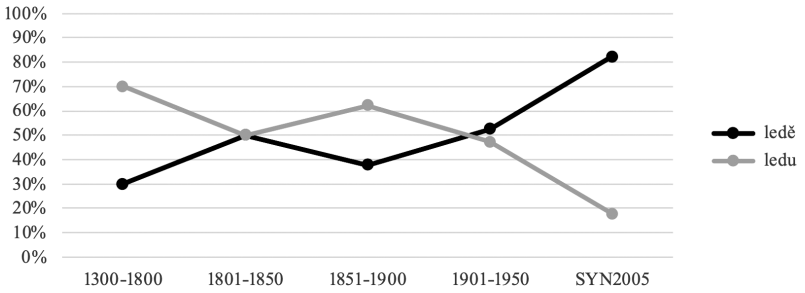

Figure 7: Percentage of loc. sg. forms for *led* 'ice' in Diakon and SYN2005 (2).

A table summarizing the data from this exercise is in the Appendix. The lexemes fall into four broad patterns of behaviour, as seen in Subsections 4.1–4.4.

4.1 Conservation of original {ě} – 16 lexemes

When we examine the data this way, we find that 11 lexemes show consistent conservation of an original {ě} exponent, as seen in Table 3.

Loc. sg. form		1300–1800		1801–1850		1851–1900		1901–1950		SYN2005	
dvoře	'courtyard'	423	96.80%	156	88.64%	535	96.75%	513	95.71%	2331	99.53%
dvoru		14	3.20%	20	11.36%	18	3.25%	23	4.29%	11	0.47%
hradě	'castle'	360	90.23%	129	75.44%	418	85.48%	256	81.79%	1518	92.56%
hradu		39	9.77%	42	24.56%	71	14.52%	57	18.21%	122	7.44%
hřbitově	'cemetery'	5	100.00%	55	96.49%	221	95.26%	194	97.98%	997	99.30%
hřbitovu		0	0.00%	2	3.51%	11	4.74%	4	2.02%	7	0.70%
kostele	'church'	719	99.58%	106	99.07%	387	98.98%	549	99.46%	2210	99.42%
kostelu		3	0.42%	1	0.93%	4	1.02%	3	0.54%	13	0.58%
koutě	'corner'	59	96.72%	59	100.00%	265	96.72%	281	96.23%	1178	96.08%
koutu		2	3.28%	0	0.00%	9	3.28%	11	3.77%	48	3.92%
lese	'forest'	176	94.12%	247	98.02%	1147	99.57%	800	99.50%	2845	99.82%
lesu		11	5.88%	5	1.98%	5	0.43%	4	0.50%	5	0.18%
mostě	'bridge'	24	75.00%	38	100.00%	96	97.96%	115	100.00%	1506	93.31%
mostu		8	25.00%	0	0.00%	2	2.04%	0	0.00%	108	6.69%
obědě	'dinner'	72	98.63%	22	100.00%	111	98.23%	178	97.27%	1361	99.56%
obědu		1	1.37%	0	0.00%	2	1.77%	5	2.73%	6	0.44%
ostrově	'island'	99	72.26%	56	90.32%	56	93.33%	202	98.06%	2730	99.06%
ostrovu		38	27.74%	6	9.68%	4	6.67%	4	1.94%	26	0.94%
případě	'case'	1	100.00%	0	0.00%	217	88.21%	707	92.06%	37059	96.24%
případu		0	0.00%	4	100.00%	29	11.79%	61	7.94%	1447	3.76%
stole	'table'	109	89.34%	80	98.77%	307	95.34%	416	97.42%	4138	99.59%
stolu		13	10.66%	1	1.23%	15	4.66%	11	2.58%	17	0.41%
světě	'world'	2538	94.49%	799	99.25%	1961	99.09%	2351	99.62%	24189	99.90%
světu		148	5.51%	6	0.75%	18	0.91%	9	0.38%	25	0.10%
táboře	'camp'	25	96.15%	5	45.45%	119	92.25%	204	96.23%	2282	99.43%
táboru		1	3.85%	6	54.55%	10	7.75%	8	3.77%	13	0.57%
venkově	'village'	0	0.00%	12	100.00%	160	99.38%	318	96.36%	1443	99.45%
venkovu		0	0.00%	0	0.00%	1	0.62%	12	3.64%	8	0.55%
voze	'vehicle'	91	94.79%	53	89.83%	174	93.05%	236	96.33%	1558	98.11%
vozu		5	5.21%	6	10.17%	13	6.95%	9	3.67%	30	1.89%
životě	'life'	856	87.44%	68	87.18%	576	95.36%	1201	98.60%	14300	99.71%
životu		123	12.56%	10	12.82%	28	4.64%	17	1.40%	41	0.29%

Table 3: Lexemes conserving original {ě}.

Almost all the lexemes in Table 3 have over 90% convergence on the {ě} exponent in all time frames. Three lexemes show a generally high preference for {ě} but in the 19th c. there is a short-lived increase in {u} forms: *tábor* ‘camp’, *hrad* ‘castle’, *případ* ‘case’. In two instances, {ě} shifts from being the more frequent exponent to the only commonly used exponent: *ostrov* ‘island’, *život* ‘life’. Conservation of an original {ě} exponent thus occurs in just over a quarter of the lexemes examined.

4.2 Maintenance of {u} – 10 lexemes

The ten lexemes which maintain {u} from the earliest attestations are found in Table 4.

Loc. sg. form		1300–1800		1801–1850		1851–1900		1901–1950		SYN2005	
areále	‘campus’	0	0.00%	0	0.00%	0	0.00%	0	0.00%	1	0.05%
areálu		0	0.00%	0	0.00%	0	0.00%	0	0.00%	2014	99.95%
festivale	‘festival’	0	0.00%	0	0.00%	0	0.00%	0	0.00%	5	0.42%
festivalu		0	0.00%	0	0.00%	0	0.00%	0	0.00%	1177	99.58%
obvodě	‘district’	0	0.00%	0	0.00%	4	6.90%	10	11.36%	243	18.56%
obvodu		0	0.00%	1	100.00%	54	93.10%	78	88.64%	1076	81.44%
pádě	‘case, fall’	1	1.05%	1	2.04%	7	3.95%	28	23.93%	32	2.03%
pádu		94	98.95%	48	97.96%	170	96.05%	89	76.07%	1548	97.97%
pořadě	‘program’	0	0.00%	4	66.67%	3	27.27%	2	14.29%	1	0.10%
pořadu		4	100.00%	2	33.33%	8	72.73%	12	85.71%	1014	99.90%
přechodě	‘crossing, transfer’	0	0.00%	0	0.00%	0	0.00%	0	0.00%	2	0.18%
přechodu		0	0.00%	1	100.00%	9	100.00%	20	100.00%	1113	99.82%
procese	‘process, trial’	1	33.33%	0	0.00%	0	0.00%	1	3.03%	4	0.15%
procesu		2	66.67%	1	100.00%	2	100.00%	32	96.97%	2655	99.85%
provoze	‘traffic, operation’	0	0.00%	0	0.00%	0	0.00%	0	0.00%	19	0.79%
provozu		0	0.00%	0	0.00%	0	0.00%	10	100.00%	2395	99.21%
úvodě	‘introduction’	1	100.00%	0	0.00%	40	45.45%	58	56.31%	4	0.29%
úvodu		0	0.00%	3	100.00%	48	54.55%	45	43.69%	1380	99.71%
způsobě	‘manner’	29	9.60%	12	17.39%	22	8.27%	9	3.61%	2	0.10%
způsobu		273	90.40%	57	82.61%	244	91.73%	240	96.39%	1981	99.90%

Table 4: Maintenance of {u}.

Five lexemes make their first appearance with, and maintain throughout, a strong preference for the {u} exponent: *areál* ‘campus’, *festival* ‘festival’, *přechod* ‘crossing’, *proces* ‘trial’, *provoz* ‘traffic’, although none appear in the corpus in the earliest period. With an additional three lexemes – *pád* ‘fall, case’, *pořad* ‘program’, *způsob* ‘manner’ – there is a temporary drop in popularity of the {u} form in

respectively the beginning of the 19th and 20th centuries.[10] One lexeme, *úvod* 'introduction', shows signs of shifting to {ě} but this changes dramatically after 1950 and there are few examples of it in SYN2005. A further, *obvod* 'district, circumference' is a recent coinage but shows a gradual increase in {ě} forms since its introduction in the early 19th c.

4.3 Shift from {u} to {ě} – 12 lexemes

This group is summarized in Table 5. It is surprisingly numerous and includes both single-syllable words that are often said to be characteristic of this group as well as deverbal nouns and borrowings, which should according to canonical definitions not have shifted.

Loc. sg. form		1300–1800		1801–1850		1851–1900		1901–1950		SYN2005	
bytě	'flat'	0	0.00%	1	3.57%	33	35.11%	140	86.42%	4229	99.23%
bytu		64	100.00%	27	96.43%	61	64.89%	22	13.58%	33	0.77%
domě	'house'	846	53.54%	313	96.60%	749	93.86%	915	93.85%	10032	99.60%
domu		734	46.46%	11	3.40%	49	6.14%	60	6.15%	40	0.40%
hlase	'voice'	8	16.33%	9	56.25%	122	75.78%	87	80.56%	1461	89.52%
hlasu		41	83.67%	7	43.75%	39	24.22%	21	19.44%	171	10.48%
klíně	'lap'	34	85.00%	33	45.21%	333	60.55%	228	77.82%	1025	94.47%
klínu		6	15.00%	40	54.79%	217	39.45%	65	22.18%	60	5.53%
ledě	'ice'	6	30.00%	7	50.00%	20	37.74%	40	52.63%	1422	82.29%
ledu		14	70.00%	7	50.00%	33	62.26%	36	47.37%	306	17.71%
okrese	'county'	0	0.00%	1	50.00%	32	22.07%	56	44.44%	1609	94.54%
okresu		0	0.00%	1	50.00%	113	77.93%	70	55.56%	93	5.46%
sále	'hall'	1	50.00%	17	85.00%	104	94.55%	122	96.06%	2216	96.85%
sálu		1	50.00%	3	15.00%	6	5.45%	5	3.94%	72	3.15%
státě	'state'	1	100.00%	1	10.00%	15	17.44%	112	84.85%	2470	87.43%
státu		0	0.00%	9	90.00%	71	82.56%	20	15.15%	355	12.57%
východě	'east, exit, sunrise'	3	5.77%	13	27.66%	158	71.17%	99	76.15%	2309	94.79%
východu		49	94.23%	34	72.34%	64	28.83%	31	23.85%	127	5.21%
základě	'basis'	9	25.00%	9	45.00%	374	95.90%	397	96.36%	14899	98.45%
základu		27	75.00%	11	55.00%	16	4.10%	15	3.64%	234	1.55%

[10] An alternate fem. lexeme *způsoba* overlapping with the masc. *způsob* was common through the 19th c. We removed fem. forms from our data where detectable (gender is shown frequently in adjective and verb agreement), but it was not always possible to tell.

západě	'west, turn, sunset'	1	3.13%	10	25.00%	130	57.02%	112	57.44%	2162	87.74%
západu		31	96.88%	30	75.00%	98	42.98%	83	42.56%	302	12.26%
zápase	'struggle, match'	0	0.00%	2	8.00%	100	56.32%	73	72.28%	2677	92.63%
zápasu		2	100.00%	23	92.00%	83	43.68%	28	27.72%	213	7.37%

Table 5: Shift from {u} to {ě}.

There are two additional oddities that can be seen in Table 6: *ústav* 'institute', *parlament* 'parliament'. In both, we see an increasing use of {ě}, reaching c. 70–90% in the interwar period, but these {ě} forms are rarely encountered in SYN2005.[11]

Loc. sg. form		1300–1800		1801–1850		1851–1900		1901–1950		SYN2005	
parlamentě	'parliament'	0	0.00%	0	0.00%	4	80.00%	57	89.06%	116	7.30%
parlamentu		12	100.00%	5	100.00%	1	20.00%	7	10.94%	1473	92.70%
pořadě	'program'	0	0.00%	4	66.67%	3	27.27%	2	14.29%	1	0.10%
pořadu		4	100.00%	2	33.33%	8	72.73%	12	85.71%	1014	99.90%

Table 6: {u} becomes {ě} and then again {u}.

4.4 No clear resolution towards {ě} or {u} – 11 lexemes

Most of the lexemes in this group, seen in Table 7, show a gradual shift towards {ě} that is not complete.

Loc. sg. form		1300–1800		1801–1850		1851–1900		1901–1950		SYN2005	
bodě	'point'	0	0.00%	0	0.00%	2	25.00%	25	64.10%	980	75.73%
bodu		0	0.00%	1	100.00%	6	75.00%	14	35.90%	314	24.27%
časopise	'magazine'	0	0.00%	7	26.92%	238	88.48%	111	79.86%	1077	81.65%
časopisu		0	0.00%	19	73.08%	31	11.52%	28	20.14%	242	18.35%
dopise	'letter'	2	100.00%	1	33.33%	12	27.91%	68	80.00%	1623	88.11%
dopisu		0	0.00%	2	66.67%	31	72.09%	17	20.00%	219	11.89%
jazyce	'language, tongue'	14	7.25%	0	0.00%	15	7.85%	46	21.80%	1427	62.26%
jazyku		179	92.75%	138	100.00%	176	92.15%	165	78.20%	865	37.74%
koncertě	'concert'	0	0.00%	1	33.33%	6	37.50%	5	45.45%	572	49.48%
koncertu		0	0.00%	2	66.67%	10	62.50%	6	54.55%	584	50.52%

[11] In the case of *ústav*, this may be connected with the existence of the paronym *ústava* 'constitution', a feminine noun with the loc. sg. form *ústavě*.

obchodě	‘shop, trade’	3	8.11%	1	3.70%	9	20.45%	30	54.55%	1529	68.87%
obchodu		34	91.89%	26	96.30%	35	79.55%	25	45.45%	691	31.13%
obraze	‘picture’	16	29.09%	22	51.16%	99	83.90%	79	75.24%	1045	78.99%
obrazu		39	70.91%	21	48.84%	19	16.10%	26	24.76%	278	21.01%
papíře	‘paper’	10	52.63%	23	69.70%	68	61.82%	114	75.50%	932	72.81%
papíru		9	47.37%	10	30.30%	42	38.18%	37	24.50%	348	27.19%
úřadě	‘office’	39	36.11%	14	50.00%	61	73.49%	117	86.67%	1039	59.20%
úřadu		69	63.89%	14	50.00%	22	26.51%	18	13.33%	716	40.80%
zákoně	‘law’	500	92.59%	15	62.50%	40	42.11%	59	64.13%	1206	80.45%
zákonu		40	7.41%	9	37.50%	55	57.89%	33	35.87%	293	19.55%
závodě	‘competition, factory’	4	66.67%	1	25.00%	11	52.38%	22	78.57%	1421	79.56%
závodu		2	33.33%	3	75.00%	10	47.62%	6	21.43%	365	20.44%

Table 7: No clear resolution towards either ending.

One lexeme shows a marked change in the last 75 years. The frequency of {ě} for *úřad* ‘office’ had been on the rise, but drops markedly in the post-war period.

4.5 Summary

We hypothesized that the prevalence of the recessive exponent {ě} might have been due to the fact that irregular endings are more easily conserved for high-frequency lexemes. However, our data show this not to be the case. In fact, a substantial group of nouns show an ongoing or partial shift from the expansive ending to the recessive ending, contrary to what we would have expected. We thus turn to a second possibility in the next section.

5 Potential confounding factors

A possible explanation for the predominance of {ě} is that our data are unrepresentative of the lexicon. Non-representativity might be caused by other factors said to influence the choice of loc. sg. ending. These are examined in Subsections 5.1–5.4.

5.1 Phonological factors

Standard handbooks suggest that different stem-final consonants are associated with {u} and {ě}. The most exhaustive study examining this contention is Štícha (2009), which identifies a single default {u} morph and some phonological contexts where {ě} appears. He finds that stems ending in /h, g, f, k, ch, r, p, b, m/ take {u}, the first three without exception; stems ending in /d, t, n, s, z, l/ may have either {ě} or {u},

with {ě} being most likely to appear for the final three, as summarized in Table 8 (note that our data do not include all the phonological environments discussed in Štícha (2009)).

Stem	Exponent (Štícha, 2009)	Lexemes from our study
-l	sometimes {ě}	4.1{ě}: kostel, stůl 4.2{u}: areál, festival 4.3{u>ě}: sál
-z	sometimes {ě}	4.1{ě}: vůz 4.2{u}: provoz 4.4{u~ě}: obraz
-s	sometimes {ě}	4.1{ě}: les 4.2{u}: proces 4.3{u>ě}: hlas, okres, zápas 4.4{u~ě}: časopis, dopis
-d	occasionally {ě}	4.1{ě}: hrad, oběd, případ 4.2{u}: obvod, pád, pořad, přechod, úvod 4.3{u>ě}: led, východ, základ, západ 4.4{u~ě}: bod, obchod, úřad, závod
-t	occasionally {ě}	4.1{ě}: byt, kout, most, svět, život 4.3{u>ě}: parlament, stát 4.4{u~ě}: koncert
-n	occasionally {ě}	4.3{u>ě}: klín 4.4{u~ě}: zákon
-v	predominantly {u}	4.1{ě}: hřbitov, ostrov, venkov 4.3{u>ě}: ústav
-m	predominantly {u}	4.3{u>ě}: dům
-b	predominantly {u}	4.2{u}: způsob
-r	predominantly {u}	4.1{ě}: dvůr, tábor 4.4{u~ě}: papír
-k	predominantly {u}	4.4{u~ě}: jazyk

Table 8: Phonological environments.

As there are no phonological environments that specifically condition the use of {ě}, it seems unlikely that this would provide a complete answer. The words in our survey fall into the environments as detailed in Table 8. Of these, 15 (28.8%) are in environments where {ě} is sometimes found; 27 (51.9%) are in environments where {ě} is occasionally found; and 10 (19.2%) are in environments where {ě} is rarely found.

Among the highly frequent words in our survey, there is no clear alignment with the environments from Štícha (2009): they do not contradict Štícha's findings over

the entirety of the nominal lexicon, which are phrased as tendencies, but neither do they reflect them consistently. Some environments, such as *-l*, *-s*, *-z*, *-d*, are in fact found 'sometimes', but in environments where {ě} should be found only 'occasionally' or {u} found 'predominantly', we find the recessive ending more frequently than we might expect.

5.2 Word-formational and etymological factors

These two factors are different in nature, but as the word-formational tendencies are said to apply only to native words, and etymologically, borrowings are said to have their own specific tendencies, we can treat them together and cover the entire lexicon. Deverbal nouns (formed from verbal roots) are judged likely to take {u} (PMČ, p. 253) as are borrowings (MČ2, p. 305). The {ě} variant is said to be found more often with nouns from the basic lexicon indicating things or places.

We can identify four corresponding categories. Eighteen nouns have simplex stems of native Czech origin; three further nouns have the suffix {ov} but are formed on native roots. A further 22 are formed from native Czech verbs and all except *pád*, *vůz* have the form *prefix* + *verbal root*. There are eight borrowed words, of which three are of such ancient standing that we have put them in a separate category (*nativized*). The list is summarized in Table 9.

Feature	Expected outcome	Actual outcome
Native simplex + {ov}	usually {ě}	4.1{ě}: hřbitov, ostrov, venkov
Native simplex	sometimes {ě}	4.1{ě}: dvůr, hrad, les, kout, led, most, oběd, stůl, svět, tábor, život 4.3{u>ě}: byt, dům, hlas, klín, sál 4.4{u~ě}: bod, jazyk
Nativized borrowing	sometimes {ě}	4.1{ě}: kostel 4.3{u>ě}: stát 4.4{u~ě}: papír
Native deverbal	usually {u}	4.1{ě}: případ, vůz 4.2{u}: obvod, pád, pořad, provoz, přechod, úvod, způsob 4.3{u>ě}: okres, ústav, východ, základ, západ, zápas 4.4{u~ě}: časopis, dopis, obchod, obraz, úřad, zákon, závod
Borrowing	almost always {u}	4.2{u}: areál, festival, proces 4.3{u>ě}: parlament 4.4{u~ě}: koncert

Table 9: Word-formational and etymological features.

The final column of Table 9 shows the actual development sketched in Subsections 4.1–4.4 vis-à-vis the expected outcome in the second column. If structural and etymological reasons were behind these developments, then we would see a predominance of simplex roots and nativized borrowings, as these should show a consistent preference for the {ě} exponent.

The picture is in fact more complex. Native simplexes do show a preference for {ě}, but in many instances, it is the result of a shift, rather than maintenance. The deverbal and borrowing groups are both relatively large and their development is even more heterogeneous, with many of them showing a shift over time towards the recessive ending. Neither word-formational factors nor etymological factors can thus account on their own for the behaviour of our sample lexicon.

5.3 Semantic factors

According to some handbooks (e.g. Grepl et al., 1995, p. 253; Cvrček et al., 2010, p. 164), where a noun exhibits polysemy, it is more likely to have {ě} when referring to a thing or place than when referring to a process. This affects a subset of the deverbals discussed above in Subsection 5.2:

- *obchod* 'trade/shop' – Diakon has {u} as the predominant ending in all senses of the word; the occasional uses of {ě} also cover the range of meanings as late as the end of the 19th c. Its early 20th c. examples mainly show a division between the concrete ('shop') meaning and the abstract ('trade') meaning.
- *pád* 'fall/instance, (grammatical) case' – Diakon attests {ě} only once before the mid-19th c.; all other examples, mostly of 'fall', have {u}, although some of these also represent the meaning 'case, instance'. The recessive ending seems to flourish briefly in the late 19th and early 20th c., but even there, {u} predominates in all meanings.
- *provoz* 'operation/traffic' – Diakon has no examples of {ě} before 1950; all examples are {u}, including two from the early 20th c. in the locational meaning 'traffic'. Modern examples of {ě} from SYN2005 show both meanings: 'traffic' and 'operation'.
- *přechod* 'transit/crossing' – Diakon has no examples of {ě} in any period. The existing examples cover the range of meanings.
- *východ* 'exit/sunrise, east' – prior to the 19th c. {u} seems to have been the default exponent; in Diakon we have three examples of {ě} from 1417 and then nothing until 1829, after which point it is reserved for the meaning 'east'. The {u} ending is attested throughout, mostly in the meaning 'exit' or 'sunset', and continues to be used in all meanings through the 19th c. and early 20th c. Compare this with only four examples of *na východu* in SYN2005, vis-à-vis 1,358 examples of *na východě*.

– *západ* ‘turning/sunset, west’ – similar to *východ*, there is only one example with {ě} in Diakon prior to 1840. From the mid-19th c. it gradually becomes the default in the meaning ‘west’. Examples of {u} in the meaning ‘west, sunset’ are attested throughout and continue to be frequent through the 19th c.

To sum up, there is sporadic historical evidence for this differentiation between location and process, but for the most part it seems to have evolved recently and not entirely consistently. We thus need to look elsewhere to motivate our findings.

5.4 Syntactic factors

Bermel (1993; 2004) suggested that canonical locative phrases with the prepositions *v*, *na* ‘in, on’ increase the chances of {ě} appearing or being highly rated by speakers; conversely, non-canonical locative phrases (with other prepositions or interposed adjectives) increase the occurrence and ratings of {u}. This is also mentioned in handbooks (MČ2, p. 305, Cvrček et al., 2010, p. 164). It is in theory possible that syntactic constructions favoring {ě} might be more popular in later time periods; however, this does not seem all that likely.

We can check this assumption quickly with data from lexemes showing the unexpected shift from {u} > {ě}. Two examples will be given below. If this hypothesis is correct, then we should see more non-canonical contexts in earlier texts, favoring the {u} ending, while more recent contexts should be more canonical, favoring {ě}.

The noun *zápas* ‘match, struggle, clash’ shows no evidence of such a contextual shift. Overwhelmingly the most popular context in all time periods is the canonical one with the preposition *v*/*ve* ‘in’, with or without an interposed adjective. Prior to 1800, it appears only twice, with {u}. The first example of {ě} appears in 1800–1850, but by the second half of the 19th century, it constitutes more than half of the examples (53%) and in 1900–1950 it is 70% of examples of this construction. The recessive form furthermore prospers with non-locative prepositions and non-canonical construction shapes as well; from no examples in 1800–1850, it constitutes 62% in 1850–1900 and 76% in 1900–1950.

The noun *led* ‘ice’ shows some indication of contextual shift in that PREP + ADJ + NOUN constructions tend to have {u} and some periods have fewer of these than others. The real story, however, is the gradual increase in {ě} forms for the canonical *na ledě~ledu* ‘on the ice’. In earlier texts each exponent appears 50% of the time; this rises to 55% for {ě} in 1800–1850, then 88% in 1850–1900 before dropping back to 68% in the first decades of the 20th c. There is some interesting analysis of the contexts to be done here, but it does not explain the shift to {ě}.

We therefore reject the hypothesis above in favor of the null hypothesis: Shifts in the contexts found in Diakon do not explain the shift from {u} to {ě}.

5.5 Analysis

All of the potential confounding factors examined contribute to the contemporary picture of variation in the loc. sg., but none have the potential to explain the maintenance of the recessive ending by themselves and, in some instances, the shift of so many high-frequency nouns to the recessive exponent. The single nugget we can extract here is a move towards {ě} occurring first in certain syntactic contexts, specifically those licencing 'canonical' locativity: the marker of location in space or, metaphorically, its extension to the 'situation' of an action. At this point in Russian, the S-curve 'derails' and one exponent becomes a marker of locativity. In Czech, however, the resolution is not so clear-cut and the exponent marking locativity continues to expand its remit into non-locative contexts.

The only common feature of this group is thus frequency. We return therefore to our original question: could frequency be related to this unexpected shift, and are there any other factors we have neglected to consider?

6 Alternative hypothesis and conclusions

Sociolinguistic factors – including the massive expansion of literacy in Czech during the National Revival of the 19th c. – may play a role in the overall context behind these changes, and may explain some of the particulars in what we see in Czech, but for lack of space we will have to leave those aside for future consideration. We know that the vast bulk of masc. inan. nouns in Czech currently take the expansive ending {u} in the loc. sg.; from this investigation, we can see that high-frequency nouns, many of which had begun with {u} or shifted to it earlier in history make a move back towards {ě} in the 19th c. or later.

The {u} morph appears in this paradigm in four cases of the singular (dat., gen., voc., loc.). In Czech, we observe overall a reduction in the functional load of the noun's case marking – as in German, where the burden of case marking is left to adjectives and other syntactic markers. In the singular of masc. nouns, German only marks the genitive case separately from its remaining three cases. The expansion of Czech {u} can potentially reduce the number of distinct masc. sg. inan. noun case forms to as few as three: one direct case form (nom., acc. sg.) and two indirect case forms ({u} gen., dat., voc., and loc. sg. and {em} instr. sg.). This could be explained as a *Sprachbund* feature attributable to linguistic coexistence in close proximity, which has historically applied across much of the Czech language area.[12]

[12] The tendency to increased syncretism and fewer exponents within Czech declension patterns extends beyond the masc. hard paradigms to the neut. hard paradigm, the old consonant-stem paradigms, and soft paradigms of all genders.

However, as distinct from German, some Czech nouns, many of them high-frequency, can have differentiated case marking on the noun itself – even increasing its use, while many other nouns default to a higher-order construction with a generic oblique ending: [*na*, *v*, *při*, *o*] … NOUN + {u}. This has a *motivation* in that constructions involving individual forms are more likely to be entrenched with repetition. It seems in addition that historically there are higher-level schemas promoting the {ě} morph that can be accessed and assist it in spreading; these concern locativity and canonicity of construction type (i.e., certain sorts of preposition + noun constructions are more 'typically' locative than those with less-canonical locative prepositions, interposed adjectives, etc.).

Users thus seem to have at their disposal multiple conflicting constructions that lead to fluctuations in the use of locative forms and a pattern that varies over time, with forms coming in and out of vogue at different periods. We will need to examine larger historical corpora of Czech to confirm this, but when the current data are compared to the overall data on the loc. sg. in this paradigm, it suggests that high-frequency nouns maintain and even shift towards {ě} more than other nouns in the class, making frequency a further contributing factor in the development of the delineation between these two exponents.

The historical data we have examined here do not fit the pattern of a change that is moving to completion or being diverted onto a single different track, as the changes in the lexicon are spread over many hundreds of years and move in both directions ({u} > {ě} as well as {ě} > {u}). Furthermore, the mix of features that have been proposed as motivators for the choice of one or the other exponent is complex, and we have now potentially added a further one (frequency). This pattern of change that moves slowly and sometimes in retrograde fashion through the system, generating multiple counterexamples and variant usages that persist for centuries, would seem to be characteristic of Czech and other Slavonic languages with high amounts of morphological material. Not all variation thus moves to completion or resolution; some – when balanced in a complex system – can evolve slowly and in various directions at once, a process which Nichols and Timberlake (1991) termed *retextualization*. The endpoint of this change – a shift to the use of the {u} morph – remains visible, but always on the horizon, receding as we draw closer to it.

OPEN DATA STATEMENT

Data for this project will be made available at: the TROLLing repository: https://site.uit.no/trolling/.

REFERENCES

Bermel, Neil (1993): Sémantické rozdíly v tvarech českého lokálu [Semantic differences in forms of the Czech locative]. *Naše řeč*, 76(4), pp. 192–198.

Bermel, Neil (2004): *V korpuse* nebo *v korpusu*? Co nám řekne (a neřekne) ČNK o morfologické variaci v tvarech lokálu [*V korpuse* or *v korpusu*? What the Czech National Corpus can (and can't) tell us about variation in the locative singular]. In: Zdeňka Hladká – Petr Karlík (eds.), *Čeština – univerzália a specifika: 5, Sborník 5. Mezinárodního setkání bohemistů v Brně 13.–15. 11. 2003* [The Czech language – universals and specifics]. Praha: Nakladatelství Lidové noviny, pp. 163–171.

Bermel, Neil (2010): Variace a frekvence variant na příkladu tvrdých neživotných maskulin [Variation and frequency exemplified on hard masculine inanimate nouns]. In: Světla Čmejrková – Jana Hoffmannová – Eva Havlová (eds.), *Užívání a prožívání jazyka: K 90. narozeninám Františka Daneše*. Praha: Karolinum, pp. 135–140.

Bermel, Neil – Knittl, Luděk (2012a): Morphosyntactic variation and syntactic constructions in Czech nominal declension: corpus frequency and native-speaker judgments. *Russian Linguistics* 36(1), pp. 91–119. <http://dx.doi.org/10.1007/s11185-011-9083-x>.

Bermel, Neil – Knittl, Luděk (2012b): Corpus frequency and acceptability judgments: a study of morphosyntactic variants in Czech. *Corpus Linguistics and Linguistic Theory*, 8(2), pp. 241–275. <https://doi.org/10.1515/cllt-2012-0010>.

Bermel, Neil – Knittl, Luděk – Russell, Jean (2015): Morphological variation and sensitivity to frequency of forms among native speakers of Czech. *Russian Linguistics*, 39(3), pp. 283–308. <http://dx.doi.org/10.1007/s11185-015-9149-2>.

Bermel, Neil – Knittl, Luděk – Russell, Jean (2018): Frequency data from corpora partially explain native-speaker ratings and choices in overabundant paradigm cells. *Corpus Linguistics and Linguistic Theory*, 14(2), pp. 197–231. <https://doi.org/10.1515/cllt-2016-0032>.

Blythe, Richard A. – Croft, William (2012): S-curves and the mechanisms of propagation in language change. *Language*, 88(2), pp. 269–304. <http://dx.doi.org/10.1353/lan.2012.0027>.

Bräuer, Herbert (1969): *Slavische Sprachwissenschaft: II, Formenlehre: 1. Teil*. Berlin: Walter de Gruyter.

Brown, Dunstan (2007): Peripheral functions and overdifferentiation: the Russian second locative. *Russian Linguistics*, 31(1), pp. 61–76. <http://dx.doi.org/10.1007/s11185-006-0715-5>.

Bybee, Joan (2006): From usage to grammar: the mind's response to repetition. *Language*, 82(4), pp. 711–733. <http://dx.doi.org/10.1353/lan.2006.0186>.

Croft, William (2000): *Explaining Language Change: An Evolutionary Approach*. London: Longman.

Cvrček, Václav a kol. (2010): *Mluvnice současné češtiny: 1, Jak se píše a jak se mluví* [A Grammar of Czech I]. Praha: Karolinum.

Denison, David (2003): Log(ist)ic and simplistic S-curves. In: Raymond Hickey (ed.), *Motives for Language Change*. Cambridge: Cambridge University Press, pp. 54–70.

DIAKORP: *DIAKORP: Diachronní korpus, verze 6 z 18. 12. 2015* [online] (2011). Praha: Ústav Českého národního korpusu FF UK. Cit. 6. 11. 2023. <http://www.korpus.cz>.

Janda, Laura (1996): *Back from the Brink: A Study of How Relic Forms in Language Serve as Source Material for Analogical Extension*. Munich – Newcastle: Lincom Europa.

Kučera, Karel – Řehořková, Anna – Stluka, Martin (2015): DIAKORP diachronic corpus, version 6 [online]. Praha: Ústav českého národního korpusu FF UK. Cit. 6. 11. 2023. <http://www.korpus.cz>.

Lamprecht, Arnošt – Šlosar, Dušan – Bauer, Jaroslav (1986): *Historická mluvnice češtiny* [A Historical Grammar of Czech]. Praha: Státní pedagogické nakladatelství.

MČ2: Komárek, Miroslav – Kořenský, Jan – Petr, Jan – Veselková, Jarmila (eds.) (1986): *Mluvnice češtiny: 2, Tvarosloví* [A Grammar of Czech: 2, Morphology]. Praha: Academia.
Meillet, Antoine (1965): *Le Slave commun*. Paris: Librairie Honoré Champion.
Nichols, Johanna – Timberlake, Alan (1991): Grammaticalization as retextualization. In: Elizabeth Closs Traugott – Bernd Heine (eds.), *Approaches to Grammaticalization: Vol. I, Focus on Theoretical and Methodological Issues*. Amsterdam – Philadelphia, PA: John Benjamins, pp. 129–146. <http://dx.doi.org/10.1075/tsl.19.1.07nic>.
PMČ: Karlík, Petr – Nekula, Marek – Rusínová, Zdenka (eds.) (1995): *Příruční mluvnice češtiny* [Handbook Grammar of Czech]. Praha: Nakladatelství Lidové noviny.
Štícha, František (2009): Lokál singuláru tvrdých neživotných maskulin (*ve vlaku* vs. *v potoce*): úzus a gramatičnost [The locative singular of hard inanimate masculine nouns (*ve vlaku* vs. *v potoce*): usage and grammaticality]. *Slovo a slovesnost*, 70(3), pp. 193–220.
SyD: Cvrček, Václav – Vondřička, Pavel (2011): *SyD – Korpusový průzkum variant* [online]. Praha: Ústav Českého národního korpusu FF UK. Cit. 6. 11. 2023. <http://syd.korpus.cz>.
SYN2005: *SYN2005: Žánrově vyvážený korpus psané češtiny* [online] (2005). Praha: Ústav Českého národního korpusu FF UK. Cit. 6. 11. 2023. <http://www.korpus.cz>.
Thornton, Anna M. (2012): Reduction and maintenance of overabundance: a case study on Italian verb paradigms. *Word Structure*, 5(2), pp. 183–207. <http://dx.doi.org/10.3366/word.2012.0026>.
Townsend, Charles – Janda, Laura A. (1996): *Common and Comparative Slavic: Phonology and Inflection with Special Emphasis on Russian, Polish, Czech, Serbo-Croatian, and Bulgarian*. Columbus, OH: Slavica.

APPENDIX

Forms and periodization in the Diakon and SYN2005 corpora (in order of frequency of the {u} morph).

Form	Gloss	1300–1800		1801–1850		1851–1900		1901–1950		SYN2005	
areále	campus	0	0.00%	0	0.00%	0	0.00%	0	0.00%	1	0.05%
areálu		0	0.00%	0	0.00%	0	0.00%	0	0.00%	2014	99.95%
způsobě	manner	29	9.60%	12	17.39%	22	8.27%	9	3.61%	2	0.10%
způsobu		273	90.40%	57	82.61%	244	91.73%	240	96.39%	1981	99.90%
pořadě	program	0	0.00%	4	66.67%	3	27.27%	2	14.29%	1	0.10%
pořadu		4	100.00%	2	33.33%	8	72.73%	12	85.71%	1014	99.90%
procese	process, trial	1	33.33%	0	0.00%	0	0.00%	1	3.03%	4	0.15%
procesu		2	66.67%	1	100.00%	2	100.00%	32	96.97%	2655	99.85%
přechodě	crossing, transfer	0	0.00%	0	0.00%	0	0.00%	0	0.00%	2	0.18%
přechodu		0	0.00%	1	100.00%	9	100.00%	20	100.00%	1113	99.82%
úvodě	introduction	1	100.00%	0	0.00%	40	45.45%	58	56.31%	4	0.29%
úvodu		0	0.00%	3	100.00%	48	54.55%	45	43.69%	1380	99.71%
festivale	festival	0	0.00%	0	0.00%	0	0.00%	0	0.00%	5	0.42%
festivalu		0	0.00%	0	0.00%	0	0.00%	0	0.00%	1177	99.58%

provoze	traffic, operation	0	0.00%	0	0.00%	0	0.00%	0	0.00%	19	0.79%
provozu		0	0.00%	0	0.00%	0	0.00%	10	100.00%	2395	99.21%
ústavě	institute	0	0.00%	1	9.09%	18	39.13%	32	69.57%	13	1.10%
ústavu		0	0.00%	10	90.91%	28	60.87%	14	30.43%	1170	98.90%
pádě	case, fall	1	1.05%	1	2.04%	7	3.95%	28	23.93%	32	2.03%
pádu		94	98.95%	48	97.96%	170	96.05%	89	76.07%	1548	97.97%
parlamentě	parliament	0	0.00%	0	0.00%	4	80.00%	57	89.06%	116	7.30%
parlamentu		12	100.00%	5	100.00%	1	20.00%	7	10.94%	1473	92.70%
obvodě	district	0	0.00%	0	0.00%	4	6.90%	10	11.36%	243	18.56%
obvodu		0	0.00%	1	100.00%	54	93.10%	78	88.64%	1076	81.44%
koncertě	concert	0	0.00%	1	33.33%	6	37.50%	5	45.45%	572	49.48%
koncertu		0	0.00%	2	66.67%	10	62.50%	6	54.55%	584	50.52%
úřadě	office	39	36.11%	14	50.00%	61	73.49%	117	86.67%	1039	59.20%
úřadu		69	63.89%	14	50.00%	22	26.51%	18	13.33%	716	40.80%
jazyce	language, tongue	14	7.25%	0	0.00%	15	7.85%	46	21.80%	1427	62.26%
jazyku		179	92.75%	138	100.00%	176	92.15%	165	78.20%	865	37.74%
obchodě	shop, trade	3	8.11%	1	3.70%	9	20.45%	30	54.55%	1529	68.87%
obchodu		34	91.89%	26	96.30%	35	79.55%	25	45.45%	691	31.13%
papíře	paper	10	52.63%	23	69.70%	68	61.82%	114	75.50%	932	72.81%
papíru		9	47.37%	10	30.30%	42	38.18%	37	24.50%	348	27.19%
bodě	point	0	0.00%	0	0.00%	2	25.00%	25	64.10%	980	75.73%
bodu		0	0.00%	1	100.00%	6	75.00%	14	35.90%	314	24.27%
obraze	picture	16	29.09%	22	51.16%	99	83.90%	79	75.24%	1045	78.99%
obrazu		39	70.91%	21	48.84%	19	16.10%	26	24.76%	278	21.01%
závodě	competition, factory	4	66.67%	1	25.00%	11	52.38%	22	78.57%	1421	79.56%
závodu		2	33.33%	3	75.00%	10	47.62%	6	21.43%	365	20.44%
zákoně	law	500	92.59%	15	62.50%	40	42.11%	59	64.13%	1206	80.45%
zákonu		40	7.41%	9	37.50%	55	57.89%	33	35.87%	293	19.55%
dne	day	441	98.66%	149	93.71%	392	92.67%	380	90.05%	1604	81.17%
dnu		6	1.34%	10	6.29%	31	7.33%	42	9.95%	372	18.83%
časopise	magazine	0	0.00%	7	26.92%	238	88.48%	111	79.86%	1077	81.65%
časopisu		0	0.00%	19	73.08%	31	11.52%	28	20.14%	242	18.35%
ledě	ice	6	30.00%	7	50.00%	20	37.74%	40	52.63%	1422	82.29%
ledu		14	70.00%	7	50.00%	33	62.26%	36	47.37%	306	17.71%
státě	state	1	100.00%	1	10.00%	15	17.44%	112	84.85%	2470	87.43%
státu		0	0.00%	9	90.00%	71	82.56%	20	15.15%	355	12.57%
západě	west, turn, sunset	1	3.13%	10	25.00%	130	57.02%	112	57.44%	2162	87.74%
západu		31	96.88%	30	75.00%	98	42.98%	83	42.56%	302	12.26%
dopise	letter	2	100.00%	1	33.33%	12	27.91%	68	80.00%	1623	88.11%
dopisu		0	0.00%	2	66.67%	31	72.09%	17	20.00%	219	11.89%

hlase	voice	8	16.33%	9	56.25%	122	75.78%	87	80.56%	1461	89.52%
hlasu		41	83.67%	7	43.75%	39	24.22%	21	19.44%	171	10.48%
hradě	castle	360	90.23%	129	75.44%	418	85.48%	256	81.79%	1518	92.56%
hradu		39	9.77%	42	24.56%	71	14.52%	57	18.21%	122	7.44%
zápase	struggle, match	0	0.00%	2	8.00%	100	56.32%	73	72.28%	2677	92.63%
zápasu		2	100.00%	23	92.00%	83	43.68%	28	27.72%	213	7.37%
mostě	bridge	24	75.00%	38	100.00%	96	97.96%	115	100.00%	1506	93.31%
mostu		8	25.00%	0	0.00%	2	2.04%	0	0.00%	108	6.69%
klíně	lap	34	85.00%	33	45.21%	333	60.55%	228	77.82%	1025	94.47%
klínu		6	15.00%	40	54.79%	217	39.45%	65	22.18%	60	5.53%
okrese	county	0	0.00%	1	50.00%	32	22.07%	56	44.44%	1609	94.54%
okresu		0	0.00%	1	50.00%	113	77.93%	70	55.56%	93	5.46%
východě	east, exit, sunrise	3	5.77%	13	27.66%	158	71.17%	99	76.15%	2309	94.79%
východu		49	94.23%	34	72.34%	64	28.83%	31	23.85%	127	5.21%
koutě	corner	59	96.72%	59	100.00%	265	96.72%	281	96.23%	1178	96.08%
koutu		2	3.28%	0	0.00%	9	3.28%	11	3.77%	48	3.92%
případě	case	1	100.00%	0	0.00%	217	88.21%	707	92.06%	37059	96.24%
případu		0	0.00%	4	100.00%	29	11.79%	61	7.94%	1447	3.76%
sále	hall	1	50.00%	17	85.00%	104	94.55%	122	96.06%	2216	96.85%
sálu		1	50.00%	3	15.00%	6	5.45%	5	3.94%	72	3.15%
voze	vehicle	91	94.79%	53	89.83%	174	93.05%	236	96.33%	1558	98.11%
vozu		5	5.21%	6	10.17%	13	6.95%	9	3.67%	30	1.89%
základě	basis	9	25.00%	9	45.00%	374	95.90%	397	96.36%	14899	98.45%
základu		27	75.00%	11	55.00%	16	4.10%	15	3.64%	234	1.55%
ostrově	island	99	72.26%	56	90.32%	56	93.33%	202	98.06%	2730	99.06%
ostrovu		38	27.74%	6	9.68%	4	6.67%	4	1.94%	26	0.94%
bytě	flat	0	0.00%	1	3.57%	33	35.11%	140	86.42%	4229	99.23%
bytu		64	100.00%	27	96.43%	61	64.89%	22	13.58%	33	0.77%
hřbitově	cemetery	5	100.00%	55	96.49%	221	95.26%	194	97.98%	997	99.30%
hřbitovu		0	0.00%	2	3.51%	11	4.74%	4	2.02%	7	0.70%
kostele	church	719	99.58%	106	99.07%	387	98.98%	549	99.46%	2210	99.42%
kostelu		3	0.42%	1	0.93%	4	1.02%	3	0.54%	13	0.58%
tábře	camp	25	96.15%	5	45.45%	119	92.25%	204	96.23%	2282	99.43%
táboru		1	3.85%	6	54.55%	10	7.75%	8	3.77%	13	0.57%
venkově	village	0	0.00%	12	100.00%	160	99.38%	318	96.36%	1443	99.45%
venkovu		0	0.00%	0	0.00%	1	0.62%	12	3.64%	8	0.55%
dvoře	courtyard	423	96.80%	156	88.64%	535	96.75%	513	95.71%	2331	99.53%
dvoru		14	3.20%	20	11.36%	18	3.25%	23	4.29%	11	0.47%
obědě	dinner	72	98.63%	22	100.00%	111	98.23%	178	97.27%	1361	99.56%
obědu		1	1.37%	0	0.00%	2	1.77%	5	2.73%	6	0.44%

stole	table	109	89.34%	80	98.77%	307	95.34%	416	97.42%	4138	99.59%
stolu		13	10.66%	1	1.23%	15	4.66%	11	2.58%	17	0.41%
domě	house	846	53.54%	313	96.60%	749	93.86%	915	93.85%	10032	99.60%
domu		734	46.46%	11	3.40%	49	6.14%	60	6.15%	40	0.40%
životě	life	856	87.44%	68	87.18%	576	95.36%	1201	98.60%	14300	99.71%
životu		123	12.56%	10	12.82%	28	4.64%	17	1.40%	41	0.29%
lese	forest	176	94.12%	247	98.02%	1147	99.57%	800	99.50%	2845	99.82%
lesu		11	5.88%	5	1.98%	5	0.43%	4	0.50%	5	0.18%
světě	world	2538	94.49%	799	99.25%	1961	99.09%	2351	99.62%	24189	99.90%
světu		148	5.51%	6	0.75%	18	0.91%	9	0.38%	25	0.10%

Trajektorie ve změnách paradigmat českých substantiv

Tento článek se zabývá známým jevem ve vývoji českého deklinačního systému substantiv: postupným nahrazováním původní *o*-kmenové koncovky v lokálu singuláru koncovkou *u*-kmenovou, které probíhá již tisíciletí. Článek ukazuje, že u velmi frekventovaných substantiv, pro která máme dostatek dat, je proti očekáváním, jež vycházejí z literatury věnované primárně angličtině, často zřejmý opačný trend: *o*-kmenová koncovka proniká k lexémům, u nichž se dříve nevyskytovala. Jelikož neexistuje jediná, převažující příčina, která by mohla způsobovat, že tato překvapivá změna nekopíruje klasický průběh ‚S-křivky', článek navrhuje prověřit, zda pro složitou interakci faktorů a forem, které nacházíme v jazycích s bohatou flektivní morfologií, není vhodnější model retextualizace.

School of Languages and Cultures, University of Sheffield
1 Upper Hanover Street, Jessop West Building
S3 7RA Sheffield, United Kingdom
n.bermel@sheffield.ac.uk
l.knittl@sheffield.ac.uk