# The transfer of former **ja*-stems to a mixed declension in Old Czech and Slovak

2025-07-04
**Matea Haverlík**
s3483568

BA thesis *Comparative IE Linguistics*
Supervised by Dr. Tijmen Pronk
Universiteit Leiden

## Abstract

Czech and Slovak are two closely related West Slavic languages that show an innovative declension type not present in any other Slavic language, traditionally named the *píseň*- and *dlaň*-pattern in Czech and Slovak respectively. The type is populated by nouns derived from Proto-Slavic **ja*- and **i*-stem forms. The forms derived from **ja*-stems have been the subject of description especially in grammars of Old Czech. This article describes the situation in the modern language and extends observations found in descriptions of Old Czech to Slovak on the basis of historical Old Czech and Old Slovak texts.

---

## 1 Introduction

Czech and Slovak are closely related Slavic languages, themselves part of a larger West-Slavic subgroup. The languages show many features of a dialect continuum, though standardization of the languages has pushed them further apart (Nábělková, 2016, 142; 164).

One of the shared features of the languages is the declension pattern exhibited by words such as *pouš*ť – traditionally named the *píseň* and *dlaň* patterns in Czech and Slovak respectively. The declension applies to feminine nouns characterised by a nominative singular with a zero-ending (*pouš*ť*-∅*) after the stem's final consonant – which is always historically soft – and a genitive singular ending in *-ě* (*pouš*ť*-ě*).

Other Slavic languages generally tend to have two major types of declension patterns for feminine nouns with front vowel endings: the thematic **ja*-stems, here named A-type, and the athematic **i*-stems, or B-type.[1] The paradigm displayed by Czech and Slovak *pouš*ť-type nouns does not appear outside of the two.

The three affected declension patterns share the feature of having a soft consonant directly before the inflectional endings, and will be collectively referred to as soft feminine declension patterns in this document.

This paradigm seems to be an innovation by Czech and Slovak, as the words belonging to it can generally be reconstructed to fall within one of two other declension patterns – A or B. Both of these patterns are preserved in Czech and Slovak, there is no sound law present in either language that would cause regular transfer of words into this new declension.

Rather than deriving from a declension that can be traced back to Proto-Slavic, the words' endings seem to show a mixed declension with some endings from type A and some from type B. Because of this, the declension is here referred to as the AB-type pattern. The formal makeup of this declension is pretty clear:

[1] Sometimes additional patterns appear, but remain restricted to specific suffixes, such as the specific declension applied to words in **-yni*.

> The *růže* [A] and *píseň* [AB] declension patterns have identical endings in the whole paradigm except N, A, and V sg. The subtype *kost* [B] is identical with the subtype *píseň* [AB] in the entirity of the sg with the exception of the G; in the pl it has its own forms.[2]
>
> — Petr (1986, 324)

If this is indeed an innovation, the origin of the distribution seems pretty clear. It seems that the incoming original **ja*-stems (usually resulting in modern A-type nouns) would have lost the ending vowel in the nominative and accusative singular, which are already strongly linked together, being direct cases. This would make them formally similar to the B-type nominative and accusative forms inherited from **i*-stems, and the spread of **i*-stem endings into the rarely used vocative – as this tranfer only seems to happen with inanimate nouns (Lamprecht et al., 1986, 158) – is quite natural, due to the connection with the direct cases (Gebauer, 1960, 205-206).

The original **i*-stems suffer from homonymy in the genitive and dative-locative in the singular, with both forms ending in *-i*: G.sg *kost-i*, DL.sg *kost-i*, so it seems plausible that some nouns might take over the A-type genitive in *-e* to clearly mark this function. After that, these nouns might take over the A-type plural endings due to similarity in the oblique cases of the singular and the general tendency of the feminine nouns to have identical endings between G.sg and NAV.pl.

In this paper, I will first present a frequential analysis of the situation in Modern Czech and Slovak concerning the distribution of the lexemes among the various declension types. In the next section I provide an overview on the current literature on this topic, mostly concerning Old Czech. After that I pose two main questions to answer to extend the already extant descriptions and incorporate etymological research into the conversation. I then propose answers to these questions on the basis of over 600 attestations of the relevant words in Old Czech and Slovak texts, and conclude what further steps to take on this topic might be.

| | | **Polish** | | **Old Church Slavonic** | |
|---|---|---|---|---|---|
| | | A: *dusza* | B: *kość* | A: доушда | B: коⷭть |
| sg | N | *-a* | *-∅* | -ѧ | -ь |
| | G | *-y/i* | *-y/i* | -ѫ | -и |
| | DL | *-y/i* | *-y/i* | -и | -и |
| | A | *-ę* | *-∅* | -ѫ | -ь |
| | V | *-o, -u (-∅)* | *-y/i (-e)* | -ѥ | -и |
| | I | *-ą* | *-ą* | -ѣⷨ | -ьⷨ/-иⷨ |
| pl | NAV | *-e* | *-y/i* | -ѫ | -и |
| | G | *-y/i, -∅* | *-y/i* | -ь | -ьи/ии |
| | D | *-om* | *-om* | -ѧмъ | -ьмъ, -ѥмъ |
| | L | *-ach* | *-ach* | -ѧхъ | -ьхъ, -ѥхъ |
| | I | *-ami* | *-ami* | -ѧми | -ьми |

Table 1: The soft feminine declensions patterns in other Slavic languages, exemplified by Polish (Grzegorczykowa et al., 1984, 231-234), closesly related to Czech and Slovak, and Old Church Slavonic (Nandriș, 1959, 55-56; 65-66)

[2] "Shodné koncovky mají jména vz. „růže" a „píseň" v celém paradigmatu mimo N, A a V sg. Podtyp „kost" je shodný s podtypem „píseň" v celém sg s výjimkou G; v pl má své samostatné tvary."

| | | Czech | | | Slovak | | |
|---|---|---|---|---|---|---|---|
| | | A: *duše* | AB: *poušť* | B: *kost* | A: *duša* | AB: *púšť* | B: *kosť* |
| sg | N | *-ě* | *-∅* | *-∅* | *-a* | *-∅* | *-∅* |
| | G | *-ě* | *-ě* | *-i* | *-e* | *-e* | *-i* |
| | DL | *-i* | *-i* | *-i* | *-i* | *-i* | *-i* |
| | A | *-i* | *-∅* | *-∅* | *-u* | *-∅* | *-∅* |
| | V | *-ě* | *-i* | *-i* | | | |
| | I | *-í* | *-í* | *-í* | *-ou* | *-ou* | *-ou* |
| pl | NAV | *-ě* | *-ě* | *-i* | *-e* | *-e* | *-i* |
| | G | *-í, -∅* | *-í* | *-í* | *-í, -∅* | *-í* | *-í* |
| | D | *-ím* | *ím* | *-em* | *-iam* | *-iam* | *-iam* |
| | L | *-ích* | *-ích* | *-ech* | *-iach* | *-iach* | *-iach* |
| | I | *-ěmi* | *-ěmi* | *-mi* | *-ami* | *-ami* | *-ami* |

Table 2: The soft feminine declensions patterns in Czech (Petr, 1986, 288) and Slovak (Letz, 1950, 262-266)

# 2 The synchronic situation in Modern Czech and Slovak

The analogical transfer of A- and B-type nouns to the new AB declension is irregular, and happened differently across the Czech and Slovak territory, becoming more or less fixed after the standardisation of the languages. This resulted in some lexemes which belong to different declensions between the two.

Finding examples of such words is not difficult, but an overall view of the differences between the languages is less trivial to obtain, as I am not aware of any digital databases which sort the languages' words by declension pattern. Such an overview must therefore be obtained by going through a dictionary and manually filtering for feminine soft declension nouns.

## 2.1 Data collection

To collect this data, two dictionaries were consulted, those being a reverse dictionary of Czech and a reverse dictionary of Slovak.

The first decision which needed to be made is what data is relevant to this question. To compare the distribution of lexemes within the three relevant declension patterns, at least all the lexemes which belong to one of these declension patterns in both languages must be collected. To ease the data collection process, the decision to also include words that fit into one of the relevant declensions in only one of the languages was made.

Because the number of words included in either of these dictionaries is very large and the entries all need to be evaluated manually, more restrictive criteria for inclusion needed to be included to make the investigation feasible. The specific criteria differ for each of the dictionaries, because the specific details about the data each dictionary presents are different.

### 2.1.1 *Retrográdny slovník slovenčiny*

Mistrík (1976) lists the declension pattern(s) of each lexeme included in the dictionary. This makes it easy to filter out specifically the words belonging to one of the three soft feminine declensions, referred to as nominal declensions 6, 7, and 8 in the dictionary (Mistrík, 1976, 23).

The dictionary is based on a previous frequency analysis of a Slovak corpus given by an earlier title. However, for completeness, it also includes lexemes not found in this corpus, added from a more

comprehensive dictionary without frequency data. To each of these lexemes they assign the minimum possible frequency value of 1, and do not distinguish them from words that do appear in that corpus (Mistrík, 1976, 21).

This results in the dictionary containing a very large number of especially uncommon words, especially words associated with productive derivational processes in the language, particularly deadjectival nouns in *-ostʹ* (Letz, 1950, 169). The large amount of words with artificially inflated frequency would skew the dataset and be extremely tedious to collect.

As it is impossible to distinguish such added lexemes from lexemes that did occur in the corpus but in the frequency class 1, and the vast majority of the lexemes with frequency 1 are words that did not occur in the dataset (Mistrík, 1976, 22), only lexemes in the frequency class 2 or more were included in the final dataset.

#### 2.1.2 *Retrográdní slovník současné češtiny*

In contrast, Těšitelová et al. (1986) only includes words present in a specific Czech corpus. Unlike the Slovak reverse dictionary, however, it does not include any information about the declension patterns of the lexemes included in the dictionary.

Fortunately, alongside the lemma forms of each included word the dictionary also contains words in the form they appeared in the original corpus, with information about the grammatical gender and case of the forms. Thus, by carefully choosing specific case forms to look for, a smaller section of the dictionary can be searched.

Specifically, the forms chosen were the G.sg and NAV.pl forms. These three forms are always identical within every feminine declension paradigm, distinguish the three target declension classes from other feminine nouns, and appear in texts very commonly,[3] thus still providing a representative sample of words in the language.

This means only the sections of the dictionary listing forms ending in *-e*, *-ě*, and *-i* need to be searched, as that is the only three graphemes in which a soft declension feminine noun in one of these case forms can end. This does mean that certain otherwise applicable lexemes will be excluded if they do not appear in the corpus in one of these forms, but this should only affect the lexemes with the lowest frequency, effectively creating a similar effect as excluding lexemes in the lowest frequency class in the Slovak reverse dictionary.

It is possible that this would create a slight semantic skew which might have an impact on the final data, since the words in different declension classes have minor semantic differences due to for example the aforementioned productive derived words in the *kost*-type declension. There is, however, no good way to correct this without a much lengthier data collection process. Additionally, any such excluded lexeme has a chance of being included in the data regardless, as long as it belongs to a frequency class of at least 2 in the Slovak dataset, so the effects should be minimal.

The N.sg form is also collected from the first section of the dictionary to determine the declension class, along with the lexeme's frequency in the corpus.

### 2.2 Data filtering and processing

Some of the collected data was excluded from the dataset despite superficially passing the inclusion criteria. The reasons for this were various:

- One Czech and nine Slovak lexemes which are *plurale tantum* were excluded, as the difference between AB-type and A-type declensions cannot be distinguished. Also, the fact that these words are never used in the singular likely interferes with the analogical processes being investigated.

[3] In fact, forms generally occur *more* frequently in the genitive than in the nominative (Těšitelová et al., 1986, XIV).

- Some words that passed the inclusion criteria intended to limit the data to soft declension feminine nouns despite not falling in any of these declension classes. Specifically, this includes:
  - two loanwords into Czech of declension class *idea* (*idea*, *Korea*),
  - two Czech proper nouns of declension class *žena* that have a genitive spelled in *-i* due to their stem ending in a soft consonant (*Kája*, *Prievidza*),
  - the Czech word *dítě* 'child', a neuter noun which has an irregular feminine plural,
  - the Slovak neuter noun *šteňa* 'puppy', which is incorrectly listed as an A-type word,
  - the Slovak masculine noun *výmoľ* 'pothole', which is incorrectly listed as an AB-type word.
- The Slovak words *stať* 'passage of text' and *hneď* 'brown colorant', which, while being feminine soft declension nouns, have their occurence frequency vastly inflated by being homonyms of a more frequent verb ('become') and adverb ('immediately'), respectively.
- The Czech word *Země* 'Earth', which is not listed in the first section of the dictionary and thus does not provide a frequency value (as opposed to *země* 'earth, ground', which does).

Because the intention is to investigate the correspondence of words of different declension patterns in the two different languages, the words which need to be paired will not have identical morphology. Because of this, a consistent rule of which words may be paired together and which may not is necessary.

The decision was made to pair words if their morphology matches exactly after excluding all the characteristics of the different declensions, i.e. after removing inflectional endings and possibly a historical **-j-*, even if the exact phonetic forms do not match what would be expected for direct cognates. This means a pair like *studna* : *studňa* would be accepted, despite the difference in whether the final consonant is palatalised (a difference of **-j-*), as would *zářivost* : *žiarivosť* with irregular correspondence, but a pair like *spravedlnost* : *spravodlivosť* would be rejected (different adjectival suffix).

A lexeme only needs to fulfill the inclusion criteria in one language. If one or more candidates valid according to the above rules appear in the other language, they are added to the dataset. Specifically, the words were looked up in the reverse dictionary of the other language, and noted down along with their frequency. As Mistrík (1976) also includes words that do not occur in the corpus but do within the language, the same concept was applied to the Czech data, including it if it is found either in the corpus or in Havránek et al. (1989), again assigning it to the lowest frequency class.

## 2.3 Data and presentation

Finally, the data was sorted into declension classes. For Slovak lexemes, this was done based on the declension pattern numbers provided by Mistrík (1976). For Czech lexemes, the declension pattern was deduced from the endings of the nominative and genitive singular of the entry. The data was then grouped by the pairs of declension patterns in each of the languages.

The grouped data is presented in two ways – the first by amount of qualifying lexemes in each group (Table 3), the second weighted by the frequency with which each lexeme appears (Table 4).

The main part of the table is the 3×3 region in the top left. This includes the investigated correspondence between the soft feminine declension patterns to which lexemes belong in Czech and Slovak.

Data was still included if one of the languages preserves the stem in a different declension pattern. These are presented separated from the soft feminine declensions' rows and columns, separated into words whose grammatical gender is still feminine and those that have moved into another.

The last row and column represents words which appear in the corpus of one language but are completely missing from the dictionary of the other.

There is no data in the bottom right quadrant of the table because these are words that do not fulfill the inclusion criteria in either language.

### 2.3.1 A note on frequency data

Due to the different nature of the two consulted dictionaries, the frequency information from one can not be directly compared to the other. Mistrík (1976, 21) includes frequencies normalized to occurences per million, whereas Těšitelová et al. (1986, XI) includes lexemes' absolute frequency in a corpus of 540,000 words.

Because of this reason the frequency data in the frequency-weighted table cannot correspond to the sum of the frequency in Czech and Slovak. Instead, the geometric mean of the two values is used – this has the added benefit of assigning higher weights to words common in both Czech and Slovak, favoring lexemes with inherited, lexicalised meaning as opposed to lexemes created by productive processes.

The computed geometric means are then summed together within each group. The resulting value is presented rounded to 2 decimal places, but the exact values do not matter, rather the approximate magnitude should be considered.

Comparing the distribution of the absolute amount of lexemes and the relative frequency, there is one main difference, that being the cell representing type A nouns in Czech corresponding to type AB nouns in Slovak. This cell contains a small amount of lexemes but has a disproportionately large total relative frequency. Looking at the given words, one stands out, that being Cz. *země* : Sk. *zem* 'earth, ground'. This is the third most common word in the dataset, accounting for a relative frequency value of 473.31, or about two thirds of 692.51, the value in that cell. This means this word specifically skews the value, since otherwise the cell would have a frequency value more appropriate for the small amount of assigned lexemes.

With the largest exception explained, the frequency and absolute lexeme count data seems to show apprximately the same distribution between the different cases. As there is no large discrepancy between the values, it seems safe to assume that this data represents the entire languages as a whole and that no significant discrepancy was introduced by limiting the data to what was included in the source dictionaries.

### 2.3.2 Interpreting the data

Unsurprisingly, the majority of the data falls on the main diagonal of the table, representing words that belong to the same declension pattern in both Czech and Slovak. This shows us that the transition of words to the new declension is definitely a shared process, and cannot be explained separately for the two languages.

| | Czech | Soft feminine | | | Other feminine | | Other | | |
|---|---|---|---|---|---|---|---|---|---|
| Slovak | | type A | type AB | type B | *žena* | *Kremnica* | *hrad* | *stavení* | - |
| Soft f. | type A | 705 | 19 | 2 | 22 | 4 | 0 | 3 | 25 |
| | type AB | 10 | 105 | 26 | 18 | 0 | 1 | 1 | 8 |
| | type B | 1 | 8 | 489 | 0 | 0 | 1 | 0 | 13 |
| F. | *žena* | 22 | 5 | 1 | | | | | |
| Other | *chlap* | 1 | 1 | 1 | | | | | |
| | *stroj* | 0 | 8 | 0 | | | | | |
| | *mesto* | 0 | 3 | 0 | | | | | |
| | *kanoe* | 1 | 0 | 0 | | | | | |
| | - | 72 | 35 | 47 | | | | | |

Table 3: Correspondences between the declension classes of Czech and Slovak soft declension feminine nouns, as counts of qualifying lexemes

| | Czech | Soft feminine | | | Other feminine | | Other | |
|---|---|---|---|---|---|---|---|---|
| Slovak | | type A | type AB | type B | *žena* | *Kremnica* | *hrad* | *stavení* |
| Soft f. | type A | 6819.64 | 192.18 | 5.46 | 199.28 | 9.15 | 0 | 5.97 |
| | type AB | 692.51 | 1390.6 | 408.6 | 119.23 | 0 | 2.45 | 3.16 |
| | type B | 2 | 142.42 | 6437.21 | 0 | 0 | 16.91 | 0 |
| F. | *žena* | 168.9 | 40.83 | 5 | | | | |
| Other | *chlap* | 1 | 2 | 1.41 | | | | |
| | *stroj* | 0 | 67.86 | 0 | | | | |
| | *mesto* | 0 | 19.49 | 0 | | | | |
| | *kanoe* | 1 | 0 | 0 | | | | |

Table 4: Correspondences between the declension classes of Czech and Slovak feminine nouns, weighted by their relative frequency

There is a substantial amount of lexemes in the four cells corresponding to one language having a *poušť*-type form and the other having a different soft feminine declension type. This represents the analogical transfer of words into the innovative *poušť*-type declension.

A couple other cells indicate a similarly notable amount of lexemes. They are, from most to least, the following:

**Cz. type A : Sk. *žena*** 22 lexemes: mostly loanwords. Also includes *záře* : *žiara* 'glow' and *koupě* : *kúpa* 'purchase'.

**Cz. *žena* : Sk. type A** 22 lexemes: mostly location nouns in *-na* : *-ňa* (e.g. *koupelna* : *kúpeľňa* 'bathroom'). Also includes some personal names with a soft consonant created as shortenings of other personal names.

**Cz. *žena* : Sk. type AB** 18 lexemes: exclusively location nouns in *-na* : *-eň*, except for *bělizna* 'white cloth' and *podobizna* 'likeness, portrait', which contain the same morpheme.

**Cz. type AB : Sk. *stroj*** 8 lexemes: some city names, *křeč* 'cramp', *výstroj* 'equipment', *koupel* 'bath', *růž* 'lipstick, rouge'. These two declension patterns have identical forms in the nominative/accusative and genitive singular, as well as the nominative/accusative plural. It is not likely that the same process occured here, while not preserving the grammatical gender this time.

Perhaps the most interesting are the cells which contain almost no lexemes. Specifically this concerns the cells corresponding to where one of the languages has a type A form and the other has a type B form. These cases only concern three lexemes, two of which two have a type A form in Slovak and a type B in Czech (*past* : *pasca* 'trap', *rez* : *hrdza* 'rust').

In the case of *past*, it is plausible that it would shift from an intermediate type AB form to the type B paradigm because of its ending in *-st* (*-sť-* in many forms), which is very characteristic of *kost*-type nouns.[4]

As for *rez*, the situation in other Slavic languages is clear about the original form likely being **rъďa* (Vasmer, 1955, 520). The reason for the transfer to a type B paradigm is unclear. Etymological dictionaries reconstruct the Proto-Slavic form as a **ja*-stem, but do not address the morphological changes (Holub & Kopečný, 1952, 312; Holub & Lyer, 1967, 420; Machek, 1968, 513). Machek (1968, 312) points out the similarity between the oblique stems of *rez*, *rz-* and *rež*, *rž-* 'rye', being one of the only words beginning in syllabic *r-*. It is possible that this could have prompted *rez* to acquire type B forms on the basis of *rež*.

[4] At least 430 of the 489 lexemes which are type B in both languages are productively derived abstract nouns in *-ost*/*-osť*, 24 others end in *-st*/*-sť*, 14 still in *-t*/*-ť*, and only the remaining 19 have another consonant at the end of the word.

One lexeme has the reverse situation, with Czech having the type A form and Slovak type B (*haluze* : *haluz*). However, both languages in fact also have the alternative form (Cz. *haluz*, Sk. *haluza*). Since both alternatives were included in the dataset, it appears in both cells. Interestingly, this is a word that appears in both **i*-stem and **ja*-stem forms across Slavic (Трубачев, 1979, 95). Perhaps this could be an example where both variations were inherited into the two languages. It does however need to be noted the word could not have had a suffix **-ja* in the shared pre-form, as in that case one would expect the final consonant to yield ***-z-ja* > ***-že/a*.

In general, it seems that the transfer between these declension patterns seems to be mostly one-directional, with influx into the new AB-type declension, but with barely any backflow back into the original declension patterns. The result of this is that inherited nouns of the AB-type in Czech or Slovak may correspond to any of the three types in the other language, but A-type words corresponding to B-type words and vice versa is extremely rare and in some cases might have a secondary explanation.

This means that while a type AB lexeme in Czech or Slovak can go back to either a **ja*-stem or an **i*-stem, type A attestations can still be considered good evidence for a **ja*-stem reconstruction and in the same way type B attestations can support **i*-stem reconstructions.

# 3 The diachronic situation in previous literature

The process of transfer to a different declension class seems to be preceded by a period of variation, where the word may be used in either its original declension pattern or in the new *poušť* type. It seems that we can indeed observe such a phenomenon occuring in Old Czech and Old Slovak sources.

The most obvious context to see this in is in the nominative and accusative singular of supposed original **ja*-stems. Indeed, the fact that in some words the nominative and accusative singular is attested both with zero- and thematic endings was previously noted and briefly discussed in multiple Old Czech and Old Slovak grammars.

Marešová (2008, 43) mentions that some words (such as *kuchyně*//*kuchyň*) still vary between both types in Modern Czech.

The most comprehensive list of lexemes affected by this is provided by Gebauer (1960, 203-204; 207), which gives attestations of these words with forms which differ from either the Modern Czech descendant or from the outcome expected for a **ja*-stem.

The oldest source I can find regarding this topic is Stanislav (1958, 170-171), which describes the situation in Old and Modern Slovak. Other than describing the diachronic process of some **ja*-stems losing endings in the nominative and accusative singular "analogically based on the type *kosť* [B]"[5], he mentions *nádej* (Cz. *naděje*), a lexeme not listed by Gebauer (1960).

Both of these sources also briefly touch on the synchronic situation of these words across different dialects, which often preserve the non-standard form, or sometimes even both forms side by side (Gebauer, 1960, 203-204; Stanislav, 1958, 170-171)

## 3.1 Reasons for morphological change

There seems to be a consensus that the transfer of **ja*-stems into the mixed declension happens due to the merger of various endings caused by the early Czech shift of **a* to *ě* following palatal consonants. Specifically, speakers would prefer the endingless variants because they mark the distinction between the nominative (originally ending in **-ʾa*) and genitive singular and nominative plural (originally ending in **-ʾe*) – the variant *poušť* would be preferred over the original **pouště* < **púšča* due to being clearly

[5] "analógiou podľa typu *kosť*"

marked for the N sg, distinguishing it from the three other case-number pairs with the same ending (*pouště* < **púšče*).

> The clipping of the inconvenient ending *-e* (causing the homonymy of the nominative and genitive sg., and additionally the nominative sg. and nominative-accusative pl.) happened under the influence of i-stems; both declensions were brought together by their identical grammatical gender and palatalising endings.[6]
>
> — Lamprecht et al. (1986, 158)

Under such a narrative it is of course necessary to motivate why we should see examples of this in Old and Modern Slovak, indeed even in cases where Modern Czech does not show the type AB forms. As shown in Section 2, these cases are far from absent in the two languages, with Table 3 showing there are 10 such lexemes, including some fairly common words like *zem* 'ground, earth' (Cz. *země*) or *nádej* 'hope' (Cz. *naděje*).

An explanation for this is provided by Mareš (2001, 72), arguing that this can be sufficiently explained by Slovak simply exhibiting less forms of this type:

> The rise of this new type in Cz is probably due to the difficulties in both declensions: in the paradigm *duše* [A] there is undesirable homophony of N sg. = G sg. = V sg. (*duše*) and G sg. = D & L sg. (*kosti*), which has been removed in the newly created type *dlaň* NA sg. — *dlani* VDL sg. vs. *dlaně* G sg. […] In Sk the rise of this type is justified by the above system difficulties with the paradigm *kosť* [B] only; indeed, in Sk the original **i̯ā*-stems of this type are relatively rare (Cz *věž*, *mez*, *tvrz*, but Sk *veža*, *medza*, *tvŕdza*||*tvŕdz* 'tower, limit, stronghold').
>
> — Mareš (2001, 72)

The claim is motivated by examples where Czech has a type AB form, with Slovak preserving the supposed original A-type. In fact, more such examples could be found (*mříž*, *hráz*…). This, however, glosses over the numerous type AB forms *shared* by both Czech and Slovak – the phenomenon in these languages must be viewed as a single process.

It is not sufficient to assume that in Slovak the only source of modern type AB words were original **i*-stems whereas Czech has both original **i*- and **ja*-stems, unless one assumes that *all* of the shared type AB words in Czech and Slovak (e.g. *báň*, *houšť*/*húšť*, *poušť*/*púšť*, *rohož*, *souš*/*súš*) are originally **i*-stems. This seems to be against current consensus on the etymology of these words.

## 3.2 Type AB words as descended from a separate declension class

Mann (1977, 72-74) claims that rather than analogy with the **i*-stems the words are instead taking on characteristics of a so-called "simple **e*-stem" declension, a type of athematic declension inherited from Proto-Indo-European. This declension would comprise root nouns inherited from the proto-language.

He proposes words like *žeň* 'harvest' or *Velikonoce* 'Easter' would be an example of an original type AB forms inherited from the proto-language (Mann, 1977, 73), possibly originally also including the word *noc* 'night'. These words would then themselves be susceptible to analogy with the **i*-stems, *noc* being preserved as an **i*-stem. The affected **ja*-stem nouns could then acquire new forms by analogy to this separate declension with a genitive in *-e*.

[6] "K odsutí nevýhodného koncového *-e* (způsobujícího homonymii nominativu a genitivu sg., a nadto nominativu sg. a nominativu-akuzativu pl.) zde došlo pod vlivem i-kmenů; obe deklinace sbližoval shodný gramatický rod a měkčící koncovky."

Important to note is that this proposed declension class is not restricted to feminine nouns – the few examples which Mann (1977, 74) connects with related lexemes from other Indo-European languages also include the masculine words *den* 'day' and *tisíc* 'thousand'.

It is somewhat unclear to me what the idea behind listing this as a separate declension is. While the origins of each declension are described, the words seems to be grouped by which declension they fall into synchronically (Mann, 1977, 59-83). This declension seems to be the exception, however. While type AB nouns definitely do form a synchronic declension class, it does not seem to be this one, as the endings listed are exclusively masculine (only listing the I.sg *dnem*, as opposed to the AB-type *-í*) (Mann, 1977, 74). As for the masculine nouns, I am personally unconvinced that these words form a cohesive declension class together rather than being two examples of words with a couple irregular forms, likely belonging to another declension.

If the idea is to posit that these words formed a cohesive declension class at some point in the prehistory of Czech, much more data would be required even to posit such a thing at the Proto-Slavic stage. He admits himself that "no IE language maintains the distinction on the basis of purely mechanistic phonology" (Mann, 1977, 73), and the evidence that he does present regarding other languages is not particularily convincing.

### 3.3 Declension transfer due to phonetic reasons

Some sources discuss the gradual appearance of forms without a vowel in the ending using terms evocative of descriptions of apocope due to sound change. These authors use words like *odpadnutí* lit. 'dropping' (Stanislav, 1958, 170) or *odsutí* lit. 'pushing away' (Gebauer, 1960, 203-204; Lamprecht et al., 1986, 158) or *truncation* (Mann, 1977, 72) to describe these changes, and it is unclear whether they mean to be implying that this change is due to a sound law or rather just describing the formal result.

One such source contrasts the 'dropping' of the vowel in type A nouns against the 'analogical transfer' from type B, but once again what the exact meaning is is not explicitly adressed:

> Nouns of type *dlaň* are of one of two origins: either they were born by dropping of the stem vowel *-a* in certain words of the *ulica* [A] declension (e.g. **zeḿa*) > *zem*, or by analogical transfer of nouns of the *kosť* [B] declension (e.g. *pec – pešťь*) to the aforementioned words which lost their nominative in *-a*.[7]
>
> — Letz (1950, 263)

It seems more reasonable to assume that this is an entirely analogically motivated process, since there seems to be no sound law posited to create such a declension type while not applying to all former **ja*-stems without exception.

Indeed, Mann (1977), who is explicitly arguing for the process to be morphologically rather than phonetically motivated seems to avoid such terminology. In the rest of the article it will be assumed that this is the only kind of process happening here.

## 4 Historical data

Most cited literature addresses the transfer of the original **ja*-stems to the new mixed declension as a matter of course. As these are mostly grammars of Old Czech and Old Slovak, omitting a process of such a large scale would of course result in an incomplete description. Sometimes this is included as a note on nouns of the type A declension, sometimes a separate section is dedicated to the type AB declension.

[7] "Podstatné mená typu *dlaň* sú dvojakého pôvodu: alebo skrsly odpadnutím kmeňového *-a* pri niektorých slovách vzoru „ulica" (napr. **zeḿa* > *zem*), alebo analogickým prechodom podstatých mien vzoru *kosť* (napr. *pec – pešťь*) k uvedeným slovám, čo stratily nominatívne *-a*."

The most detailed descriptions are provided Gebauer (1960) and Mareš (2001), but neither (understandably) goes into much detail about specific lexemes or the exact historical data.

## 4.1 Research goals

There are various phenomena associated with the transfer of words between paradigms. Some arise from being able to watch the general trends unfold over time in the period before the standardization of the languages, some from comparing the difference between usage in the Old Czech and Slovak period when compared to the modern languages.

### 4.1.1 The frequency of AB-type forms in original **ja*-stems

The first matter is that of varying frequency over time. As these words supposedly go back to **ja*-stems, one would expect them to be used with type A endings early in the (pre)history of the language, before the analogical processes happen.

The process can then give rise to observable changes during the Old Czech and Slovak period, whether due to the process still being actively ongoing and productive during the attestation period, both variants being acceptable and some gradually taking over in intra-speaker variation, or different sets of lexemes being affected in different regions, just to name a few possibilities.

According to Gebauer (1960, 203), "The beginnings of this appear already in the old language, over time the amount of these instances increases."[8] This seems to be corroberated by Lamprecht et al. (1986, 158): "Already in the old age is the endingless nominative and accusative singular common. In Modern Czech these nouns are then much more frequent."[9].

The idea here would be to verify whether the available data supports this claim and to check if this is also the case across Old Slovak sources.

### 4.1.2 The etymology of words varying between A-type and AB-type forms

As mentioned in Section 3, Gebauer (1960, 203-204; 207) lists the largest amount of forms of the consulted sources. The forms are discussed in sections 171.2 and 173.2, discussing the aberrant nominative and accusative forms respectively. In total, he lists about 140 entries concerning 32 different lexemes. Additionally, he lists 9 entries belonging to words with the **-yni* suffix, which may show a clipped form in the accusative, though these will not be considered as they are somewhat out of scope for this investigation.

Listing these 32 lexemes as part of the section on **ja*-stems seems to imply that all of them are indeed descended from Proto-Slavic **ja*-stems. However, the etymology of many of these words is not entirely clear. Some of the words listed appear in both type A and type B forms across Slavic languages, implying both forms might be reconstructible for the proto-language, for example *roveň/rovně* < **orvьn-ja*, **orvьn-ь* (Трубачев & Журавлев, 2005a, 223; 228-229). Other words' reflexes seem to differ between languages, such as in the case of *hráz(e)* or *koudel(e)* (Трубачев, 1980, 36-37; 1985, 48-50).

It would not at all be surprising if words with competing forms in the early stages of Czech were particularily susceptible to ending up the mixed AB-paradigm, so the interaction between unclear etymology and variable Old Czech forms should be carefully examined.

## 4.2 Data collection

The data was collected from three sources, compiled into a single table and annotated with attestation dates and case and number markings. More detailed descriptions of the process for each source follow here.

[8] "[z]ačátky toho jsou již v jazyku starém, časem pak případů takových přibývá[.]"

[9] "Už ve staré době bývá častý nominativ a akuzativ singuláru bez koncovky [...]. V češtině nové je pak těchto jmen daleko více."

#### 4.2.1 *Historická mluvnice jazyka českého*

As mentioned in Section 3, Gebauer (1960, 203-204; 207) lists the largest amount of lexemes of the consulted sources. They are discussed in sections 171.2 and 173.2, discussing the aberrant nominative and accusative forms respectively, as they appear in Old Czech. In total, he lists about 140 entries concerning 32 different lexemes. Additionally, he lists 9 entries belonging to words with the **-yni* suffix, which may show a clipped form in the accusative. The time period comprises mostly the 14-16th centuries A.D., with some attestations outside of that.

This is a good dataset in that it is has attestations of a wide variety of lexemes. Unfortunately, it is very curated in that the examples presented only provide a couple attestations of any given lexeme, usually when it occurs in different forms in the various attestations. This means that the data is ill-suited to analysing changes in frequency over time for individual lexemes.

The list of 32 lexemes in Gebauer (1960) will be reused as the main focus of the remainder of the article, and will be referred to multiple times. It is not exhaustive, but while there might of course be more insights to be gained with other data, using the same list everywhere enables comparing results between the different sources.

#### 4.2.2 *Historický slovník slovenského jazyka*

In terms of time range, Blanár et al. (n.d.) offers a more or less complementary view, in that the data included concerns approximately 16-18th century Slovak (except standard) texts. As this is a full dictionary, it has a couple advantages over Gebauer's list.

The first point is the aim of the dictionary. As it is more concerned with presenting a semantic overview of the attested words and with providing a general overview of attested forms, it is not as focus as being exemplary of a specific phenomenon as Gebauer is. Where Gebauer lists examples only in the nominative and accusative singular, Blanár et al. list a much more diverse sample of attestations. Having access to attestations in the G.sg/NA.pl allows for the identification of potentially surprising B-type forms, while seeing attestations of identical surface forms in different cases offers more insight into the context in which these paradigm changes happen (as discussed in Section 3.1).

The second is maybe more obvious: simply the amount of data. Across the 32 lexemes on the list Blanár et al. list 375 attestations, as opposed to Gebauer's 140. If we restrict the counted data to only forms in the N.sg and A.sg, to place the data on even footing with Gebauer, Blanár et al. still provide over 200 such attestations.

Of course, the goal of the dictionary is not to provide a comprehensive list of attestations. The intent is to cover as many lexemes as possible, so there is a practical limit on the amount of attestations listed for any given lexeme (Blanár et al., 1991, 14). Because of this, the collected data will not be completely representative of what an average speaker might have used. This might be both due to the comparatively smaller amount of data that a more detailed investigation into just the relevant words might yield, as well as the bias introduced by only selecting attestations that serve as clear examples of the provided lexicographic information.

As the dictionary does not provide a morphological gloss of the forms presented, all attestations of the affected words must be done manually. Sometimes determining the exact case and number is difficult due to the degree to which the usage of specific constructions (genitives in the direct object, certain prepositional phrases) has changed over the past couple centuries, rendering my native intuition of Modern Slovak not entirely applicable.

In some cases, this is not an issue for this investigation – as an example, the usage of genitives with direct objects especially of collective and uncountable meaning makes distinguishing the homonymic G.sg and A.pl forms very difficult and possibly subjective. Fortunately, as this homonymy applies within

all declension patterns no information that allows their identification is lost in this way. In such cases, the specific case/number annotation does not matter as much, so less care has been put into making sure the annotations there are accurate.

In other cases the interpretation of the specific case intended can have consequences on which declension class the word may belong to. This kind of distinction often carries subtle semantic differences between the forms, appearing mostly when distinguishing type A accusatives from forms of other cases. The first is contrasting A-type A.sg attestations in *-u* and I.sg attestations of any type in *-ú* (the length is not always written), e.g. *pod strazu* 'in custody (I.sg)' or 'into custody (A.sg)'. The other is A-type A.sg attestations in *-i* – influenced by the Czech fronting of vowels after soft consonants – and L.sg attestations of any type in *-i*, e.g. *na zemi* 'on the ground (L.sg)' or 'onto the ground (L.sg)'. In the case of these kinds of attestations an educated guess has to be made, so the exact numbers might differ slightly.

#### 4.2.3 *Chrestomatie k vývoji českého jazyka (13.-18. století)*

In response to the curated nature of the previous two sources, for the third I looked for a source that publishes entire Old Czech texts. This ended up being Porák (1979). This allows for the collection of data to be more or less unbiased, the factors skewing the distribution shifting from the semantic and morphological domain to the stylistic, as the author is more likely to decide which texts to include based on the higher level features of the text.

This format has a couple of quirks that make the data very different from that collected from the other two sources. One factor is intra-text cohesion. As each document generally deals with the same topic for longer passages of text, lexemes are likely to repeat multiple times in each text. This allows for analysis of synchronic variation within such texts. The different frequencies of different lexemes also play a much larger role in such a sample, leading to a very rich sample of data on some and barely any on others.

Because the search is not targeted specifically at entries concerning the given lexemes, it is possible to find relevant data that would otherwise not be considered. For example, during the collection of the data for this project, I noticed an A-type form for the word *rozkoš* 'delight' which only appears in AB-type forms in Modern Czech and Slovak.

The major downside of this manner of data collection is the fact that it has a much lower time efficiency than the other two sources listed. So while this source was consulted specifically to allow for frequency analysis over time, in reality a smaller amount of usable data was collected than from the other sources, specifically from the sections up to and including the 15th century.

### 4.3 Discussion

After the data was collected, the answers to the aforementioned questions were tested on the acquired data. Following is a discussion of the various investigated topics and what was found.

#### 4.3.1 Frequency of AB-type forms over time in Old Czech and Slovak

To investigate the changing frequency of A-type forms over the course of the attestation period the forms first have to be sorted by attestation date. Luckily, the text in which the original attestation appears is marked quite clearly in Gebauer (1960) and Porák (1979), and Blanár et al. (n.d.) gives a specific year directly after each attestation.

The only issue is that some abbreviations presented as source texts of items in Gebauer (1960) are missing from the abbreviation list – these attestations had to unfortunately be excluded from the dataset.

On the basis of this data a chart is produced plotting the total amount of attestations of a given form until the date given by the other axis. This is preferable over a non-cumulative chart because it prevents identical datapoints (e.g. from the same text) from coinciding. Two such cumulative counts are shown,


Figure 1: The cumulative total count of forms with given ending in the N and A.sg by the given date. If the date is not certain, the possible time range is presented, and the data point is sorted as if falling into the middle of the range.

one showing forms with a full vowel ending, which is characteristic of A-type forms, contrasted against forms without such an ending, which reflect one of the other two declension types.

On such a chart we can estimate the approximate frequency of the usage of a given form around a given time by the slope the datapoints take – a higher slope means more datapoints in a short amount of time. This is of course meaningless on its own, as it will correlate heavily with the amount of texts preserved from a given time period – more data preserved total means more data of a specific kind preserved. The frequency can however be compared between the two series of data – in Figure 1 we can see Gebauer (1960)'s data shows a higher frequency of type A attestations in general, but a higher frequency specifically near the end of the 16th century.

We can immediately see a large difference between the two datasets – while Blanár et al. (n.d.) shows an equal amount of both types of attestations, (Gebauer, 1960)'s data has significantly less attestations of AB- or B-type forms. As Gebauer's data mostly concerns data dated to 200 years earlier than Blanár et al., this is consistent with the assertion that the frequency increases over time, discussed in Section 4.1.1.

For a better view of the changing frequency within each source it is useful to normalise the maximum totals of each series to the same value.

In Figure 2 we can see that the increasing frequency of AB- or B-type forms applies even within the data collected from Blanár et al. (n.d.). By around 1600, the slope shown by AB- or B-type forms is only half a steep as that shown by A-type forms. Between 1600 and the middle of the 18th century, the slopes seem to be moving in parallel, and after that the AB- or B-type forms outnumber and catch up to A-type forms.

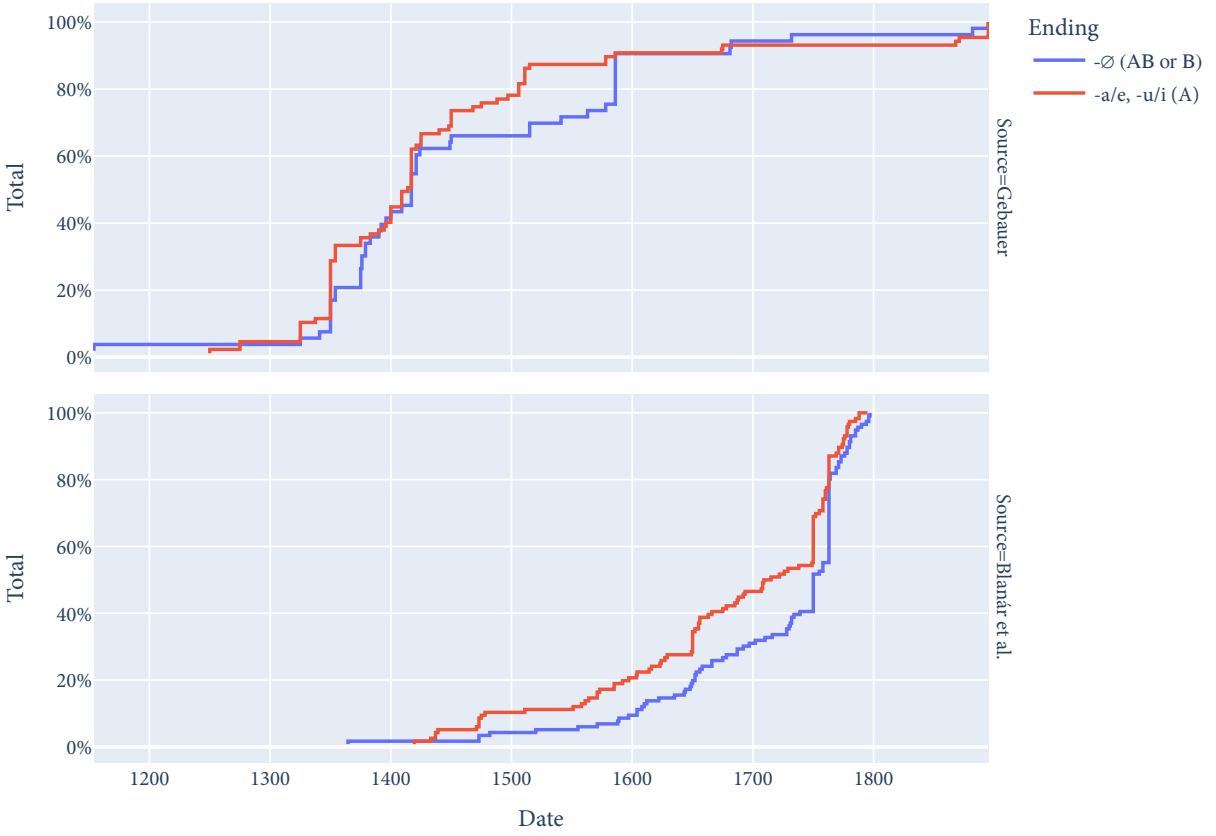


Figure 2: The cumulative fraction of forms with a given ending in the N and A.sg by the given date. Time ranges are no longer shown.

Gebauer (1960)'s data shows a much more uniform slope, though AB- or B-type forms also seem to be less common in the 15th and 16th centuries. This would parallel the situation shown by Blanár et al. (n.d.)'s data, but the lower frequency compared to before this period is strange. Due to the source's heavily curated nature, however, it is not clear if this can be considered significant.

#### 4.3.2 The relationship between attested forms and proposed etymology

While Gebauer (1960)'s list supposedly consists of original **ja*-stems, newer etymological dictionaries contest the accuracy of some of these reconstructions, as the cognates of some across Slavic languages seem to point to **i*-stem reconstructions.

We can also find evidence for **i*-stem forms in attested B-type endings *-i* in the G.sg case. These occur across a range of various words, but (except for *obec*, which is treated separately in Section 4.4.2) occur in the dataset only once or twice per lexeme.

The specific list of lexemes presenting these etymologies is listed here along with the total amount of G.sg forms collected. A brief comment on the etymology of each is provided.

***koudel*** 1/3 attestations show *-i*. Cognates in other Slavic languages show various forms, so it seems difficult to securely decide on the origin. Трубачев et al. (n.d., 12: 48-50) lists **kǫděľь*.

***mez*** 1/1 attestation shows *-i*. As the word is only attested in this form once, not much weight should be placed on this result. Cognates seem to securely pin down the origin as **ja*-stem **meďa* (Трубачев et al., n.d., 18: 45-46)

***obec*** 7/11 attestations show *-i*. This lexeme is treated separately in Section 4.4.2.

***peleš*** 1/1 attestations show *-i*. The single attestation is difficult to base a secure analysis on. Other slavic languages show various forms, some of them not even in soft declensions (Трубачев et al., n.d., 42: 65-66).

***postel*** 1/5 attestations show *-i*. Cognates show both forms (Vasmer, 1955, 416).

***rohož*** 1/2 attestations show *-i*. Other cognates seem to rather suggest a **ja*-stem form, like Russian and Ukrainian *рогожа*, Polish *rogoża*.

***stráž*** 2/4 attestations show *-i*.

***tvrz*** 2/2 attestations show *-i*. These both occur in the genitive of *tvrdz* 'stronghold', which Blanár et al. (n.d., "tvrdz"; "tvŕdz") lists as a borrowing or otherwise influenced from Old Czech, as opposed to *tvŕdza* 'hardship' which is supposedly inherited.

***vrš*** 2/4 attestations show *-i*.

***zem*** 2/29 attestations show *-i*. Seems to generally go back to a **ja*-stem, though there are some notable exceptions discussed in Section 4.4.1 (Vasmer, n.d., 1: 452-453).

In general, it seems that a lot of the words which show variation between type A and type B forms do not admit secure reconstructions at the Proto-Slavic level. A couple possible reasons might be considered: most straightforwardly, these words might just not descend regularly from a shared Proto-Slavic ancestor, only coming into use later and possibly spreading across the Slavic world horizontally. An alternativele is that our current understanding of the nominal system in Proto-Slavic is missing an essential detail that would help explain these forms.

## 4.4 Observations on individual lexemes

While the data from Porák (1979) is not very suitable for the observation of general trends due to the relatively small amount of data collected. However, it does contribute a decent amount of forms to the most common lexemes, specifically *země* 'ground, earth' and *obec* 'community'.

Because of this, more specific observations can be made about these lexemes, so an entire subsection is dedicated to each.

### 4.4.1 The contrastive use of *zem(ě)* in the locative and accusative

As already mentioned in Section 2.3.1, the word *zem(ě)* is very common, and by far the most frequently attested in the collected dataset. It is frequently used with the prepositions *v* 'in' and *na* 'on'. These prepositions are characterised by the contrasting semantics depending on what case is used in their complement – the locative is used to denote stative meaning – 'in' or 'on' – while the accusative is used for change of location – 'into' or 'onto'.

As mentioned in Section 4.2.2, while the semantics remain subtly different, the surface form between the two cases is leveled by the fronting of vowels after palatalised consonants – the original accusative ending **-ǫ* > *-u* is fronted to *-i*, which is identical to the locative ending of the soft declensions.

My initial hypothesis was that the high frequency of these constructions (affecting 32 of the collected 108 attestations of *zem(ě)* in the singular) might be a factor in the acceleration of the shift to forms of the AB-declension. However, that does not really match with what we actually observe – Czech, which has vowel fronting, preserves *země* as an A-type form *with* homonymy, while Slovak, which should not have any need to disambiguate the formally different *-u* and *-i*, has turned the lexeme into the AB-type *zem*.

So then, the opposite hypothesis might be entertained – that the similar forms could actually be a factor acting *against* the shift to another declension. Under this lens, there are a few other pieces of evidence that might be able to support it.

The first is that the accusative form *v zem(i)* 'into the ground' seems to be much less common in Old Slovak (1 occurence out of 3 with *v*; 61 total) than in Old Czech (11 occurences out of 15 with *v*; 63 total), though it is unclear whether this is enough data to put much weight on this, given that Porák (1979)'s

data is internally correlated. Old Slovak instead seems to use the preposition *do* 'into' for this instead (5 occurences, as opposed to 2 in Old Czech).

Secondly, while it is true that Czech preserves the A-type *země*, Mann (1977, 72) points out that the AB-type form *zem* does also appear in Czech as the only acceptable option specifically in the phrase *na zem* 'onto the ground'. A narrative where the semantic difference between the accusative and locative versions of *v zemi* is larger than the one between *na zem(i)* could be suggested. If semantics pushed the two forms of *na zem(i)* apart but held *v zemi* closer together, we could explain why Czech would use a form here that is rarely used otherwise.

Anecdotally, I personally had a much easier time distinguishing between the different uses of *na* than *v* while annotating – but I currently have no other evidence of this claim.

It also has to be said that Czech and Slovak are not the only languages showing unexpected forms for the word descended from **zemľa* – Old Church Slavonic has an attested alternative L.sg form **земи**, where one would expect **земл҃и** for a regular **ja*-stem (Birnbaum, n.d., 014v.8). This form looks as if reflecting an old **i*-stem locative. Russian supposedly shows some fossilised accusatives in *-ь* (Vasmer, n.d., 1: 452-453). Perhaps B-type forms for this word might go back further than the phenomenon of the AB-type in Czech and Slovak?

With all of these options, it seems difficult to say which variant is most likely. It is also very possible that my attempts to explain the difference are a case of overfitting of data. Ideally, some of these observations should be verified on independent data to check whether these are real patterns or just artifacts of randomness.

#### 4.4.2 Differing tendencies in the use of *obec*

The word *obec* seems to be following a trend completely different from the other treated words. It does have some things in common – indeed, it does occasionally appear with A-type forms in the N and A.sg, but the degree to which this is the case is much lower than the other lexemes, with A-type forms accounting for only 4 of the 31 datapoints.

Of this already low amount, only one concerns the N.sg form, that being an attestation from 1688 in Zavar, a village near Trnava in Slovakia. Gebauer (1960, 203) states the form is not attested at all in Old Czech.

The other forms of the word are even more damning – 7 out of 11 attestations in the G.sg show a form ending in *-i*, which is a form characteristic of the type B declension.

It seems that despite having a small number of A-type attestations and belonging to type AB in the modern languages the status of *obec* as an original **ja*-stem is not secure.

Consulting an etymological dictionary only serves to discredit the **ja*-stem option even more – while (Трубачев & Журавлев, 2005b, 161-162) reconstructs a variable **obьtja/*obьtjь*, the reflexes of this word seem to be restricted to West Slavic (other than a Russian dialectal *обча*), and even within the subgroup the **ja*-stem reflexes only appear in Czech and Slovak.

The rest of the Slavic languages have the adjective **obьtjь(jь)* 'common', further suggesting that the **-j-* is part of the base of the word and not just the declension stem (Трубачев & Журавлев, 2005b, 166). More discussion on this in Section 4.5.

Based on this, it seems likely the A-type forms are secondary in Czech and Slovak, possibly appearing after the lexeme entered the AB-type paradigm as an original **i*-stem.

## 4.5 Modern evidence for declension transfer

In the case of some AB-type words we would have been able to deduce they used to be **ja*-stems even without the presence of attested diachronic change. This is because they can show consonant mutation

caused by the devocalisation of suffixed short *-*i*- before the thematic *-*a*: **rъd-ja* > **rъďa* > *rzě* → *rez* (as opposed to ***rъd-ъ* > ***reď*); **pust-ja* > **pušča* > *púšťě* → *púšť* (as opposed to ***pust-ъ* > ***púsť*, cf. *pustina*).

The same outcome could be achieved with an **i*-stem pre-form if the above words contained a **j* before the declension stem: ***pustj-ъ* (***pušč-ъ*) could still give *poušť*. Sometimes this can be rejected if the underlying form is known from related words – *pustý* < **pust-ъjь* 'deserted', *rudý* < **rud-ъjь* 'red'.

The opposite situation is present in *obec* 'community', where the related adjective is already soft – **obьtj-ь(jь)* 'common' makes an **i*-stem interpretation relatively more likely, as **obьtj-ь* is sufficient to produce the expected form even without the need for a **ja*-stem to cause consonant mutation.

# 5 Conclusion

With this investigation I aimed to provide a more detailed treatment and verify the claims made about the history of **ja*-stem nouns that end up in the AB-type mixed declension, which seemed to be lacking in the literature despite widespread awareness of this phenomenon.

Based on the data gathered, it seems that while the claims on the increasing frequency of type AB forms across the history of Old Czech can be confirmed and even extended to Old Slovak, the assumption that all of the words showing A-type endings in the N and A.sg reflect former **ja*-stems is at best dubious.

There is, of course, still much more to the topic. Firstly, all of these analyses were done in interpretation after already having assembled a dataset, so – aside from verifying Gebauer (1960)'s observations in Section 4.3.1 – there is risk of overjustifying results with the quirks of the specific data collected. Ideally the statements in Section 4.4 would be checked against an independent dataset in the future.

Additionally, for both of the lexemes for which the data for a deeper analysis was available the story ended up being much more complex than a general overview could offer. Chances are that this would be possible for many of the words treated in Section 4.3.2, as long as enough high quality data can be obtained.

Lastly, an analysis of the geographic distribution of forms belonging to the different declensions was intended but ended up being infeasible due to both time constraints and the already broad scope covered in this article. As Modern Czech and Slovak display distinct semantic and morphological patterns covered in Section 2.3.2, a separate investigation into distributions of forms in pre-standardization language could be fruitful.

# Bibliography


Birnbaum, D. J. *suprasliensis.obdurodon.org*. Retrieved July 3, 2025, from http://suprasliensis.obdurodon.org/pages/supr014v.html

Blanár, V., Giger, M., Hríbiková, A., Jóna, E., Kotulič, I., Kováčová, M., Krasnovská, E., Kuchar, R., Laliková, T., Majtán, M., Majtánová, M., Ondrejková, R., Ondrejkovičová, D., Osadská, A., Peciar, Š., Ricziová, B., Szabová, A., Sitárová, M., Skladaná, J., & Valentová, I. *Historický slovník slovenského jazyka* (M. Majtán, R. Kuchar, & J. Skladaná, Eds.). Veda. Retrieved July 3, 2025, from https://slovnik.juls.savba.sk/?d=hssj

Blanár, V., Jóna, E., Kotulič, I., Krasnovská, E., Kuchar, R., Majtán, M., Majtánová, M., Peciar, Š., Riczová, B., & Skladaná, J. (1991). *Historický slovník slovenského jazyka* (M. Majtán, V. Blanár, I. Kotulič, & Š. Peciar, Eds.; Vol. 1). Veda.

Gebauer, J. (1960). *Historická mluvnice jazyka českého, Tvarosloví. III, 1: Skloňování*. Academia.

Grzegorczykowa, R., Laskowski, R., & Wróbel, H. (Eds.). (1984). *Gramatyka współczesnego języka polskiego: Morfologia* (Wydanie pierwsze). Państwowe Wydawnictwo Naukowe.

Havránek, B., Bělič, J., Helcl, M., Jedlička, A., Křístek, V., & Trávníček, F. (1989). *Slovník spisovného jazyka českého* (H. Pešatová & P. Čech, Eds.; 2., nezměněné vydání). Academia. https://ssjc.ujc.cas.cz/

Holub, J., & Kopečný, F. (1952). *Etymologický slovník jazyka českého*. Státní nakladatelství učebnic.

Holub, J., & Lyer, S. (1967). *Stručný etymologický slovník jazyka českého: se zvláštník zřetelem k slovům kulturním a cizím*. Státní pedagogické nakladatelství.

Lamprecht, A., Šlosar, D., & Bauer, J. (1986). *Historická mluvnice Češtiny*. Státní pedagogické nakladatelství.

Letz, B. (1950). *Gramatika slovenského jazyka*. Štátne nakladateľstvo.

Machek, V. (1968). *Etymologický slovník jazyka českého* (Druhé, opravené a doplněné vydání). Academia.

Mann, S. E. (1977). *Czech Historical Grammar* (Revised Edition). Buske.

Mareš, F. V. (2001). The Historic Development of the Slavic Noun Declension I (The System of Categories). In F. V. Mareš, *Diachronische Morphologie des Ur- und Frühslavischen*. Europäischer Verlag der Wissenschaften.

Marešová, H. (2008). *Základy historické mluvnice češtiny (s texty k rozboru)* (1. vydání). Vydavatelství Univerzity Palackého v Olomouci.

Mistrík, J. (1976). *Retrográdny slovník slovenčiny* (E. Pauliny, Ed.). Univerzita Komenského.

Nandriș, G. (1959). *Handbook of Old Church Slavonic. Part I: Old Church Slavonic Grammar*. The Athlone Press.

Nábělková, M. (2016). The Czech-Slovak Communicative and Dialect Continuum: With and Without a Border. In T. Kamussela, M. Nomachi, & C. Gibson (Eds.), *The Palgrave Handbook of Slavic Languages, Identities and Borders*. Palgrave Macmillan. https://doi.org/10.1007/978-1-137-34839-5_8

Petr, J. (1986). *Mluvnice češtiny (2): Tvarosloví* (M. Komárek, J. Kořenský, J. Petr, & J. Veselková, Eds.). Academia.

Porák, J. (1979). *Chrestomatie k vývoji českého jazyka (13.-18. století)* (1. vydání). Státní pedagogické nakladatelství.

Stanislav, J. (1958). *Dejiny slovenského jazyka, Tvaroslovie* (Vol. 2). Vydavateľstvo Slovenskej akadémie vied.

Těšitelová, M., Petr, J., & Králík, J. (1986). *Retrográdní slovník současné češtiny* (J. Horecký, Ed.). Academia.

Vasmer, M. *Russisches etymologisches Wörterbuch*. Carl Winter Universitätsverlag.

Vasmer, M. (1955). *Russisches etymologisches Wörterbuch* (Vol. 2). Carl Winter Universitätsverlag.

Трубачев, О. Н. (Ed.). (1979). *Этимологический словарь славянских языков: праславянский лексический фонд* (Vol. 6). Наука.

Трубачев, О. Н. (Ed.). (1980). *Этимологический словарь славянских языков: праславянский лексический фонд* (Vol. 7). Наука.

Трубачев, О. Н. (Ed.). (1985). *Этимологический словарь славянских языков: праславянский лексический фонд* (Vol. 12). Наука.

Трубачев, О. Н., & Журавлев, А. Ф. (Eds.). (2005a). *Этимологический словарь славянских языков: праславянский лексический фонд* (Vol. 32). Наука.

Трубачев, О. Н., & Журавлев, А. Ф. (Eds.). (2005b). *Этимологический словарь славянских языков: праславянский лексический фонд* (Vol. 31). Наука.

Трубачев, О. Н., Журавлев, А. Ф., & Варбот, Ж. Ж. (Eds.). *Этимологический словарь славянских языков: праславянский лексический фонд*. Наука.

## Appendix

Data collected for the purposes of this study as well as interactive versions of Figure 1 and Figure 2 can be found online at https://github.com/mm-tea/mixed-declension.