

Productive phonological and morphological generalizations in Czech inflectional morphology

Guy Tabachnick

gtabach.github.io

Faculty of Arts, Masaryk University

April 20, 2026

Outline

Introduction

Corpus study

Nonce word study 1

Nonce word study 2

Conclusion

Morphological productivity

One way to probe speakers' mental knowledge about their language is by testing their *morphological productivity*

- If speakers need to produce an inflected form of a word they know, they have two options:
 1. generate it procedurally
 2. access it from storage
- If speakers need to produce an inflected form of a word they *don't* know, only the first option is available
- Asking speakers to inflect made-up words (*nonce words* or *wugs*) is thus a good way to find out their productive patterns (Berko, 1958)
- Their behavior is often *variable*, reflecting knowledge of gradient statistical patterns (e.g. Hayes et al., 2009)

Morphological productivity

One way to probe speakers' mental knowledge about their language is by testing their *morphological productivity*

- If speakers need to produce an inflected form of a word they know, they have two options:
 1. **generate it procedurally**
 2. ~~access it from storage~~
- If speakers need to produce an inflected form of a word they *don't* know, only the first option is available
- Asking speakers to inflect made-up words (*nonce words* or *wugs*) is thus a good way to find out their productive patterns (Berko, 1958)
- Their behavior is often *variable*, reflecting knowledge of gradient statistical patterns (e.g. Hayes et al., 2009)

Productive patterns

What information do speakers use in figuring out how to make a new inflected form?

- **Phonology:** Based on the behavior of words with similar shape (cf. Albright & Hayes, 2003; Becker et al., 2011; Ernestus & Baayen, 2003; Gouskova et al., 2015; Hayes et al., 2009)
- **Morphology:** Inferred from patterns implied by the word's other inflected forms (cf. Ackerman & Malouf, 2013; Ackerman et al., 2009; Copot & Bonami, 2023; Finkel & Stump, 2007; Tabachnick, 2024; Wurzel, 1989)
- **Other factors:** Gender, semantics, etc.

Productive patterns

What information do speakers use in figuring out how to make a new inflected form?

- **Phonology:** Based on the behavior of words with similar shape (cf. Albright & Hayes, 2003; Becker et al., 2011; Ernestus & Baayen, 2003; Gouskova et al., 2015; Hayes et al., 2009)
- **Morphology:** Inferred from patterns implied by the word's other inflected forms (cf. Ackerman & Malouf, 2013; Ackerman et al., 2009; Copot & Bonami, 2023; Finkel & Stump, 2007; Tabachnick, 2024; Wurzel, 1989)
- **Other factors:** Gender, semantics, etc.

I will present two nonce word studies showing that Czech speakers apply gradient phonological (well-established!) and morphological (fairly novel!) generalizations in producing novel inflected forms

Outline

Introduction

Corpus study

Nonce word study 1

Nonce word study 2

Conclusion

Czech locative

Czech has numerous inflection classes that interact with gender and phonology – focus on one narrow corner: *masculine inanimate hard-stem nouns*.

	‘team’	‘time’	‘football’
nominative	tým	čas	fotbal
genitive	týmu	času	fotbalu
dative	týmu	času	fotbalu
accusative	tým	čas	fotbal
vocative	týme	čase	fotbale
locative	týmu	čase	fotbalu~e
instrumental	týmem	časem	fotbalem

Czech locative

Czech has numerous inflection classes that interact with **gender** and phonology – focus on one narrow corner: ***masculine inanimate hard-stem nouns***.

	‘team’	‘time’	‘football’
nominative	tým	čas	fotbal
genitive	týmu	času	fotbalu
dative	týmu	času	fotbalu
accusative	tým	čas	fotbal
vocative	týme	čase	fotbale
locative	týmu	čase	fotbalu~e
instrumental	týmem	časem	fotbalem

Czech locative

Czech has numerous inflection classes that interact with gender and **phonology** – focus on one narrow corner: *masculine inanimate hard-stem nouns*.

	‘team’	‘time’	‘football’
nominative	tým	čas	fotbal
genitive	týmu	času	fotbalu
dative	týmu	času	fotbalu
accusative	tým	čas	fotbal
vocative	týme	čase	fotbale
locative	týmu	čase	fotbalu~e
instrumental	týmem	časem	fotbalem

Czech locative

Czech has numerous inflection classes that interact with gender and phonology – focus on one narrow corner: *masculine inanimate hard-stem nouns*.

	‘team’	‘time’	‘football’
nominative	tým	čas	fotbal
genitive	týmu	času	fotbalu
dative	týmu	času	fotbalu
accusative	tým	čas	fotbal
vocative	týme	čase	fotbale
locative	týmu	čase	fotbalu~e
instrumental	týmem	časem	fotbalem

Czech locative

Czech has numerous inflection classes that interact with gender and phonology – focus on one narrow corner: *masculine inanimate hard-stem nouns*.

	‘team’	‘time’	‘football’
nominative	tým	čas	fotbal
genitive	týmu	času	fotbalu
dative	týmu	času	fotbalu
accusative	tým	čas	fotbal
vocative	týme	čase	fotbale
locative	týmu	čase	fotbalu~e
instrumental	týmem	časem	fotbalem

Czech locative

Czech has numerous inflection classes that interact with gender and phonology – focus on one narrow corner: *masculine inanimate hard-stem nouns*.

	‘team’	‘time’	‘football’
nominative	tým	čas	fotbal
genitive	týmu	času	fotbalu
dative	týmu	času	fotbalu
accusative	tým	čas	fotbal
vocative	týme	čase	fotbale
locative	týmu	čase	fotbalu~e
instrumental	týmem	časem	fotbalem

Allomorphy in the locative: **u** vs. **e**, sometimes variably.

(**e** triggers some consonant alternations and is sometimes written as **ě**)

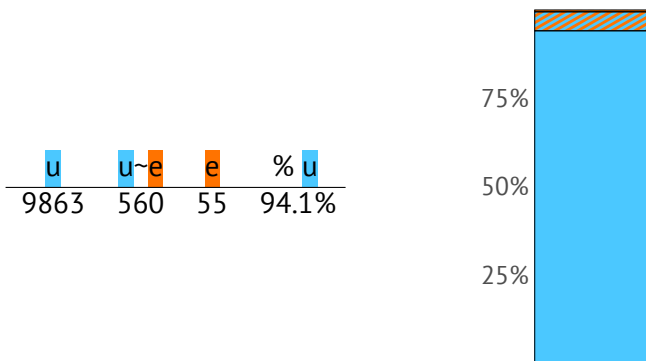
Basic design

First step: establish the patterns in the lexicon to which speakers are exposed

- SYNv11: very large corpus of mostly recent written texts in the Czech National Corpus (Křen et al., 2022)
- Searched for nouns after prepositions (ask me about this if curious) in the genitive and locative (we'll come back to this)
- Common nouns (e.g. no uppercase, numbers, or odd spellings/stem changes) appearing at least 5 times total and at least once in each case
- Nouns counted as variable if they appeared at least 1% of the time with either suffix
- Total: 10,478 nouns

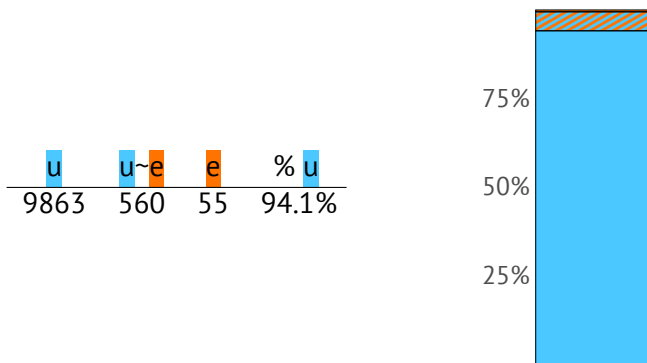
Basic distribution

A large majority of nouns take the (default) **u** in the locative



Basic distribution

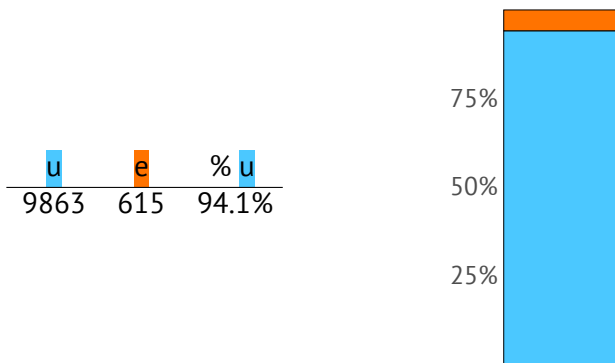
A large majority of nouns take the (default) **u** in the locative



For practical (and empirically justifiable) purposes: lumping into *always* **u** and *sometimes* **e**

Basic distribution

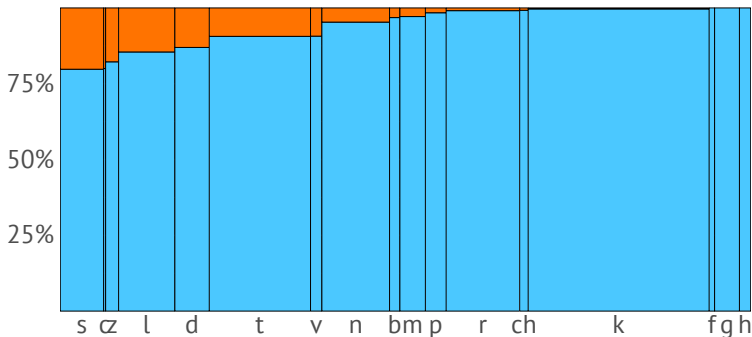
A large majority of nouns take the (default) **u** in the locative



For practical (and empirically justifiable) purposes: lumping into *always* **u** and *sometimes* **e**

Phonology

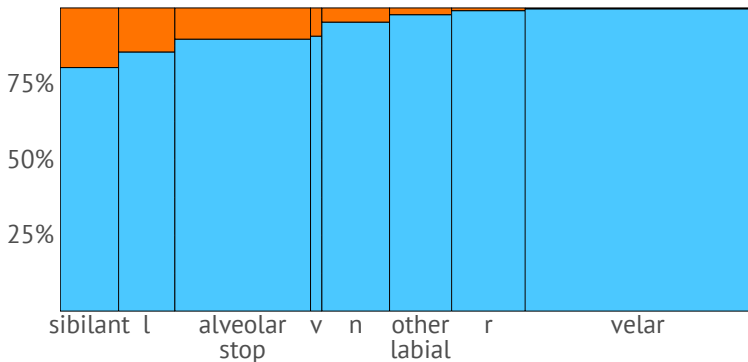
Primary effect (as is common): a noun's locative is partially predicted by its final consonant.



Other significant factors: final syllable vowel (height, backness, length), monosyllabicity, word-final cluster

Phonology

Primary effect (as is common): a noun's locative is partially predicted by its final consonant.



Other significant factors: final syllable vowel (height, backness, length), monosyllabicity, word-final cluster

Morphology

We see the same variation in the genitive as the locative, and all combinations of the two cases are possible:

	‘team’	‘evening’	‘time’	‘forest’	‘back of head’
nominative	tým	večer	čas	les	týl
genitive	týmu	večera	času	lesa	týlu~a
dative	týmu	večeru	času	lesu	týlu
accusative	tým	večer	čas	les	týl
vocative	týme	večere	čase	lese	týle
locative	týmu	večeru	čase	lese	týlu~e
instrumental	týmem	večerem	časem	lesem	týlem

Morphology

There is a correlation: when one case has the non-**u** suffix, the other is more likely to as well.



Morphology

There is a correlation: when one case has the non-**u** suffix, the other is more likely to as well.



Outline

Introduction

Corpus study

Nonce word study 1

Nonce word study 2

Conclusion

Frequency matching

Previous studies have found that adult speakers (generally) respond to nonce word tasks by “frequency matching”:

“*Law of frequency matching*: Speakers of languages with variable lexical patterns respond stochastically when tested on such patterns. Their responses aggregately match the lexical frequencies.” (Hayes et al., 2009, p. 826)

- Speakers extend gradient patterns in their lexicon to new words
- Thus, speakers must be *learning* and *storing* some amount of lexical statistics (type frequency, at least)

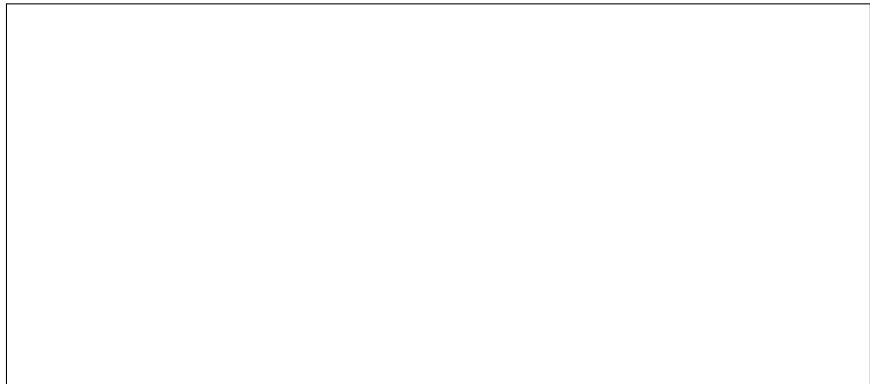
Types of lexical patterns

What sort of lexical statistics do speakers learn and apply?

- Various *phonological* properties, like the consonant adjacent to the affix (cf. Albright & Hayes, 2003; Ernestus & Baayen, 2003; Hayes & Londe, 2006)
- ... But not all – conflicting results on whether speakers learn phonologically “unnatural” generalizations, like non-local properties (cf. Becker et al., 2011; Gouskova et al., 2015; Hayes et al., 2009)
- Although it is known that relatively strong *morphological* correlations (like that between the genitive and locative) exist (cf. Ackerman & Malouf, 2013; Ackerman et al., 2009; Wurzel, 1989), only recently have we started to collect evidence that speakers learn and use them (Copot & Bonami, 2023; Tabachnick, 2024)

Study design

I extended the usual wug test: participants were shown *two forms* (nominative and genitive) and had to select a third (locative).



Study design

I extended the usual wug test: participants were shown *two forms* (nominative and genitive) and had to select a third (locative).

In our area we usually grill špod or products made ze špodu.

Please select the word's genitive form: [ze špoda / ze špodu]

Show word in sentence

Study design

I extended the usual wug test: participants were shown *two forms* (nominative and genitive) and had to select a third (locative).

In our area we usually grill špod or products made ze špodu.

*Please select the word's genitive form: [ze-špoda / **ze špodu**]*

Make sure participants are paying attention

Study design

I extended the usual wug test: participants were shown *two forms* (nominative and genitive) and had to select a third (locative).

In our area we usually grill špod or products made ze špodu.

*Please select the word's genitive form: [ze-špoda / **ze špodu**]*

That's correct! Now select the word in the appropriately inflected form according to you.

After removing the drill [z špoda / ze špodu / z špodu / ze špoda], a clean, circular hole will remain [v špodě / ve špodě / ve špodu / v špodu].

Show word in new sentence, ask for genitive again and locative

Study design

I extended the usual wug test: participants were shown *two forms* (nominative and genitive) and had to select a third (locative).

In our area we usually grill špod or products made ze špoda.

*Please select the word's genitive form: [**ze špoda** / ze-špødu]*

That's correct! Now select the word in the appropriately inflected form according to you.

After removing the drill [z špoda / ze špodu / z špodu / ze špoda], a clean, circular hole will remain [v špodě / ve špodě / ve špodu / v špodu].

Some trials show the other genitive

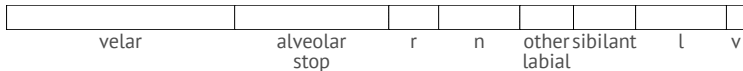
Stimuli: phonology

82 nonce words selected from an initial set of 427 generated stimuli to roughly match the distribution of phonological properties in the lexicon

types in lexicon



trials in experiment



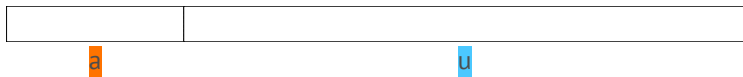
Trials: morphology

Each speaker saw 50 trials, of which 12 had genitive **a** (too many would be very unnatural)

types in lexicon



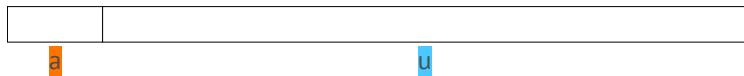
trials in experiment



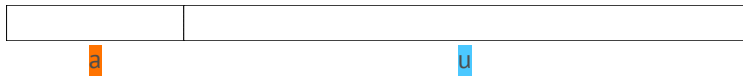
Trials: morphology

Each speaker saw 50 trials, of which 12 had genitive **a** (too many would be very unnatural, but words with genitive **a** are common)

tokens in lexicon



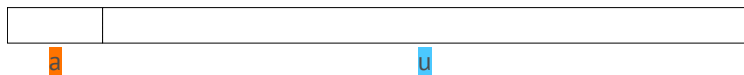
trials in experiment



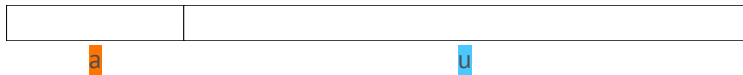
Trials: morphology

Each speaker saw 50 trials, of which 12 had genitive **a** (too many would be very unnatural, but words with genitive **a** are common)

tokens in lexicon

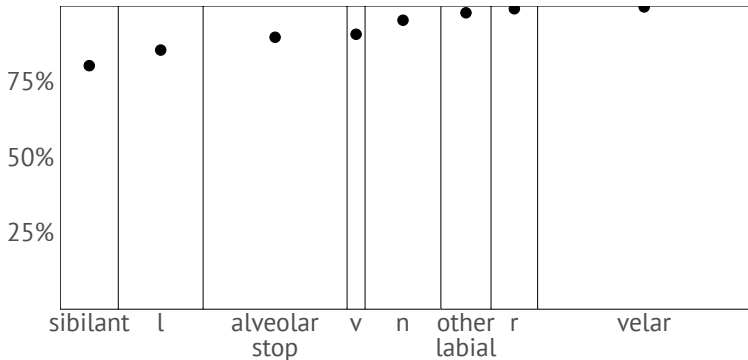


trials in experiment

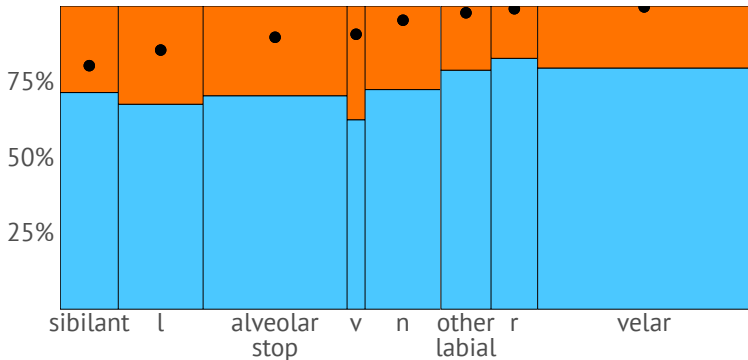


- 90 Czech speakers recruited on Prolific
- 4 discarded for demographic reasons or poor performance
- Total: 4297 trials

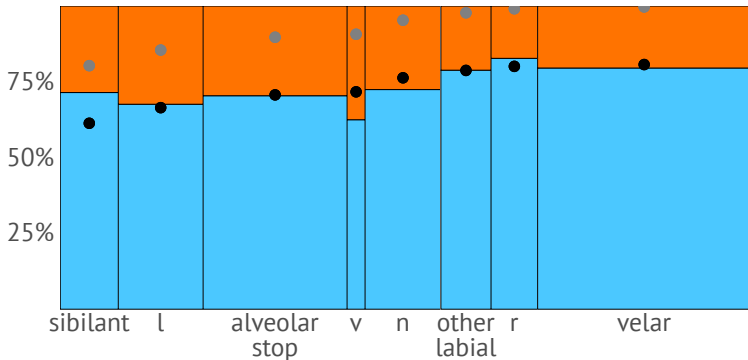
Results: phonology



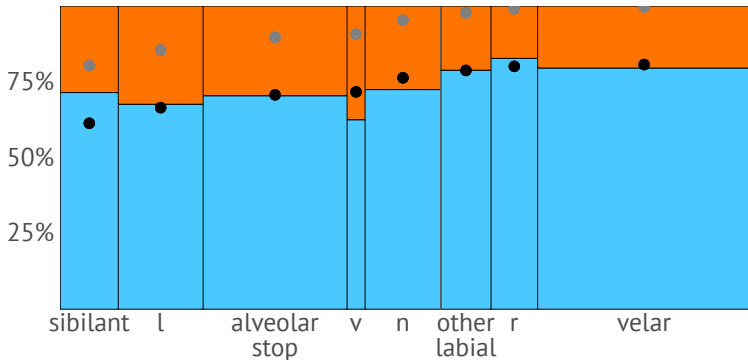
Results: phonology



Results: phonology

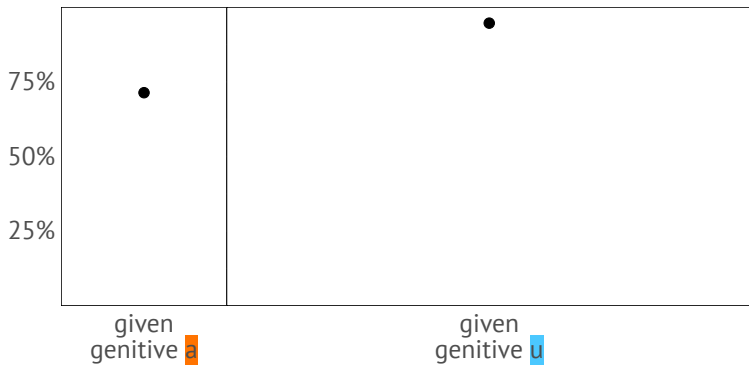


Results: phonology

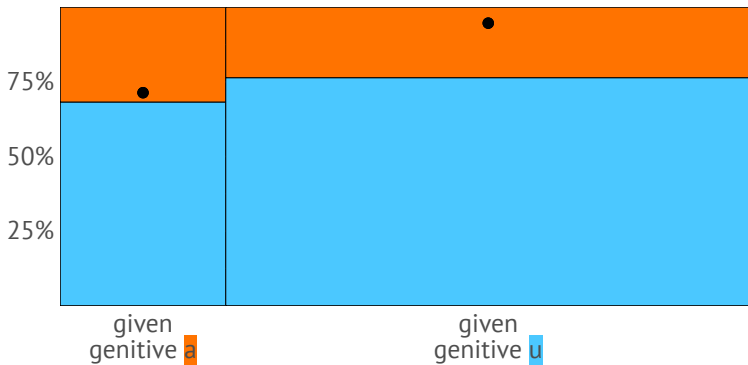


- Relatively high rates of **e** across the board – task made the alternative more salient
- Higher rates of **e** for *v* and *n* than expected – my corpus didn't include proper nouns, place names ending in *-ov* and *-ín* take **e**
- In general, the match is decent

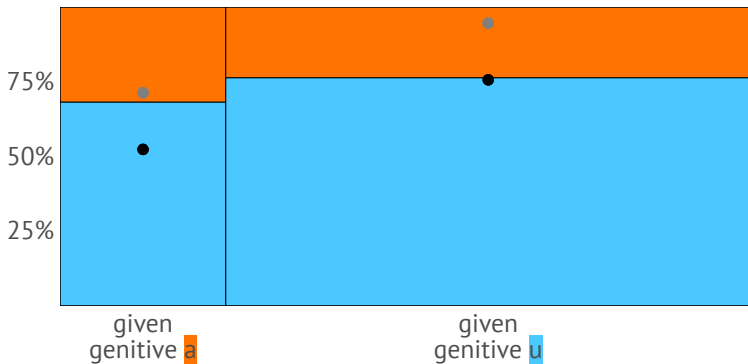
Results: morphology



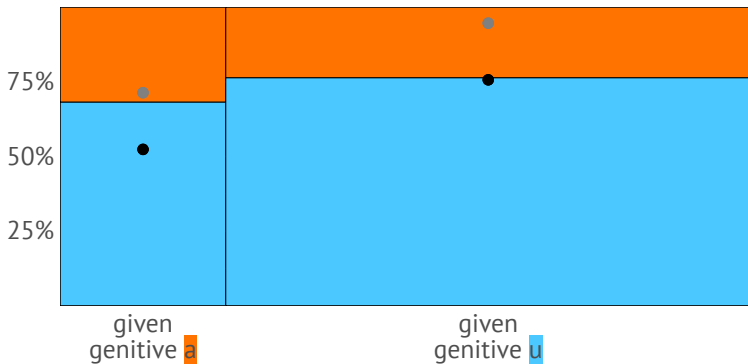
Results: morphology



Results: morphology



Results: morphology



- The difference is less extreme than in the lexicon – maybe also a task effect
- But the correlation is clear and significant

Summary

- Czech speakers have learned and apply (some kind of) phonological generalizations over the locative suffixes in their lexicon
- They also show a correlation between genitive and locative

Outline

Introduction

Corpus study

Nonce word study 1

Nonce word study 2

Conclusion

Morphological correlation

Two possible explanations of the genitive–locative correlation in Study 1:

- Productive application of learned morphological patterns to fill in gaps in storage where necessary
- Something more local to the task (e.g. priming between u)

Morphological correlation

Two possible explanations of the genitive–locative correlation in Study 1:

- **Productive application of learned morphological patterns to fill in gaps in storage where necessary**
- Something more local to the task (e.g. priming between u)

This is what I want to show

Morphological correlation

Two possible explanations of the genitive–locative correlation in Study 1:

- Productive application of learned morphological patterns to fill in gaps in storage where necessary
- **Something more local to the task (e.g. priming between u)**

Need to eliminate the alternate hypothesis

Morphological correlation

Two possible explanations of the genitive–locative correlation in Study 1:

- Productive application of learned morphological patterns to fill in gaps in storage where necessary
- **Something more local to the task (e.g. priming between u)**

Need to eliminate the alternate hypothesis

- If this is the case, it should appear in variable *real words* as well

Morphological correlation

Two possible explanations of the genitive–locative correlation in Study 1:

- Productive application of learned morphological patterns to fill in gaps in storage where necessary
- Something more local to the task (e.g. priming between u)

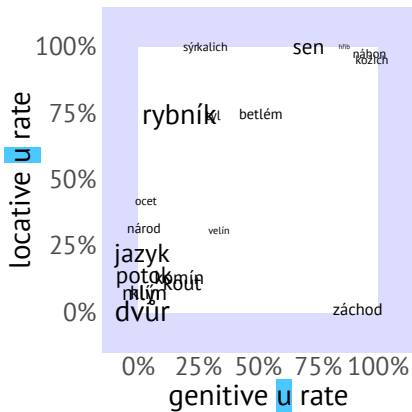
Need to eliminate the alternate hypothesis

- If this is the case, it should appear in variable *real words* as well

In either case, speakers should match each word's rates of u to the corpus

Study design

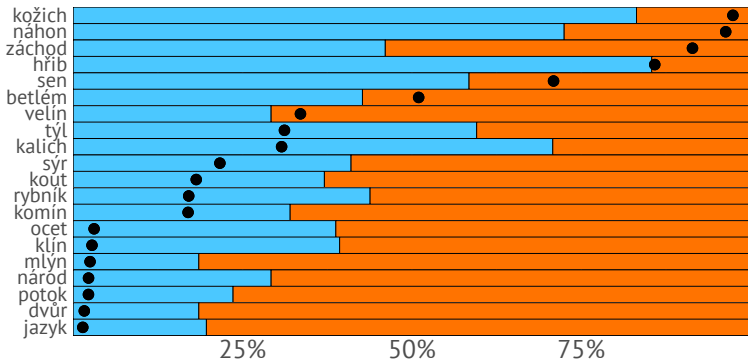
Same as the first study, but with 20 real words with variable genitive



- 15 words also have variable locative
- Not much to choose from – very different frequencies
- 90 speakers (1 removed)
- Each shown 40 trials (each word once with each genitive suffix)
- Total: 3560 trials

Results: genitive

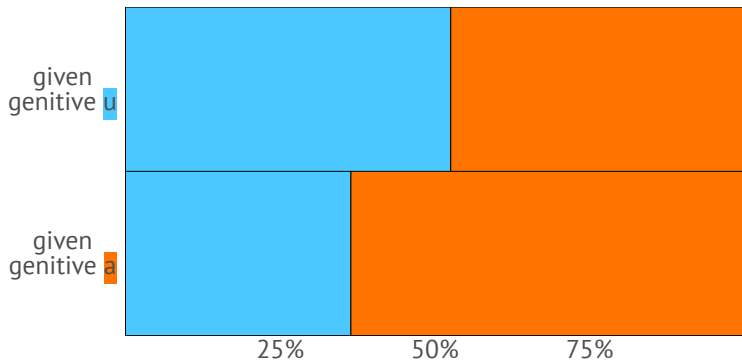
Remember: speakers were asked to choose a genitive as well as a locative.



- The range is compressed and a bit messy, as before
- But evidence that speakers know different words have different rates of genitive forms!

Results: genitive

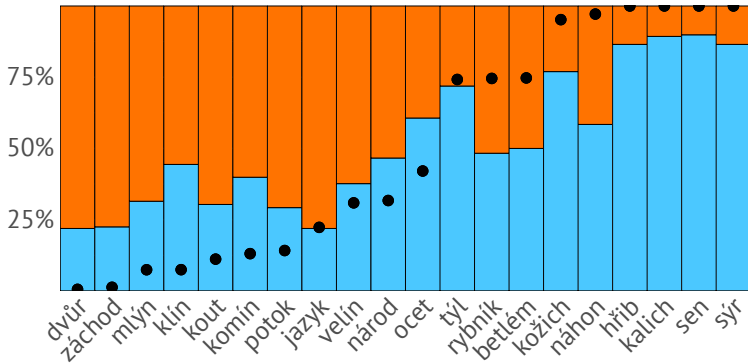
Are speakers influenced by the genitive we showed them?



Yes – evidence for some sort of local “priming”/suggestibility

Results: locative

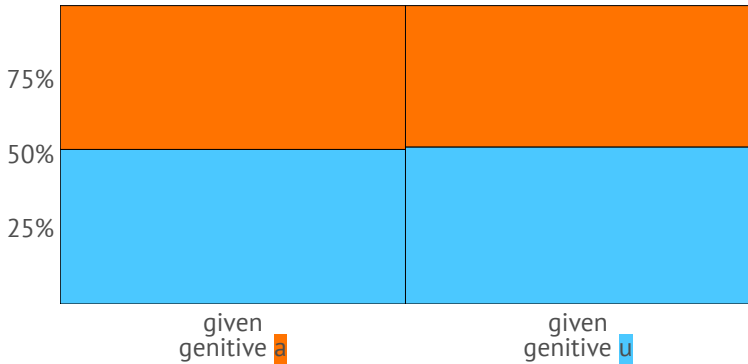
What about the locative?



As in the genitive, speakers store different rates for variable locative words

Results: locative

Are speakers influenced by the genitive we showed them?



No evidence for local “priming”/suggestibility across cases

Summary

- More evidence that speakers learn patterns from their lexicon
- The shown genitive influenced participants' choice of genitive but not locative
- Thus genitive–locative correlation found in Study 1 *cannot* be reduced to local priming – they are extending patterns in the lexicon to new words

Outline

Introduction

Corpus study

Nonce word study 1

Nonce word study 2

Conclusion

Conclusion

By asking Czech speakers to inflect real and made-up words, we find evidence for *stochastic application of gradient patterns*

- Applying *known words*' different rates of genitive and locative u
- Inferring the locative of *unknown words* from the distribution of locatives of words that share their properties
 - Phonology (e.g. final stem consonant)
 - Morphology (correlations between genitive and locative)

References I

- Ackerman, F., Blevins, J. P., & Malouf, R. (2009). Parts and wholes: Implicative patterns in inflectional paradigms. In J. P. Blevins & J. Blevins (Eds.), *Analogy in grammar: Form and acquisition* (pp. 54–82). Oxford University Press.
- Ackerman, F., & Malouf, R. (2013). Morphological organization: The low conditional entropy conjecture. *Language*, 89(3), 429–464.
- Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, 90(2), 119–161.
- Becker, M., Ketrez, N., & Nevins, A. (2011). The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language*, 87(1), 84–125.
- Berko, J. (1958). The child's learning of English morphology. *Word*, 14(2–3), 150–177.
- Copot, M., & Bonami, O. (2023). Behavioural evidence for implicative paradigmatic relations. *The Mental Lexicon*, 18(2), 177–217.
- Ernestus, M., & Baayen, R. H. (2003). Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language*, 79(1), 5–38.
- Finkel, R., & Stump, G. (2007). Principal parts and morphological typology. *Morphology*, 17, 39–75.
- Gouskova, M., Newlin-Łukowicz, L., & Kasyanenko, S. (2015). Selectional restrictions as phonotactics over sublexicons. *Lingua*, 167, 41–81.

References II

- Hayes, B., & Londe, Z. C. (2006). Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology*, 23, 59–104.
- Hayes, B., Zuraw, K., Siptár, P., & Londe, Z. (2009). Natural and unnatural constraints in Hungarian vowel harmony. *Language*, 85(4), 822–863.
- Křen, M., Cvrček, V., Hnátková, M., Jelínek, T., Kocek, J., Kovářiková, D., Křivan, J., Milička, J., Petkevič, V., Procházka, P., Skoumalová, H., Šindlerová, J., & Škrabal, M. (2022). Korpus SYN, verze 11 z 14.12.2022. Ústav Českého národního korpusu FF UK. www.korpus.cz
- Tabachnick, G. (2024). Hungarian speakers use morphological dependencies in inflecting novel forms. *Glossa: a journal of general linguistics*, 9(1).
- Wurzel, W. U. (1989). *Inflectional morphology and naturalness*. Kluwer.