

UiT

**THE ARCTIC
UNIVERSITY
OF NORWAY**

Rivalry in Russian: Three cases of morphosyntactic variation in modern Russian

Tore Nessel

CLEAR group (Cognitive Linguistics: Empirical Approaches to Russian)

**Co-authors:
Laura A. Janda
R. Harald Baayen
Julia Kuznetsova
Anastasia Makarova**



Big Questions – Small Answers

How about a cognitive linguistics/Construction Grammar cocktail?

- Rival forms:
 - Two or more forms appear to have the same function:
 - Слезы **каплют** одна за другой на клавиши. (Gončarov 1859)
 - Слезы в щи **капают**. (Bitov 1969)
- Questions:
 - Description: What are the factors motivating the choice between rival forms? Are we witnessing diachronic change?
 - Method: How can we study rival forms empirically?
 - Theory: What are the implications of rival forms for theoretical linguistics?
- Three case studies: “genitive-accusative shift”, “suffix shift” and “nu-drop”
- Tentative answers:
 - A cocktail of factors: morphophonology, morphology, semantics, frequency
 - Method: large corpora and statistical analysis
 - Theory 1: must accommodate multiple factors and statistical tendencies
 - Theory 2: morphological paradigm = radial category



Rival forms: Three case studies

- Genitive-accusative in objects:
 - Бояться жены**ы** (genitive object)
 - Бояться жен**у** (accusative object)
- “Suffix shift” («суффиксальный сдвиг»):
 - Слёзы кап**лю**т... (Verb suffix: -a)
 - Слёзы кап**аю**т ... (Verb suffix: -aj)
- “Nu-drop”:
 - Свет гас**ну**л ... (with *-nu* suffix)
 - Свет гас ... (without *-nu* suffix)



Case Study 1: Genitive-accusative of objects



- Co-author: Julia Kuznetsova
- PhD, Russian linguistics, Tromsø 2013
- Interested in
 - Construction Grammar
 - Empirical methods/quantitative analysis
 - Russian, Persian and other languages

Why FEAR? A personal note

- Nessel (2010: 46):
 - “If a verb includes [the middle voice marker] *–sja*, an accusative object is impossible.”
- Maier (2010: 144) in a review of Nessel:
 - “With *bojat’sja*, the accusative is very widespread (“mycket vanligt”) if the object is animate.”
- Disagreement in the scholarly literature:
 - Frequency
 - Grammar (declension)
 - Semantics/pragmatics (individuation, animacy)
 - Sociolinguistics (unmarked, colloquial, substandard)
 - Genre (fiction vs. non-fiction, direct speech vs. narrative)



Since the scholarly literature doesn't really bring clarity to the issue, we decided to try to find out ourselves – the hard way.

Three well-known cases of **acc-gen** variation

1. Object of negated verbs
 - *čitat' knigu*
 - read-INF book-**ACC**
 - *ne čitat' knigu*
 - not read-INF book-**ACC**
 - *ne čitat' knigi*
 - not read-INF book-**GEN**
2. Object with partitive meaning
 - *kupit' xleb-∅* 'buy bread'
 - buy-INF bread-**ACC**
 - *kupit' xleba* 'buy some bread'
 - buy-INF bread-**GEN**
3. Object of so-called weak intensional verbs
 - *ždat' avtobus-∅* 'wait for **the** bus'
 - wait-INF bus-**ACC**
 - *ždat' avtobusa* 'wait for **a** bus'
 - wait-INF bus-**GEN**



Alan Timberlake

- These three cases have received considerable attention in the literature.
- Timberlake (2004: 317): individuated reference is relevant for all three types.
- The **acc-gen** variation we are dealing with has received less attention.
- We are not aware of previous corpus studies or experimental studies.

Accusative with verbs (supposedly) governing the genitive

	Main corpus	Newspaper corpus	Total
<i>bojat'sja</i> 'fear'	35	53	88
<i>dobivat'sja</i> 'strive for'	1	1	2
<i>dožadat'sja</i> 'wait for'	73	41	114
<i>dostigat'</i> 'reach'	19	5	24
<i>izbegat'</i> 'avoid'	20	10	30
<i>kasat'sja</i> 'touch'	1	0	1
<i>opasat'sja</i> 'be afraid of'	2	0	2
<i>pugat'sja</i> 'be frightened of'	2	0	2
<i>slušet'sja</i> 'obey'	69	23	92
<i>stesnjat'sja</i> 'feel shy'	2	0	2

Only five verbs have enough **accusative** attestations to facilitate corpus analysis – and even they are not frequently attested in

How can such a low-frequent phenomenon be investigated empirically?

Method: Needles & Haystacks

- Russian National Corpus at www.ruscorpora.ru:
 - Basic corpus: 230 million words
 - Newspaper corpus: 173 million words
- Challenge:
 - Too **few** examples with *bojat'sja* + **Acc** to use a small sample from the corpus
 - Too **many** examples with *bojat'sja* + **Gen** to use the whole corpus
- Proposed methodology:
 - Search for *bojat'sja* + **Acc** in whole corpus, weed out noise manually
 - Create random sample for *bojat'sja* + **Gen** from whole corpus, weed out noise manually
 - Estimate the frequency of *bojat'sja* + **Gen** in the whole corpus on the basis of the sample

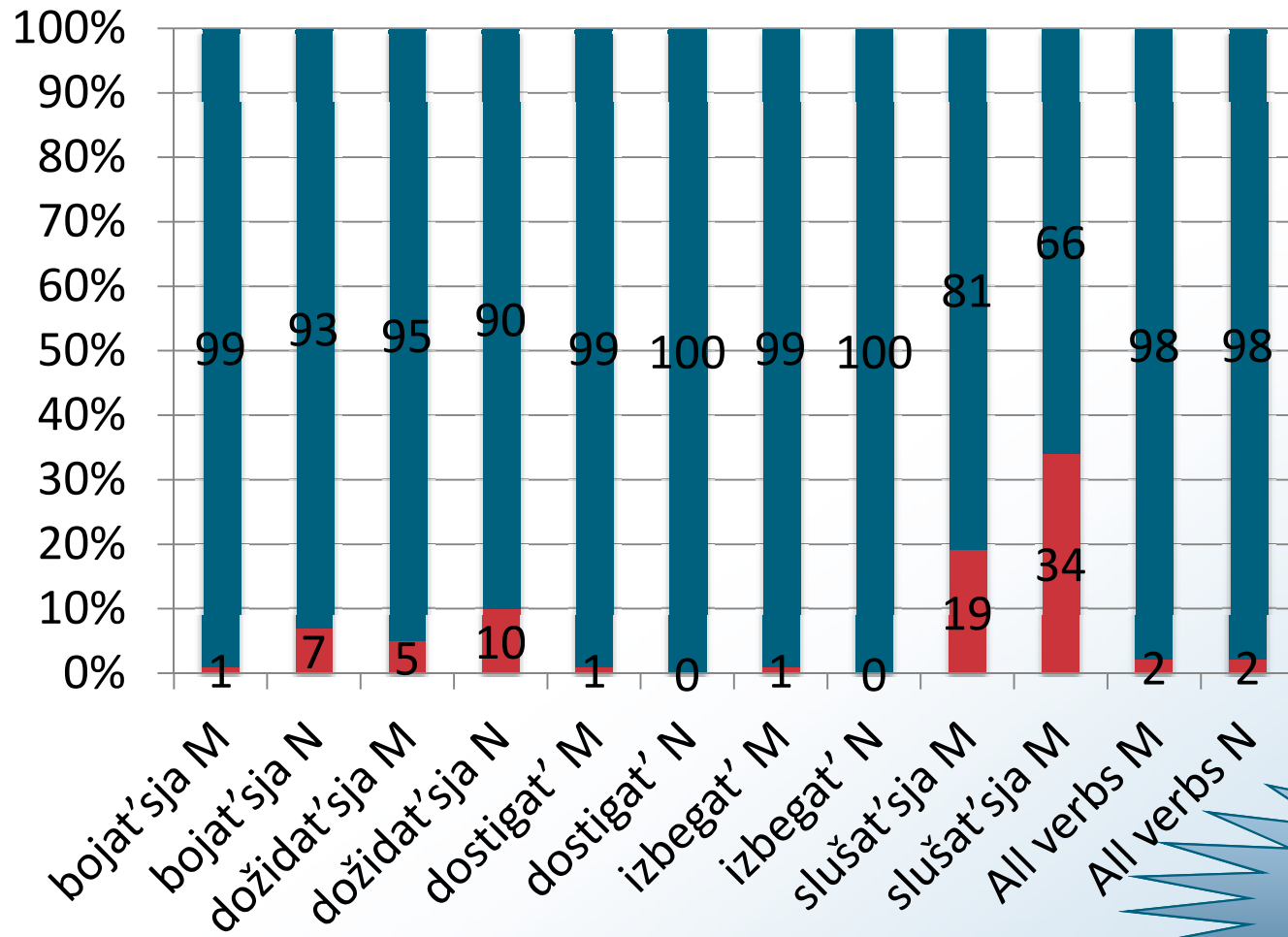


Our proposed methodology facilitates realistic comparison of the frequencies of *bojat'sja* etc. + **Acc** (attested) and *bojat'sja* etc. + **Gen** (extrapolated).

How frequent is the accusative?

Verb	Corpus	#Acc	#Gen	#Acc+Gen	%Acc
<i>bojat'sja</i>	Main	35	3437	3472	1
	Newspaper	53	757	810	7
<i>dožadat'sja</i>	Main	73	1406	1479	5
	Newspaper	41	370	411	10
<i>dostigat'</i>	Main	19	3185	3204	1
	Newspaper	5	2604	2609	0
<i>izbegat'</i>	Main	20	1765	1785	1
	Newspaper	10	2115	2125	0
<i>slušet'sja</i>	Main	70	292	362	19
	Newspaper	23	44	67	34
All verbs	Main	212	10086	10298	2
	Newspaper	132	5891	6023	2

How frequent is the **accusative**? (2)



Observations:

1. General:

- **Accusative** infrequent
- only one verb > 10%

2. Verbs:

- Accusative friendliness hierarchy:

1. *slušat'sja*
2. *dožadat'sja*
3. *bojat'sja*
4. *dostigat'/izbegat'*

**Observed differences
are statistically
significant**

So is this new? What have people said?

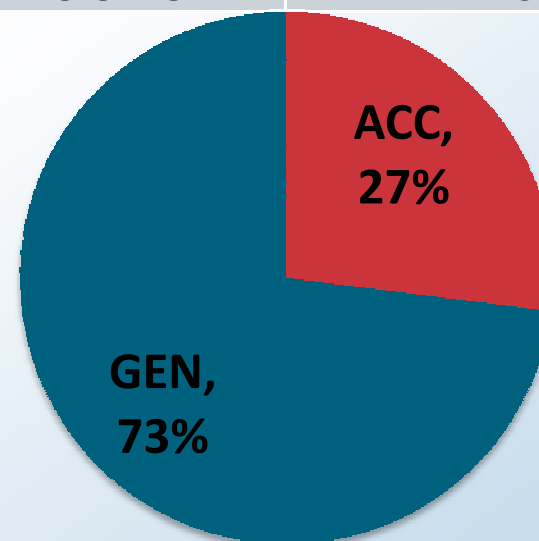
- Miloslavskij (1978: 212):
 - “**scanty exception**” (Russian: *mizernoe isključenie*)
- Prokopovič et al. (1975: 17):
 - “**isolated examples**” (Russian: *ediničnye primery*),
- Butorin (1966: 130):
 - “Writers of the 19th century **sometimes** use the construction *bojat’sja* + **accusative**” (My translation. TN)
- Gorbačevič (1971: 237):
 - “**not infrequently**” (Russian: *neredko*)
- Krys’ko (1997: 245):
 - “In addition to **the massive use** of the **accusative** with the reflexive *bojat’sja*, from the late 1800s we also find the accusative instead of the genitive for [...] *izbegat’* [‘avoid’]” (My translation. TN)

Corpus data suggest a situation close to what Butorin assumed.

Psycholinguistic experiment: *bojat'sja*

- From corpus to linguistic competence (mental grammar):
 - A long way to go!
- Psycholinguistic experiment:
 - 409 participants
 - ca. 8000 datapoints
 - Internet survey
 - Both sexes
 - Different age groups
 - Different educational backgrounds

	# examples
Total ACC	2244
Total GEN	5863
Total ACC+GEN	8107



The experimental data suggest that the accusative is somewhat more frequent in present-day colloquial Russian than indicated by corpus data.

The Individuation Hypothesis

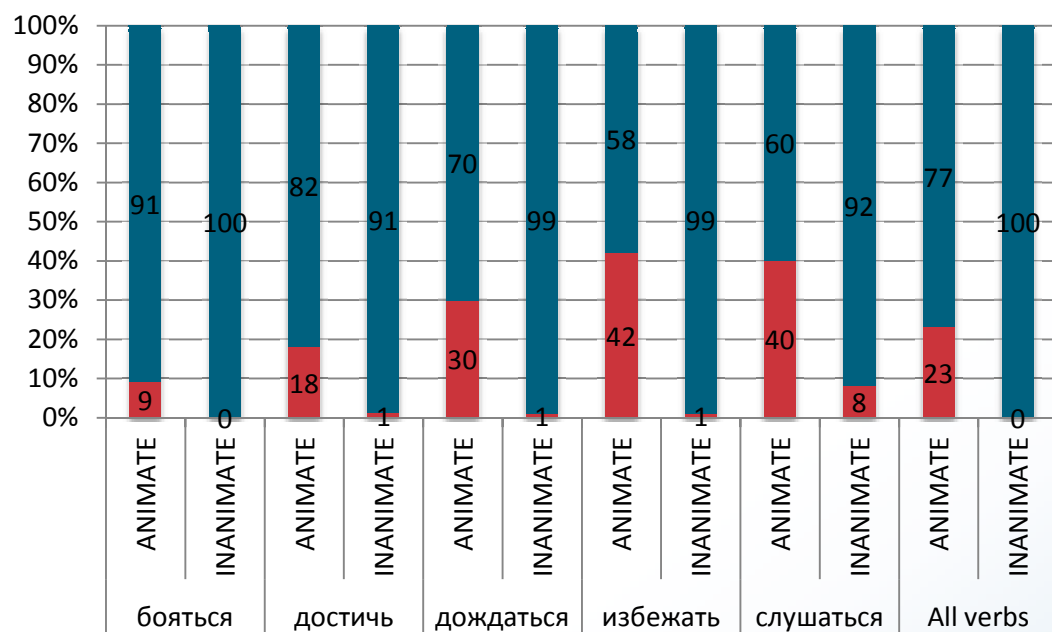
- The Individuation Hypothesis:
 - Individuation favors the **accusative** in the object
 - **Genitive** is used with low degree of individuation
 - Individuation = likelihood of being treated as an individual
 - Timberlake 2004: 319, Nichols 1993: 82, Kagan 2013
- Individuation favors **accusative** objects in three constructions
 - Object of negated verbs
 - Object with partitive meaning
 - Object of so-called weak intensional verbs
- Reasonable to expect similar effect for *bojat'sja*-type verbs
- In order to find out, we tested the effect for five verbs in corpus and one verb in experiment
- Not enough data to test for more than two categories:
 - **Animate**
 - Inanimate



Individuation hierarchy:

1. **Human**
2. **Animal**
3. **Concrete object**
4. **Abstraction/mass**

Animacy for five verbs in the corpus



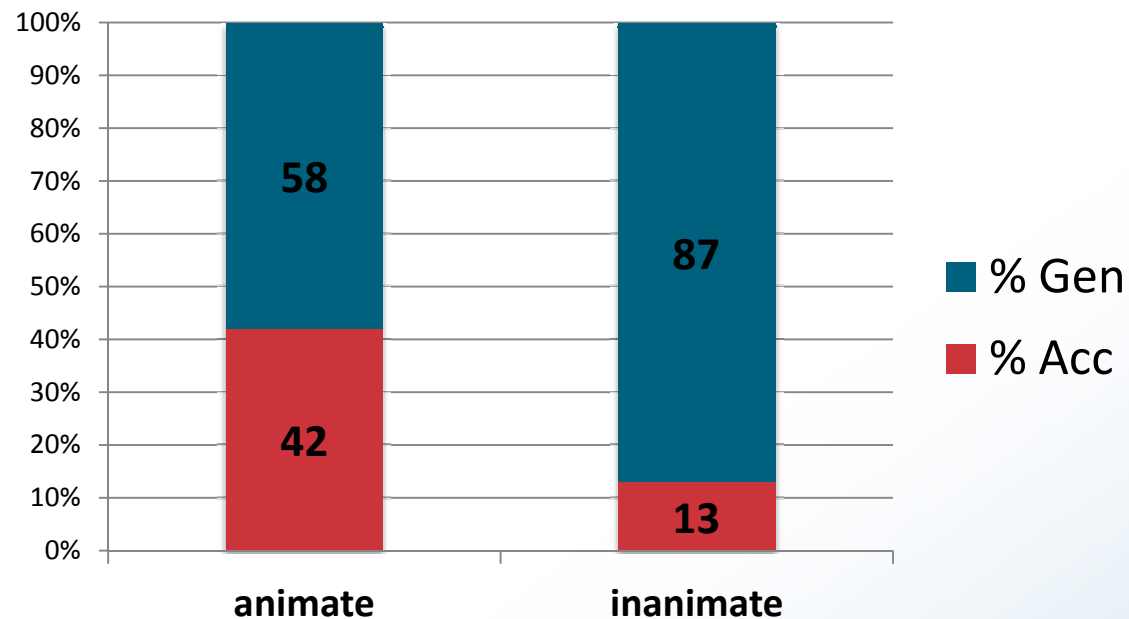
Data from main corpus.

(too small numbers for some verbs in newspaper corpus to facilitate comparison)

- Animate is always more **accusative** friendly than inanimate:
 - Animate > Inanimate
- Accusative for inanimate is very rare:
 - 0-1% accusative
- Exception: *slušaťsja* ‘obey’
 - However, *slušaťsja* is not high frequent, and therefore has limited impact on the overall picture.
- Statistically significant differences (chi-square, $p = 2.2e^{-16}$)
- Effect size: moderate to large (Cramer’s V-value = 0.4)

Individuation Hypothesis confirmed: Animate > Inanimate

Individuation: Animacy in experiment

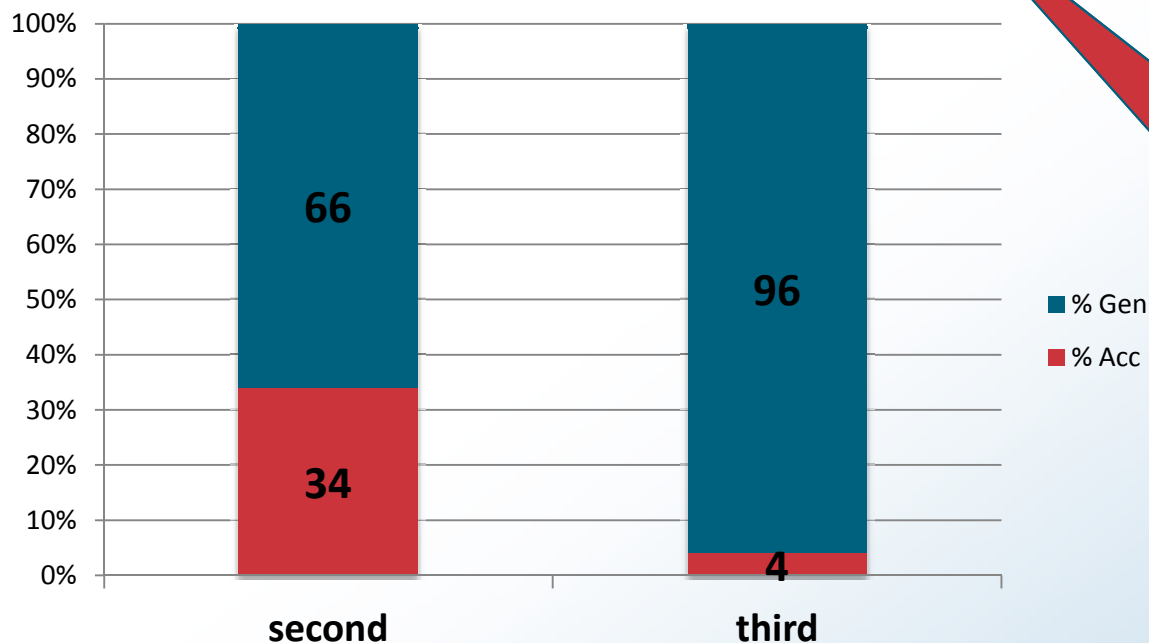


Experimental data support Individuation Hypothesis:
animate > inanimate

The declension system – a factor?

	Decl. I	Decl. II	Decl III	Plural
Animate	Acc=Gen			Acc=Gen
Inanimate				

In these parts of the system, we cannot know if the object is in the acc or gen



Most examples of *bojat'sja* etc. + ACC are from this part of the system. Is there a difference in the experiment?

Nouns of third declension are less likely to be used in the accusative.

- Most likely an epiphenomenon: very few animate nouns in third declension.

First declension (masc in C, neuter in –o)

- Is the accusative possible in the first declension?
- Animates:
 - Accusative and genitive have the same form, so we cannot tell.
 - *bojat'sja studenta* 'fear a student' = **accusative** or **genitive**?
- Inanimates:
 - 3 examples in our database (from Soviet newspaper)
 - Bojus' «**Detroit**» (2 times)
 - Boitsja «**Nefteximik**»
 - Krys'ko (1997: 242f.) cites a handful of examples.
 - Only one example is clear:
 - *Sirija ... gosudarstvom, **kotoroe** izrail'*

NB! Pronoun/proper name, OV word order, and long distance between O and V.

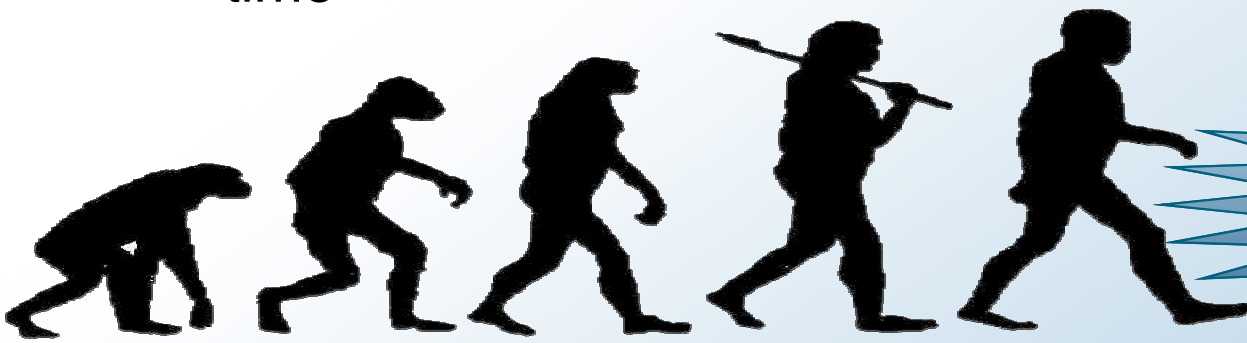


Declension:

- First declension (inanimates): accusative is extremely rare.
- Accusative only possible under otherwise optimal conditions?

Ongoing change or stable variation?

- We have seen that there is variation between accusative and genitive...
- But does this mean that we are witnessing ongoing change, or is the variation stable over time?
- Stability over time:
 - English *ask* and *aks* 'to inquire for information' have lived side by side for centuries in certain varieties of English.
- Two arguments that we are dealing with change over time:
 1. Experiment: younger informants use more accusative
 2. Corpus data (main corpus) show increase for accusative over time



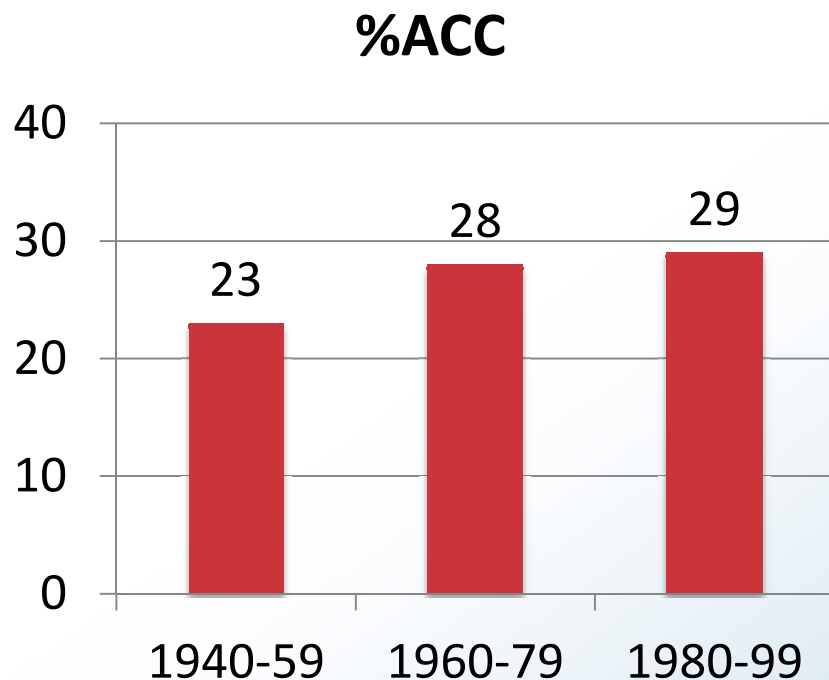
Let us take a closer look
at the experiment and
main corpus data!

Ongoing change (2): Experimental data

Year of birth	#Acc	#Gen	%Acc
1940-1959	224	744	23
1960-1979	1029	2659	28
1980-1999	957	2316	29

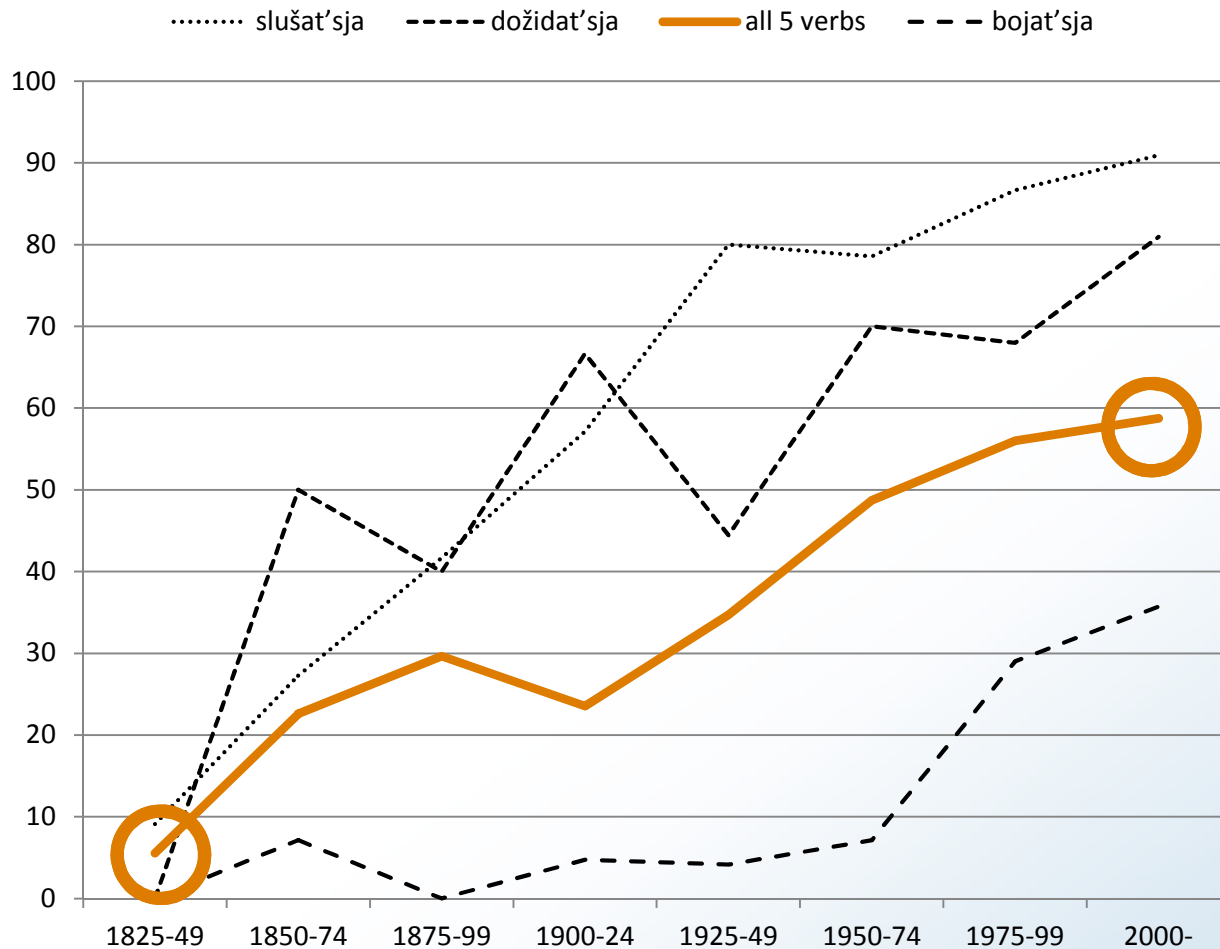
Observations:

1. Younger people are more likely to use the **accusative**.
2. Differences are statistically significant:
 - p-value = 0.0002
3. Small, but reportable effect size:
 - Cramer's V-value = 0.1



(Weak) support for the hypothesis that we are dealing with language change!

Ongoing change (3): Diachronic data (main corpus)



Observations:

1. Accusative is on the increase
2. Possible to carry out statistical analysis of “all 5 verbs”
3. Differences between 1825-49 and 2000- are significant:
 - p-value 2e-04.
4. Moderate-large effect size:
 - Cramer's V-value = 0.4

Diachronic data from main corpus clearly indicate that accusative is on the increase.

Verbs: Grammatical voice



- Russian has the suffix **–sja**, which is a marker of
 - Middle voice (in the sense of Kemmer 1993): *myt'sja* 'wash (oneself)'
 - Passive voice: *kniga čitaetsja* 'the book is being read'
- Conventional wisdom:
 - Verbs in **–sja** cannot take objects in the accusative.
- The “accusative friendliness hierarchy” goes against this:
 - *slušaťsja* > *dožidat'sja* > *bojaťsja* > *izbegat'/dostigat'*
 - the three most accusative friendly verbs have **–sja**!
- However, what does **–sja** mean in these verbs?
 - *slušaťsja* 'obey' vs. *slušať* 'listen': no transparent voice relationship
 - *dožidat'sja* 'wait' vs. **dožidat'*: only one member attested
 - *bojaťsja* vs. **bojať*: only one member attested

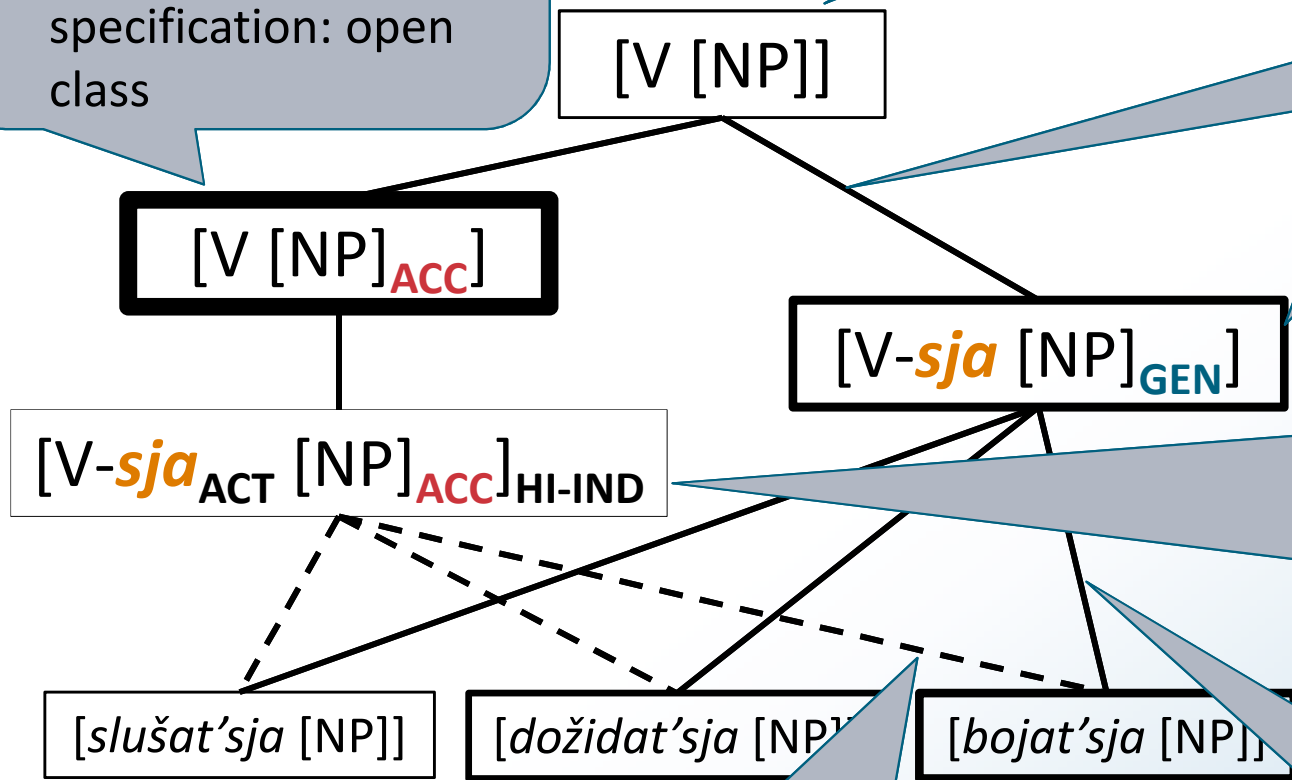
Opacity Hypothesis:

- The more opaque **–sja** is, the more likely is the verb to combine with an accusative object.
- Bleaching of **–sja** facilitates increase of accusative.

for the genitive accusative shift

- Default: **Acc** object
- Thick line = high type frequency
- No semantic specification: open class

- Boxes = Constructions
- V governs object NP



- Association lines
- Connects constructions with subtypes

- Some **sja**-verbs take **genitive** objects

- A few **sja**-verbs take **accusative** objects
- Conditions:
 - ACT = "not transparent marker of middle voice"
 - Highly individuated

- Most **accusative** friendly verb to the left.
- Thin lines: low freq.

- Dashed lines:
 - synchrony: variation
 - diachrony: emerging connection (change)

- Association lines show that verb takes **genitive** object

FEAR: Conclusions

Don't be afraid! We can sort it out (in Construction Grammar)

1. Descriptive:

- a. Accusative (still) not very frequent for verbs like *bojat'sja* 'fear'
- b. Accusative is increasing over time (at least for animate nouns)
- c. Interaction of a cocktail of factors facilitate the shift to accusative:
 - i. Individuation (animacy)
 - ii. Frequency (type and token)
 - iii. Grammatical voice (the *-sja* suffix)
 - iv. Verb semantics (inensionality, directionality, individuation compatibility)



2. Methodological – needle in haystack problem:

Possible to investigate low-frequent phenomena in large corpora through combinations of searches in whole corpus and estimates based on random samples.

3. Theoretical:

- a. Constructions change through interaction of numerous factors
- b. Construction Grammar networks facilitate unified analysis of such multifactorial variation and change



Case study 2: Suffix shift



Co-author: Laura A. Janda

An appropriate name for the boat he came on!



The king of statistics in linguistics visiting Tromsø in September 2013

Co-author: R. Harald Baayen

Suffix shift in Russian: **-a** → **-aj** in verbs

- Ongoing diachronic change
- Non-productive suffix *-a* replaced by productive suffix *-aj*:
 - /kaplʲut/ (with *-a*) → /kap**aj**ut/
‘(they) drip’
- Affects all present tense/imperative forms.
- Attested in the Russian National Corpus.
- Our database consists of about 20,000 examples.

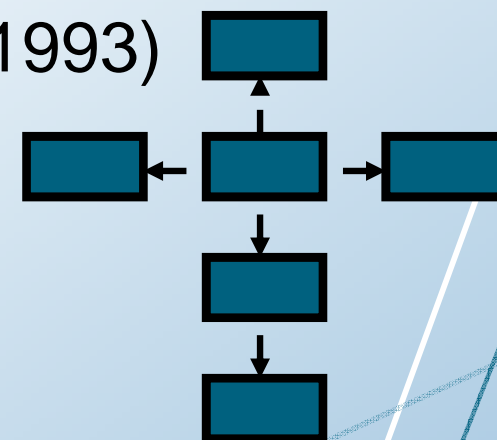


Does suffix shift affect all inflected forms of a verb to the same extent?

What is an inflectional paradigm?

- A classical, Aristotelian category
 - A list of verb forms
 - All forms have the same status
 - Word & Paradigm (Matthews 1972)
- An epiphenomenon
 - Inflected forms are results of concatenation of morphemes (inflection, derivation, syntax)
 - Distributed Morphology (Halle & Marantz 1993)
- A radial category (Lakoff 1987)
 - Paradigms have structure
 - Prototypical and peripheral forms

	Sg	Pl
1		
2		
3		

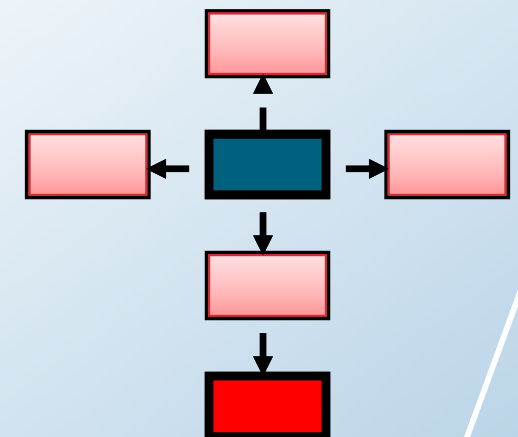


Is it possible to investigate the question empirically?

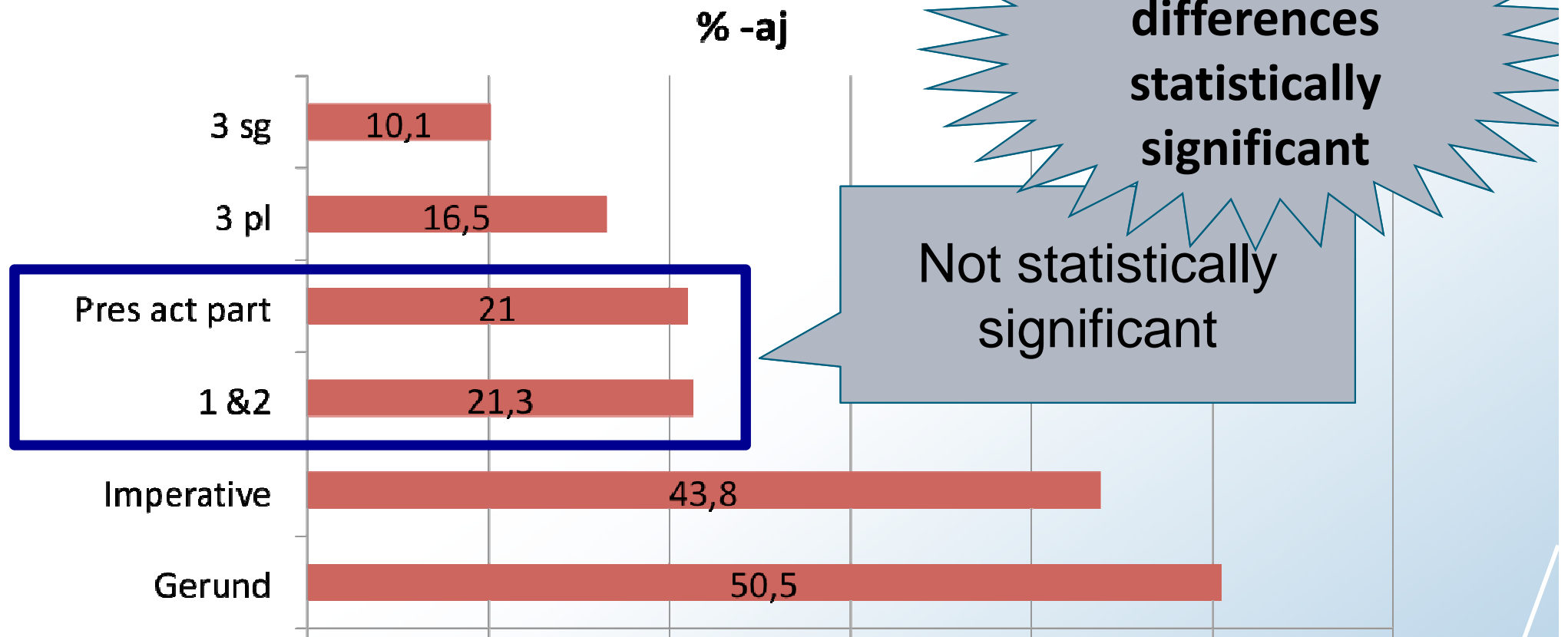
Our approach

- Language change: Predictions
 - Paradigm as Aristotelian category
 - all forms affected in the same way
 - Paradigm as epiphenomenon
 - all forms affected in the same way
 - Paradigm as radial category
 - peripheral forms most affected
- Statistical modeling
 - Logistic Mixed Effects modeling facilitates systematic analysis of several factors
 - Thanks to R. Harald Baayen for help with statistical analysis!

	Sg	Pl
1		
2		
3		



Results



Finite forms (especially 3 sg) resist regularization, while gerund is most likely to change. Suggest that paradigms are construction networks

Another question about suffix shift: Do all kinds of verbs accept it?

- Traditional questions in historical linguistics:
 - What happened?
 - Why did it happen?

How about asking the opposite questions:

- What did NOT happen?
- Why did it NOT happen?

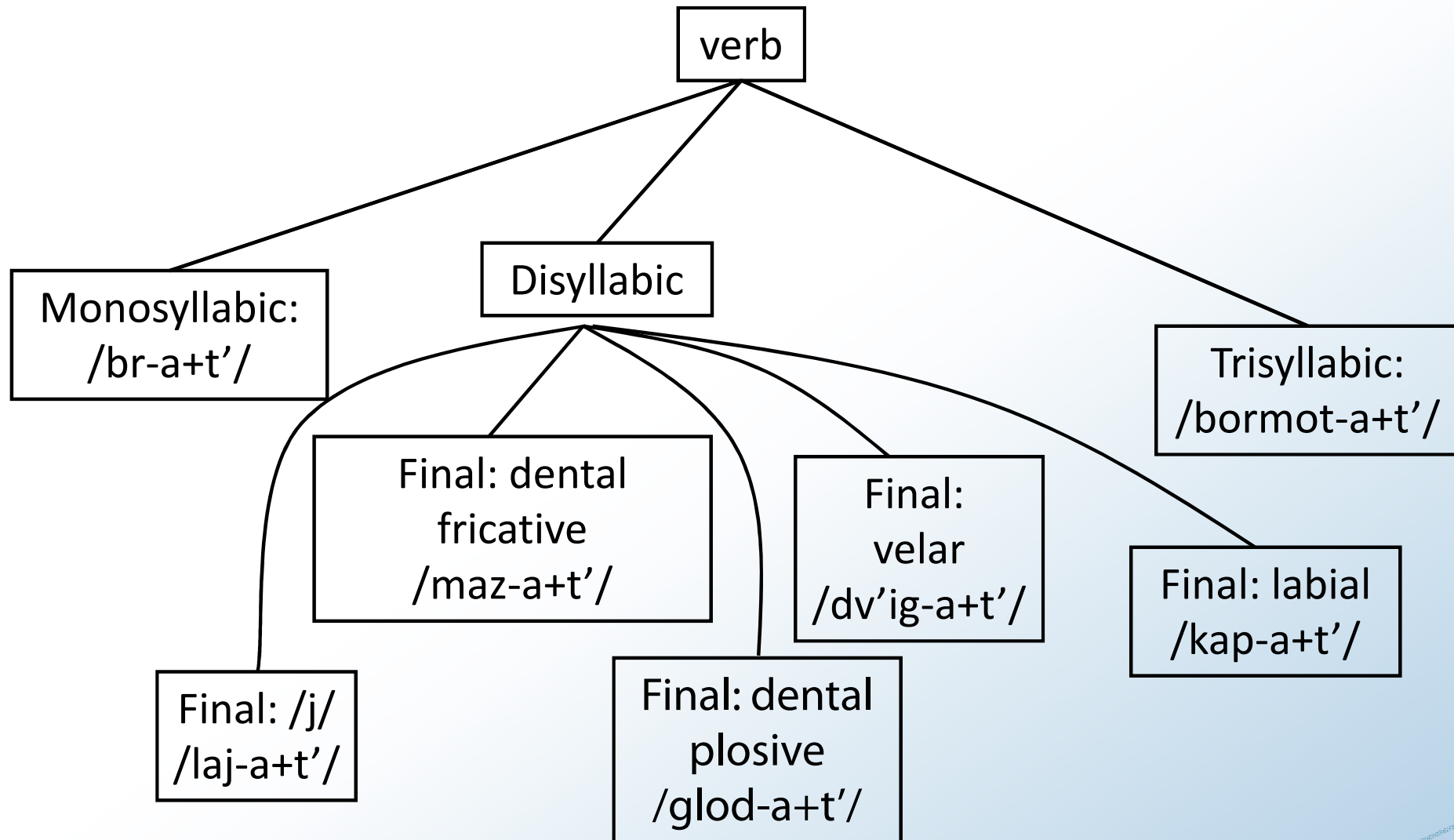
Here the two types are identical!

Suffix shift:
Based on the similarity in the past tense/infinitive speakers choose the productive way to form present tense/imp

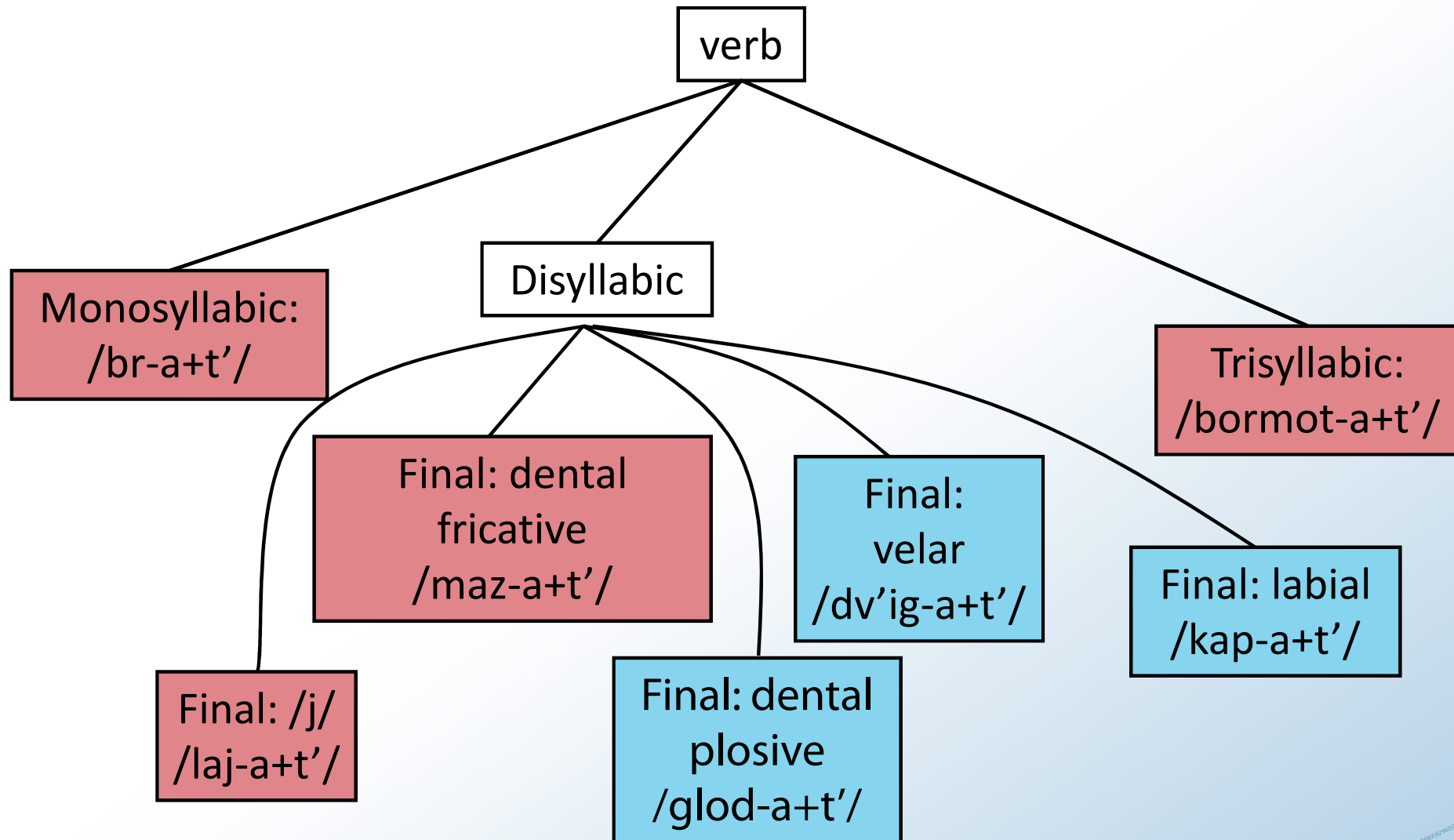
An example of “abductive change” (Andersen 1973)

	Непродуктивный	Продуктивный
Наст. вр. 3 ед.ч.	каplet	дела ает
Наст. вр. 3 мн.ч.	каплют	дела ют
Повел. наклонение	капли(те)	дела й
Акт. причастие	каплющий	дела ющий
	капля	дела я
	кап л	дела л
	кап ли	дела ли
	кап ать	дела ть


Blocking or “suffix shift”?



Blocking or “suffix shift”?



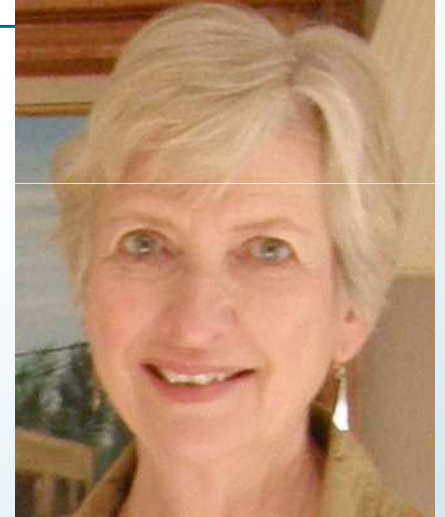
Complexity of alternations and **Blocking**

3. Trans. Soft. (plos./lab.): t~č, k~č, p'~pl'
 2. Trans. Soft. (fricatives): s~š
 1. Plain softening: s~s'
 0. No alternation: j~j
- 

Blocking correlates with low complexity.
Suffix shift eliminates most complex alternations.

Two types of generalizations

- a) Source-oriented:
 - Specify how a target is formed from a source, e.g. $A \rightarrow B$.
- b) Product-oriented:
 - Characterize a target without specifying how it is formed from a source.



Joan Bybee



Dan Slobin

Product-oriented generalization for Russian verbs

a) Present tense forms: /...Vj+V.../

Stem ends in Vj, ending begins with V

b) /aj/-verbs: /d' el-**aj+ut**/ 'they do'

/ej/-verbs: /krasn' -**ej+ut**/ 'they redden'

/ova/-verbs: /obraz-**uj+ut**/ 'they form'

/j/-final /a/-verbs: /**aj+ut**/ 'they bark'

Product-oriented hypothesis

- a) Suffix shift occurs only if a product-oriented generalization is not already satisfied.
- b) “If it ain’ t broke, don’ t fix it”
Don’ t add /j/ if you already have one.



Test of hypothesis

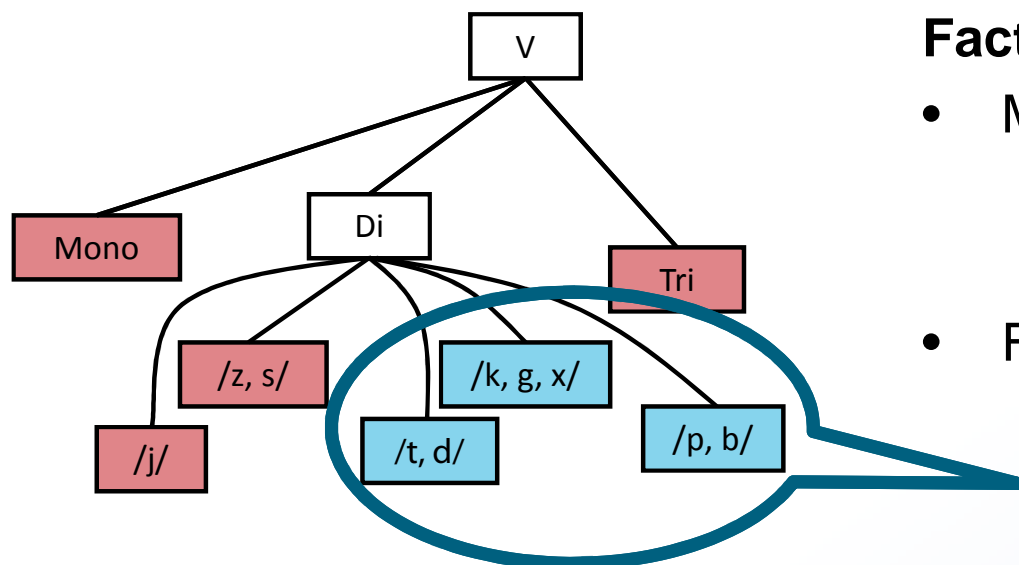
a) /j/-final roots:

- Already have /j/: /laɟ+ut/ ‘they bark’
- No need to add /j/
- We correctly predict blocking.

b) Non-Syllabic/Fricative-final roots:

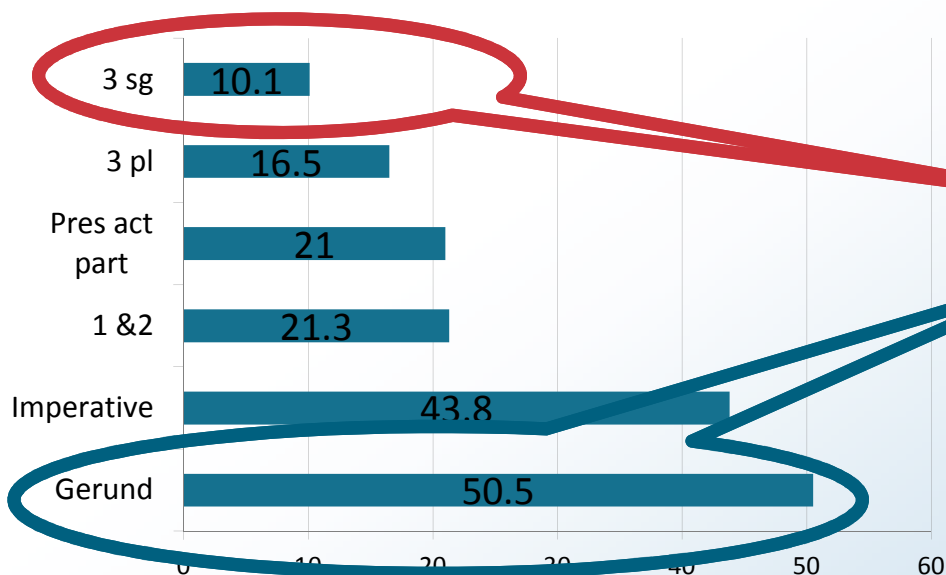
- Do not have /j/
- Product-oriented generalizations cannot explain why suffix shift is blocked for such verbs

“Suffix shift” – relevant factors



Factor cocktail:

- Morphophonology:
 - number of syllables
 - root-final consonant/alternation
- Frequency:
 - Verbs undergo suffix shift to various degrees, apparently depending on frequency (Berdicevskis & Piperski 2014 & 2015)
- Inflectional features:
 - 3 sg most conservative (e.g. **каплет**)
 - Gerund most innovative (e.g. **капая**)
- Data from Russian National Corpus
- NB! Statistical tendencies, not categorical differences



Case study 2: Suffix shift



- Co-author: Anastasia Makarova
- PhD, Russian linguistics, Tromsø 2014
- Topic:
 - Verbal diminutives in Russian

Empirical problem

	Variant 1:	Variant 2:
Infinitive	Pri-vyk-nu-t'	Pri-vyk-nu-t'
Masculine sg	Pri-vyk-∅	Pri-vyk- NU -l
Feminine sg	Pri-vyk-∅-l-a	Pri-vyk- NU -l-a
Neuter sg	Pri-vyk-∅-l-o	Pri-vyk- NU -l-o
Plural	Pri-vyk-∅-l-i	Pri-vyk- NU -l-i
Active participle	Pri-vyk-∅-š-ij	Pri-vyk- NU -vš-ij
Gerund	Pri-vyk-∅-ši	Pri-vyk- NU -vši

A group of Russian verbs display variation in the past tense forms: ∅ ~ **NU**.

What is the distribution of ∅ and **NU**?

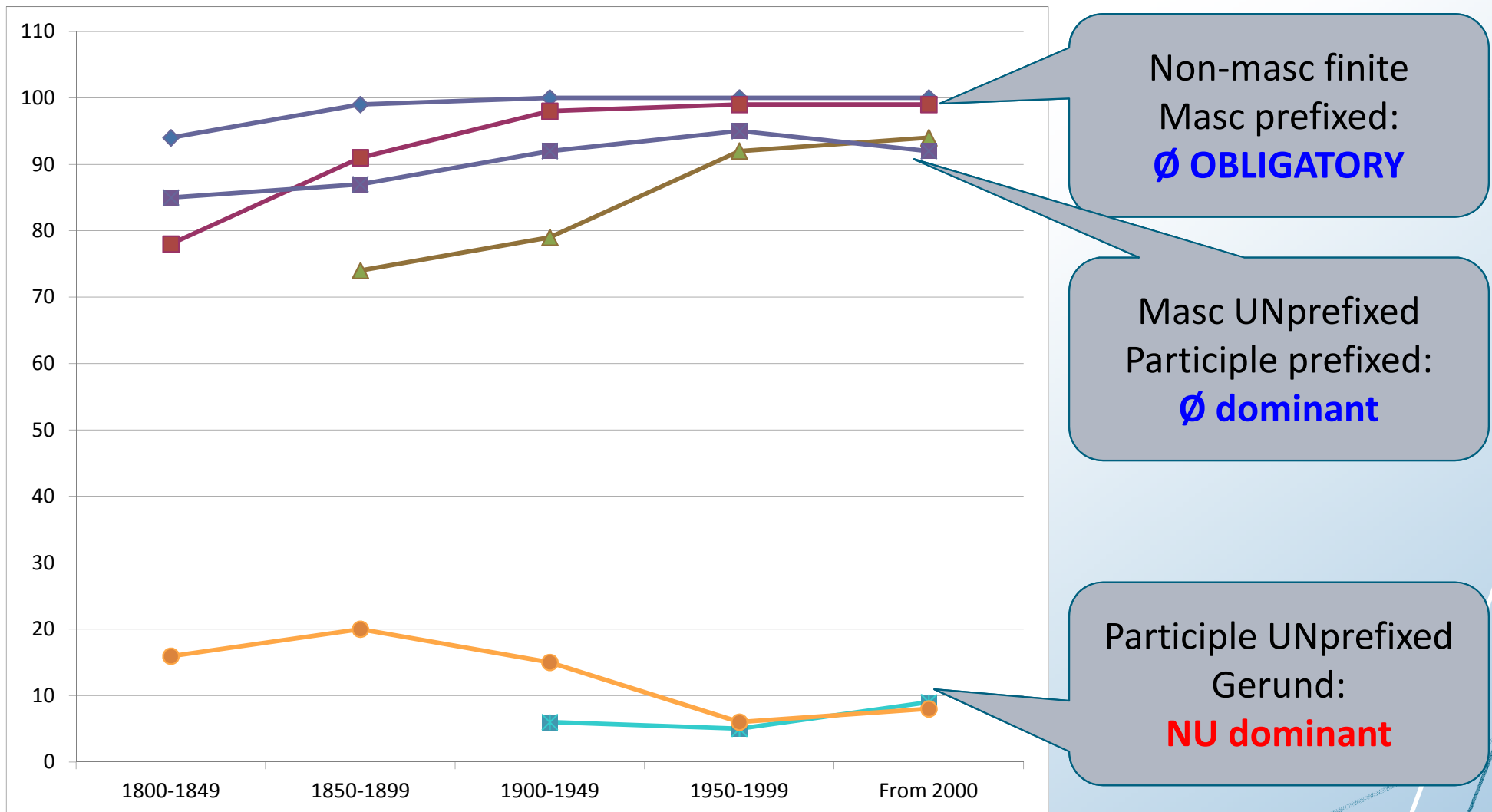
Which factors determine this distribution?

Factors at play

- Corpus investigation
 - Russian National Corpus (main corpus)
 - Investigated all verbs showing \emptyset ~ **NU** according to Academy Grammar.
 - Our database:
 - 34,026 examples
 - 74 verbs
 - Period covered: 1800-2010
- Investigated factors
 - Root-final consonant
 - Inflected forms
 - Aspectual prefixes
 - Voice marker -**sja**
 - Transitivity

Statistical analysis shows these factors are of major importance.

Historical development in a nutshell



Non-masc finite
Masc prefixed:
∅ OBLIGATORY

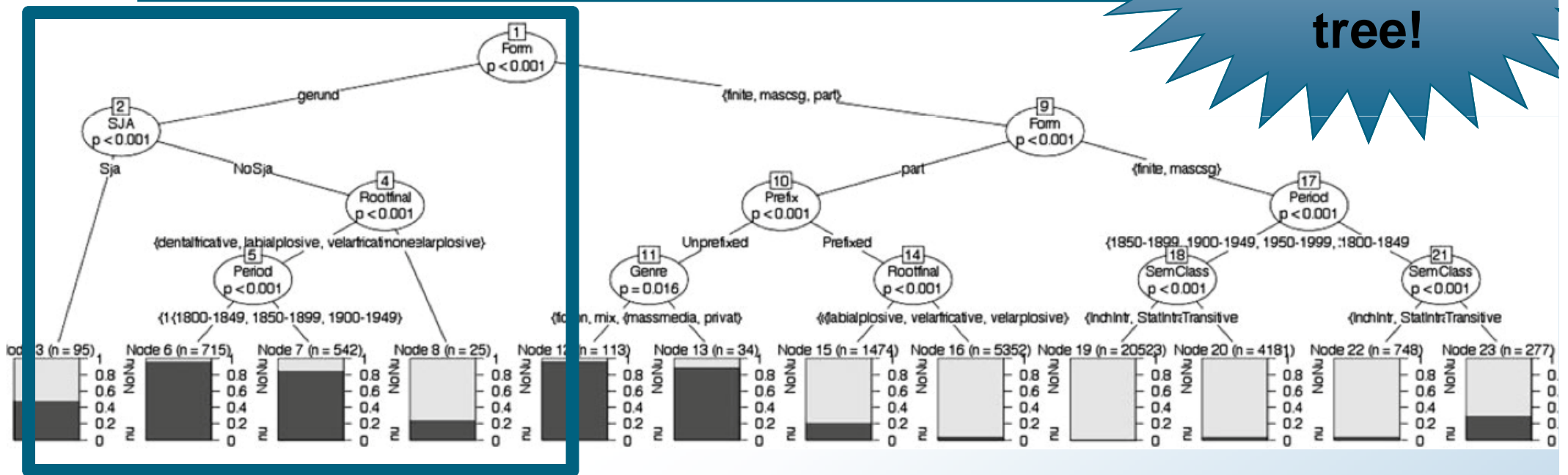
Masc UNprefixed
Participle prefixed:
∅ dominant

Participle UNprefixed
Gerund:
NU dominant

Diachrony: Polarization towards all **∅** or all **NU**

Statistical analysis: Classification tree

Let's zoom in on a part of the tree!



“Tree and forest” model (Strobl, Malley and Tutz 2009)

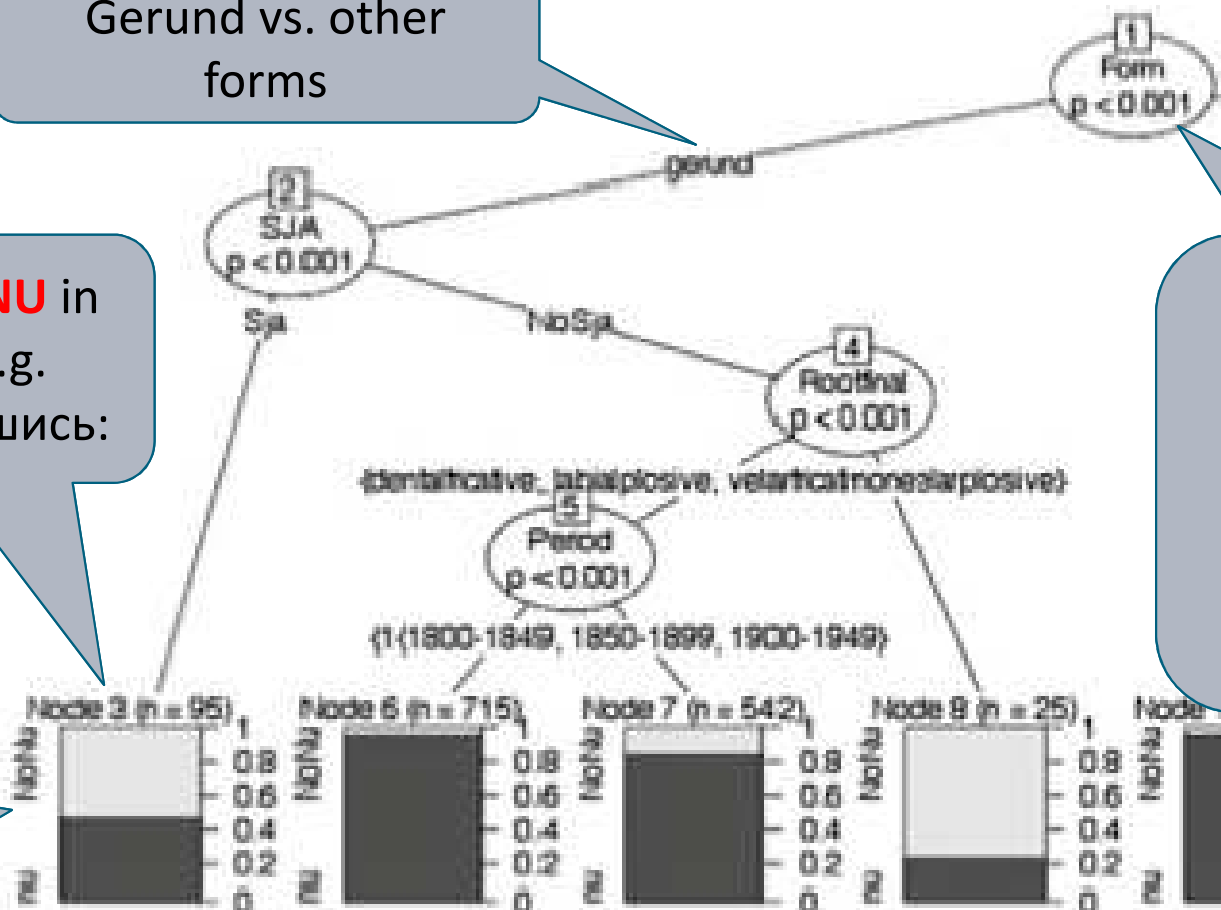
- Sorts the material, trying to predict the choice of \emptyset vs. **NU**
- Provides intuitive diagram of the outcomes that are predicted and yielded by various combinations of predictor values
- A very powerful tool for the study of rival forms in language
- A good alternative to logistical regression

Zooming in: How to make use of the tree

Gerund vs. other forms

-sja inhibits **NU** in gerunds, e.g. проник(ну)вшись:

Light: \emptyset
Dark: **NU**



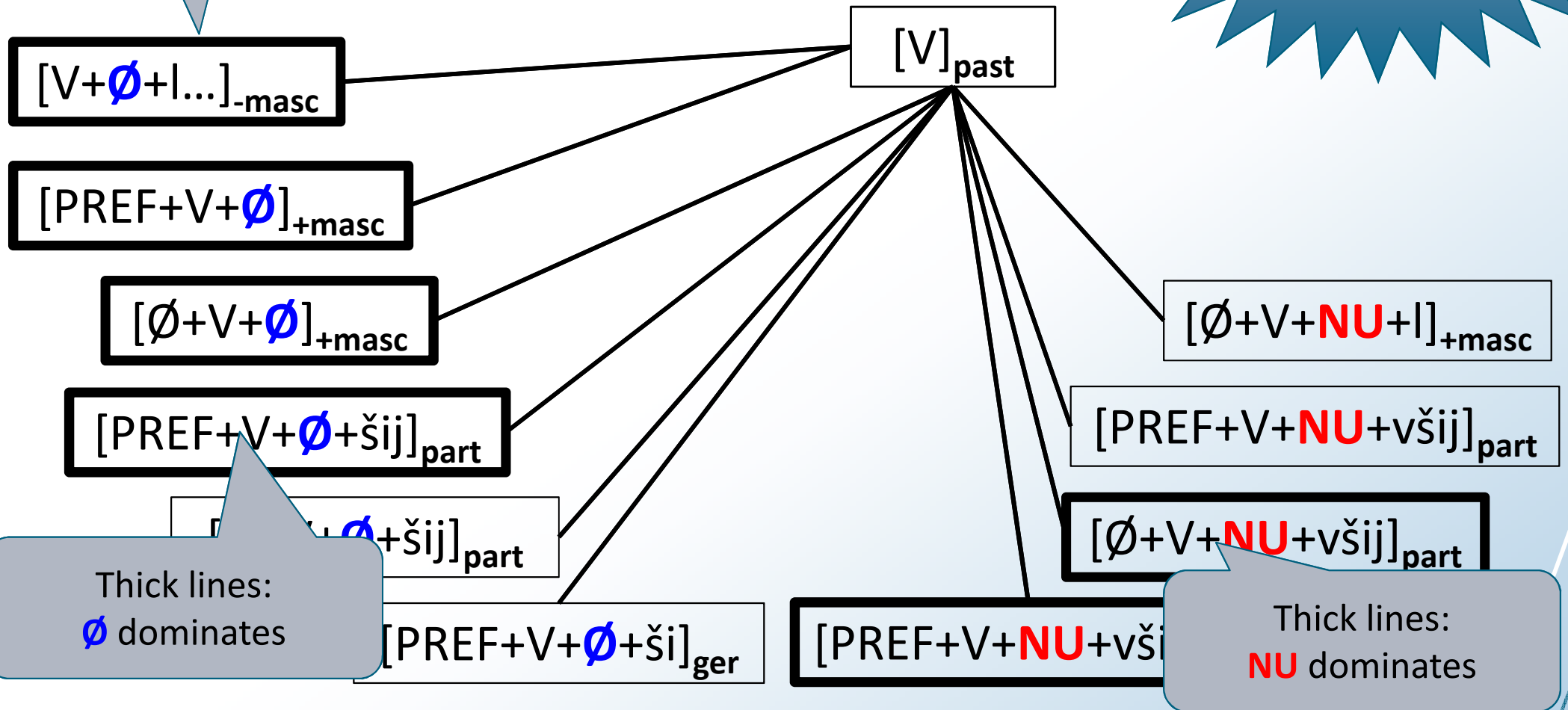
Most important parameter: Inflected form NB! provides p-values to indicate statistical significance

We had not spotted the relevance of *-sja* for \emptyset vs. **NU** until we carried out the tree and forest analysis!

∅ obligatory
No corresponding
construction with **NU**

Again:
Paradigm
=
Network

Representation as a construction network



Obligatoriness and dominance for rival forms can be represented as construction networks.

Nu-drop in a nutshell!

1. Descriptive:

- \emptyset ~ **NU** variation depends on several factors:
 - Inflected forms
 - Aspectual prefixes
 - Voice marker –*sja*

2. Methodological:

- “Tree and forest” model is a powerful tool for analysis of rival forms

3. Theoretical:

- Construction networks not limited to syntax
- Inflectional paradigms are category networks



Rival forms: Wrapping up

- Three case studies:
 - “suffix shift”
 - “nu-drop”
 - “genitive-accusative shift”
- Questions:
 - Factors: What are the factors motivating the choice between rival forms?
 - Method: How can we study rival forms empirically?
 - Theory: What are the implications of rival forms for theoretical linguistics?
- Tentative answers:
 - A cocktail of factors: morphophonology, morphology, semantics, frequency
 - Method: large corpora and statistical analysis
 - Theory 1: must accommodate multiple factors and statistical tendencies
 - Theory 2: morphological paradigm = radial category

