

UIT

NORGES
ARKTISKE
UNIVERSITET

Psycholinguistic studies of Russian aktionsarten

Anastasia Makarova

CLEAR group

(Cognitive Linguistics: Empirical Approaches to Russian)

Slavic languages in the Black Box

Tübingen, 24-26 September 2014





Why aktionsarten?

- Aktionsarten modify verbal meaning by adding quantitative/qualitative characteristics: *spat* 'sleep' → *pospat* 'sleep for a while'
- Include regular meaning-form correspondences: *za* + indeterminate motion verbs = ingressive (beginning of motion)
- Traditionally described as unrelated and very specific phenomena
- We find semantic and morphological overlap between different aktionsarten: *pri-* in comitatives: *pripevat* 'sing along' and diminutive semelfactives: *prikriknut* 'shout slightly once'
- Need a better understanding of the interaction of different types of aktionsarten as parts of the phenomenon of aktionsarten in Russian in general

Aktionsarten provide data for studying the distribution of variants in language and choices speakers make



Why study morphological variation?

- Speakers are making choices each time there are two or more forms expressing a similar meaning
- Factors at play:
 - meaning
 - environment
 - phonology
 - morphology
 - syntax
 - frequency effects
 - ...
- Different relationship between rival forms: synonymy and free variation, allomorphy, ...

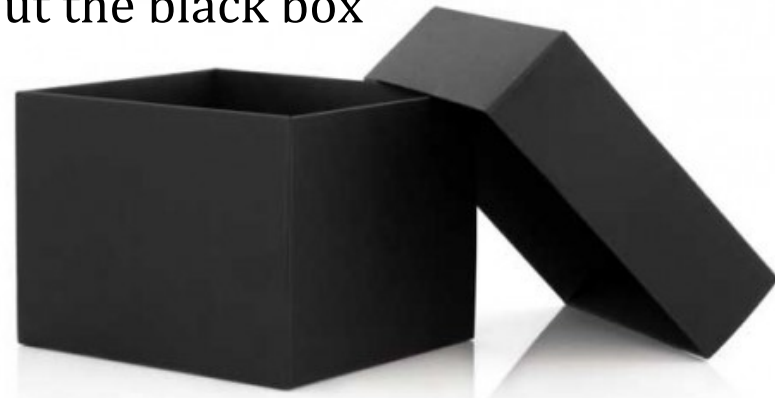
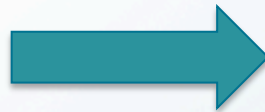


We need to understand the behavior of rival forms in order to understand the form-meaning relationship

How can we study morphological variation?



- Corpus studies
 - Russian National Corpus
 - Facilitate quantitative and qualitative analyses
 - Enable us to put forward hypotheses
- Experiments
 - Enable us to test predictions of the hypotheses
 - Enable us to make assumptions about the black box



Two case studies: attenuatives and semelfactives in Russian

Will corpus findings
be corroborated in
an experiment?

Russian attenuatives

- Refer to actions performed with lower intensity than the activity they are related to, secondary, incomplete (Isačenko 1982, Zaloznjak, Šmelev 2000):
 - *priotkryt* ‘open slightly’, *podpravit* ‘amend slightly’
- Formed via prefixation from verbs with varying semantics:
 - prefixed or unprefixed bases (perfectives and imperfectives)
 - *pri-*: *prikupit* ‘buy some’, *pod-*: *podmorozit* ‘get slightly frosty’, *po-*: *pokurivat* ‘smoke a little bit from time to time’
 - also combinations of prefixes and suffixes: *pri-...-iva-*: *priskulivat* ‘whimper along’, *pod-...-nu-*: *poddaknut* ‘say ditto to’
- A corpus study suggests that *pri-* and *pod-* are not randomly distributed across stems (Makarova 2014):
 - only *pri-*, only *pod-*, both prefixes
- The choice between *pri-* and *pod-* is based on the type of semantic interaction between the stem and the prefix.

Type and token
frequency
taken into
account

Attenuatives: experimental contexts

- We want to confirm that the distribution of two prefixes is not random
- Do corpus data mirror the mental grammars of individual speakers?
- Do responses of the native speakers mimic the RNC data?
 - verbs that only take *pri-*
 - verbs that are attested with *pod-* only
 - verbs attested with *pri-* and *pod-*
- Stimuli: existing contexts from the RNC
 - 59 sentences culled from the corpus
- All verbal prefixes replaced with gaps (164 prefixes, 59 targets):
 - Дедушка __крывает ворота и, __держивая ногой, __пускает меня.
- 105 controls

Cloze-test
tasks

Attenuatives: responses

- 122 participants, 20,008 responses
 - 7,198 for targets, 12,810 for controls
 - 14,062 (70%) matched the original prefixes from the RNC!
 - for targets, 69% responses were attenuatives

Other responses than *pri-* and *pod-* for targets are challenging for statistical analysis

Attenuatives: results

- Verbs that in the RNC are used with *pri-* or *pod-* exclusively:

	Target responses	Non-target responses	Total (<i>pri-</i> and <i>pod-</i>)
<i>pri-</i> in RNC	889 (97.7%)	21 (2.3%)	910 (100%)
<i>pod-</i> in RNC	1044 (99.4%)	6 (0.6%)	1050 (100%)

$\chi^2=9.5$, $df=1$,
 $p<0.001$,
 Cramer's V=
 0.074

- Verbs that in the RNC are attested with both prefixes:

	<i>pri-</i> responses	<i>pod-</i> responses	Total
<i>pri-</i> in RNC	981 (67.7%)	467 (32.3%)	1448 (100%)
<i>pod-</i> in RNC	417 (27.6%)	1091 (72.35%)	1508 (100%)

$\chi^2=474$, $df=1$,
 $p<0.001$,
 Cramer's V=
 0.4

- Contrastive vs. non-contrastive use of prefixes matters
 - 77.4% target responses vs. 60.7% target responses

- Frequency matters (token frequency):

- 95% of matches for high frequent verbs
- 76% of matches for low and average frequent verbs

$\chi^2=95$, $df=1$,
 $p<0.001$,
 Cramer's V=
 0.18



What motivated the choice of attenuatives?

- Lexical context:

	Lexical support	No lexical triggers	Total
<i>attenuatives</i>	782	897	1679
<i>other</i>	560	811	1371

$\chi^2=9.8$, $df=1$,
 $p<0.005$,
Cramer's V=
0.057

- Frequency:

Frequency	Attenuative	Other	Total
<i>high</i>	1414 (82.8%)	294 (17.2%)	1708
<i>average</i>	2192 (81.7%)	492 (18.3%)	2684
<i>low</i>	1826 (65.1%)	980 (34.9%)	2806

$\chi^2=268$, $df=2$,
 $p<2.2e-16$,
Cramer's V=
0.2

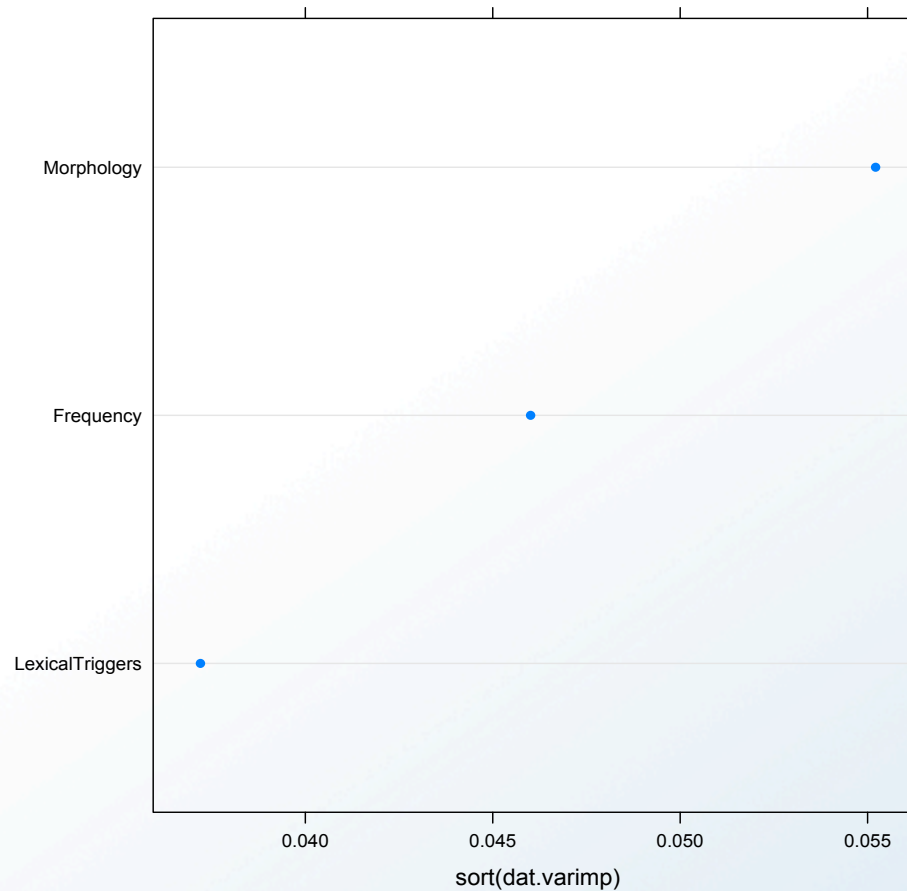
- Morphology:

	Morphological triggers	No morphological triggers
<i>attenuatives</i>	1618	1134
<i>other</i>	212	818

$\chi^2=436$, $df=1$,
 $p<2.2e-16$,
Cramer's V=
0.34

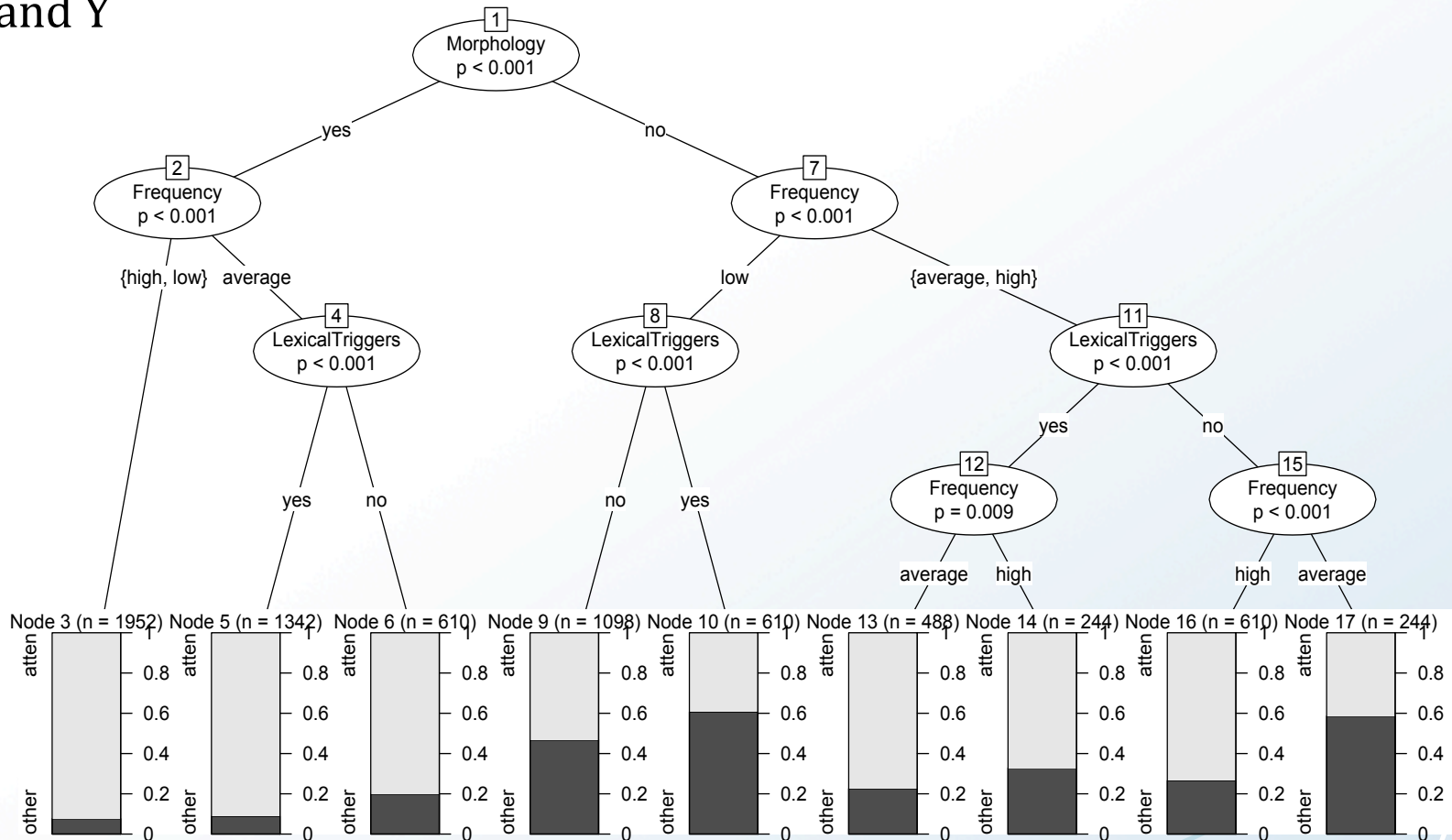
Factors in interaction -1

- Random forest (cforest): measures relative importance of variables in the choice between X and Y



Factors in interaction -2

- Classification tree (ctree): groups predictors and visualizes how well different combinations of predictors account for the choice between X and Y



Attenuatives: summary

- Hypothesis confirmed: the choices of morphological variants are not random, the distribution of *pri-* and *pod-* mimics that in the RNC
- The experiment enabled us to measure the effect of several factors, as well as the relative importance of the factors:
morphology > frequency > lexical context

Chi-squares
with effect
sizes

Ctree,
cforest

No!

Can we apply the same methodology to other aktionsarten?

Russian semelfactives

- Denote single action, select a single cycle in a repeatable series of events (Lomonosov 1755, Townsend 1968, Maslov 1948, 1965, Bondarko 1971, Švedova et al. 1980, Zaliznjak, Šmelev 2000, Janda 2007, Nessel 2013): *maxnut'* 'wave once', *sglupit'* 'be dumb once'
- Formed from verbs denoting a simple physical action, acoustical or optical phenomena by suffixation (-*nu-* / -*anu-*) and prefixation (*s-*)
- A corpus study by Dickey & Janda (2009) suggests that -*nu-* and *s-* have a near-complementary distribution:
 1. sound and impact verbs prefer -*nu-*
movement and behavior verbs use more *s-*
 2. non-productive 1st conjugation uses only -*nu-*
-**ěj* verbs use only *s-*strong tendencies for other morphological classes

Will corpus findings be corroborated in an experiment?



Semelfactives vs. attenuatives

Similarities

- Modifications of verbs
- Variation in morphological markers
- Hypothesis: the choice of the relevant morphological marker is not random

Differences

- Different types of morphemes involved:
 - semelfactives involve prefixes and suffixes
 - attenuatives involve only prefixes
- Hypotheses to be tested are different:
 - semelfactives: semantic and morphological motivation
 - attenuatives: semantic motivation



We cannot use the same type of tests for the two aktionsarten

Pseudo-randomized order for unbiased responses

Semelfactives: experimental contexts

- Two possible factors: morphological class and semantics
- Impossible to analyze independently based on existing verbs
- Nonce-verbs (Berko 1958, Bybee & Prado 1981, Chernigovskaya & Gor 2000, 2001, 2003, Svistunova 2008), made-up contexts:
 - morphological classes of the nonce verbs can be controlled
 - semantic classes of the nonce verbs can be controlled
 - contextual motivation to use semelfactive aktionsart
- Nonce-verbs presented in the context:
 - Птичка тликала все реже и реже, В последний раз и замолчала.
- Existing verbs as controls:
 - Я все равно считаю, что Пашка очень смелый и не дрейфит, То, что он повел себя так тогда, это чистая случайность: любой бы на его месте.
 - 32 targets, 12 controls

Cloze-test tasks for less biased responses

Semelfactives: responses, challenges for data analysis, results

- 63 participants, 2954 responses
 - 2,233 for targets, 721 for controls (12 ungrammatical)
 - semelfactive responses: 1239 (53%)
 - challenge: *-nu-* and *s-* are not mutually exclusive
 - we can study the behavior of semantic classes with respect to *-nu-* and with respect to *s-* and compare different semantic classes
 - several responses per stimulus
- Hypothesis confirmed: semantic classes matter, morphology matters
 - *-aj-* easiest, *-*ěj-* hardest
 - the harder the higher the number of other forms (e.g. when morphology and semantics are in conflict)
- suffixation is more productive
- informants need semantics (contextual support)

t-tests,
p<0.001

Summary



- Studies of Russian aktionsarten: specific and manageable research questions that help understand big theoretical questions
- We need tailored experiments even for seemingly similar types of data
- We do not need sophisticated methods to get reliable results
- Focus on experimental output, secure the possibility of statistical analysis:
 - the output needs to be analyzable
 - data can be summarized in .csv files and analyzed with R, each response is a data point.
- Secure availability of results:
 - for yourself and the research community
 - <http://opendata.uit.no>