

# Rival constructions in word formation

Chiara Naccarato

Linguistic Convergence Laboratory, NRU HSE

November 29<sup>th</sup> 2019

# Rivalry

- ▶ Word synonymy
  - ▶ e.g. quick, fast, swift, rapid, speedy

# Rivalry

- ▶ Word synonymy
  - ▶ e.g. quick, fast, swift, rapid, speedy
- ▶ Grammatical constructions
  - ▶ e.g. Russian decade constructions *ν dvadcatye gody* vs. *ν dvadcatyx godax* ‘in the Twenties’ (i.e. accusative vs. locative construction)

# Rivalry

- ▶ Word synonymy
  - ▶ e.g. quick, fast, swift, rapid, speedy
- ▶ Grammatical constructions
  - ▶ e.g. Russian decade constructions *ν dvadcatye gody* vs. *ν dvadcatyx godax* ‘in the Twenties’ (i.e. accusative vs. locative construction)
- ▶ Word-formation constructions
  - ▶ e.g. agent noun suffixes in Russian: *-tel’*, *-nik*, *-čik*, etc.

# Word synonymy

- ▶ In language “there are no actual synonyms”; within a set of words like “quick, fast, swift, rapid, speedy”, each one of these words “differs from all the others in some constant and conventional feature of meaning” (Bloomfield 1933: 145)
- ▶ “One thing becomes clear once we begin a serious quest for absolute synonyms, and that is that if they exist at all, they are extremely uncommon. (...) There is no obvious motivation for the existence of absolute synonyms in a language, and one would expect that either one of the items would fall into obsolescence, or that a difference in semantic function would develop.” (Cruse 1986: 270)

# Grammatical constructions

- ▶ In Construction Grammar: “Principle of No Synonymy”: “If two constructions are syntactically distinct, they must be semantically or pragmatically distinct” (Goldberg 1995: 67)

# Grammatical constructions

- ▶ In Construction Grammar: “Principle of No Synonymy”: “If two constructions are syntactically distinct, they must be semantically or pragmatically distinct” (Goldberg 1995: 67)
- ▶ So, what about the difference between *ν dvadcatye gody* and *ν dvadcatyx godax* ‘in the Twenties’?

# Grammatical constructions

- ▶ In Construction Grammar: “Principle of No Synonymy”: “If two constructions are syntactically distinct, they must be semantically or pragmatically distinct” (Goldberg 1995: 67)
- ▶ So, what about the difference between *ν dvadcatye gody* and *ν dvadcatyx godax* ‘in the Twenties’?
- ▶ No evidence of semantic or sociolinguistic differentiation (Nesset & Makarova 2018)
- ▶ A leveling process: one form is ousting the other over time (Nesset & Makarova 2018)



# Word formation

This talk:

- ▶ The question of rivalry in **word formation**
- ▶ Different constructions used to form the same type of nouns -> a case study of Russian synthetic compounds with agentive/instrumental meanings
- ▶ Is there absolute synonymy in word formation? Why do we need different constructions if their functions are the same?
- ▶ Look for possible differences:
  - ▶ semantic
  - ▶ diachronic
  - ▶ stylistic
  - ▶ related to **productivity**

# Synthetic compounds

- ▶ Co-occurrence of two word-formation processes
  - ▶ Compounding + suffixation
  - ▶ Compounding + conversion
- ▶ Constructions with **agentive/instrumental** meanings in Russian:
  - ▶ *zakon-o-da-tel'* 'legislator'
  - ▶ *prav-o-zaščit-nik* 'human rights activist'
  - ▶ *gruz-o-perevoz-čik* 'cargo company'
  - ▶ *vs-e-znaj-ka* 'know-all'
  - ▶ *gazon-o-kosi-lka* 'lawnmower'
  - ▶ *jazyk-o-ved* 'linguist'
  - ▶ etc.

# To sort things out

- ▶ RNC word-formation database: over 16,000 compounds retrieved from dictionaries of modern Russian
- ▶ Selection of all synthetic compounds with agentive/instrumental meanings
- ▶ 831 compounds
- ▶ 9 constructions (*-ec*, *-lec*, *-tel'*, *-nik*, *-(š)čik*, *-l'ščik*, *-ka*, *-lka*, suffixless)
- ▶ Analyzed according to different parameters concerning:
  - ▶ the type of bases allowed by each construction
  - ▶ the semantic features of each construction
  - ▶ their frequency in texts belonging to different time periods/genres/registers

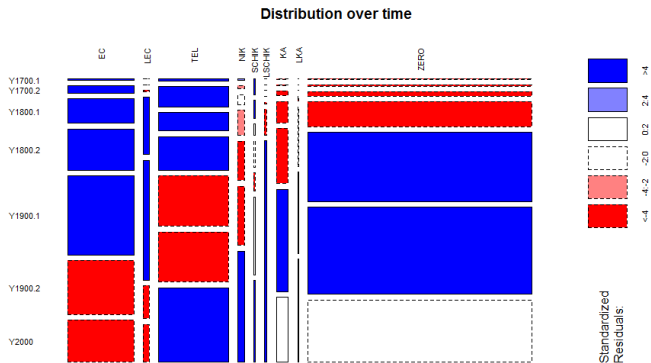
# Differences in the type of bases allowed

- ▶ Non-verbal base (*mjas-o-rub-ka*):
  - ▶ mostly nouns, but also adjectives, pronouns and numerals (e.g. *jasn-o-vid-ec*)
  - ▶ mostly functioning as the theme/patient argument of the base verbs, but other possibilities are allowed (e.g. *kanat-o-xod-ec*)
- ▶ Verbal base (*mjas-o-rub-ka*):
  - ▶ mostly transitive, but intransitive also occur (e.g. *kanat-o-xod-ec*)
  - ▶ mostly imperfective, but occasionally perfective occur too (e.g. *vozduch-o-očisti-tel'*)
- ▶ No remarkable differences among different constructions

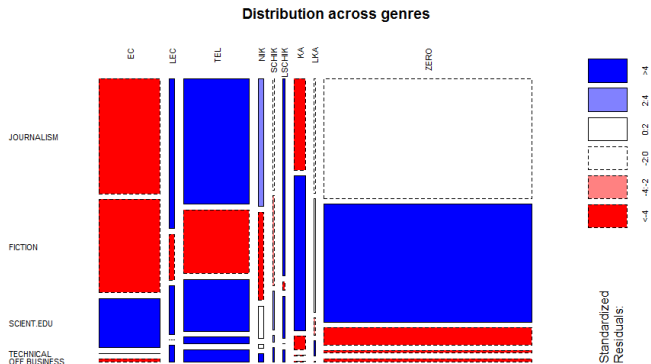
# Semantic differences

- ▶ Remarkable differences among different constructions
- ▶ Certain constructions show a preference for agentive meanings:  
*zakon-o-da-tel'*, *kanat-o-xod-ec*, *granat-o-met-čik*, *jazyk-o-ved*
- ▶ Certain constructions show a preference for instrumental meanings:  
*mjas-o-rub-ka*, *ruk-o-moj-nik*
- ▶ Certain constructions have **only** agentive meanings:  
*zempl-e-vlade-lec*, *stal-e-plavi-l'ščik*
- ▶ Certain constructions have **only** instrumental meanings:  
*gazon-o-kosi-lka*

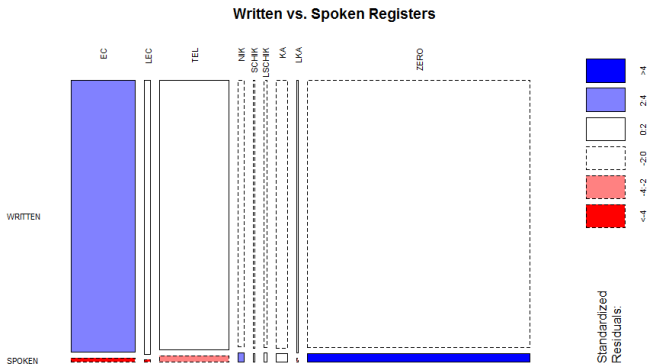
# Diachronic differences



# Stylistic differences



# Stylistic differences





## A case study

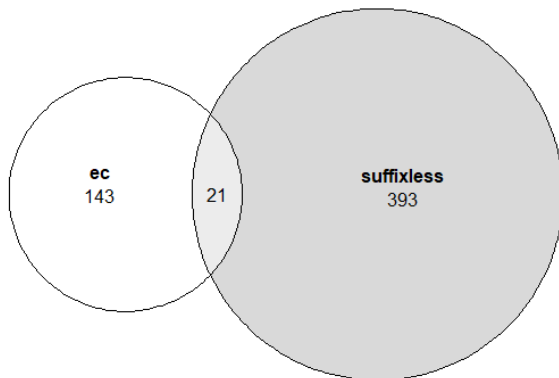
- ▶ Strongest rivalry: *-ec* vs. suffixless
- ▶ Semantically more similar to one another compared to other constructions
- ▶ Both agents and instruments (but a general preference for agents)
- ▶ Apparently, a certain degree of diachronic and stylistic differentiation
- ▶ Less specialized (i.e. technical) meanings compared to other constructions

## A side note

- ▶ The suffixless construction shows some morphonological constraints which do not affect the construction in *-ec*
  - ▶ Only one-syllable unprefixated imperfective verb stems (Tagabileva 2013: 198)
  - ▶ No verb stems ending with a consonant cluster (Sichinava 1999)
- ▶ However, such stems are not so frequent for the construction in *-ec* (e.g. *-deržec*, *-torgovec*) too, and many verb stems are common (or at least possible in principle) for the two constructions

## The data

21 common compounds, e.g. *pravd-o-ljub-ec* vs. *pravd-o-ljub*



which also suggests that the two constructions show a high degree of synonymy (this is not common for other constructions!)

**Table 1:** Common compounds

<i>bogomol</i>	<i>bogomolec</i>	<i>odnodum</i>	<i>odnodumec</i>
<i>verxovod</i>	<i>verxovodec</i>	<i>pravdoljub</i>	<i>pravdoljubec</i>
<i>vodonos</i>	<i>vodonosec</i>	<i>samoljub</i>	<i>samoljubec</i>
<i>vol'nodum</i>	<i>vol'nodumec</i>	<i>svobodoljub</i>	<i>svobodoljubec</i>
<i>duxobor</i>	<i>duxoborec</i>	<i>sebjaljub</i>	<i>sebjaljubec</i>
<i>dušegub</i>	<i>dušegubec</i>	<i>serdceved</i>	<i>serdcevedec</i>
<i>ženoljub</i>	<i>ženoljubec</i>	<i>slastoljub</i>	<i>slastoljubec</i>
<i>žizneljub</i>	<i>žizneljubec</i>	<i>trudoljub</i>	<i>trudoljubec</i>
<i>kitoboj</i>	<i>kitoboec</i>	<i>čelovekoljub</i>	<i>čelovekoljubec</i>
<i>krjučkotvor</i>	<i>krjučkotvorec</i>	<i>čestoljub</i>	<i>čestoljubec</i>
<i>morexod</i>	<i>morexodec</i>		

# Semantic differentiation

- ▶ Приток *богомольцев* постепенно увеличивался, и прежний храм становился тесен. [Христорождественский женский монастырь в Твери // «Журнал Московской патриархии», 2004.02.23]
- ▶ Ненавижу насекомых. Только *богомолы* нравятся. Не знаю почему. Странные они какие-то... [Татьяна Соломатина. Отойти в сторону и посмотреть (2011)]

## But not always...

- ▶ Как же! Может, и есть на свете *правдолюбцы*, только не моя маменька. Да и твои тоже... не слишком склонны рыться в архивах. Или я не права? [Дарья Симонова. Сорванная слива (2002)]
- ▶ Ну и что? Лжецы лгут да лгут и так залгутся, что иногда такую правду выговорят, что никакому *правдолюбцу* не снилась. Так моя мама безумная говорила. А я ей всегда верила. [Юрий Буйда. У кошки девять смертей (2000) // «Новый Мир», 2005]

# So...

- ▶ Semantically very similar
- ▶ Some diachronic and stylistic differences
  - ▶ -ec mostly occurs in written texts and is perceived as archaic, suffixless is more neutral, does not show such stylistic nuances (Tagabileva 2013)
- ▶ Is this enough to explain their coexistence?
- ▶ What about their **productivity**?
- ▶ “Wherever there are alternate processes for expressing the same categories in a language, there are differences in the degree of productivity of the processes” (Bybee 1985: 132)

# What is *productive*?

- ▶ According to (Švedova 1980):
  - ▶ -ec exhibits limited productivity and new compounds are only formed by analogy with existing compounds (e.g. *vertoletonosec* > *bronosenosec*)
  - ▶ the suffixless construction is highly productive both in technical terminology and in everyday language



# What is *productive*?

- ▶ According to (Švedova 1980):
  - ▶ -ec exhibits limited productivity and new compounds are only formed by analogy with existing compounds (e.g. *vertoletonosec* > *bronenosec*)
  - ▶ the suffixless construction is highly productive both in technical terminology and in everyday language
- ▶ But what are the criteria for determining productivity?
- ▶ Is it simply native speakers' intuition?

# What is *productive*?

- ▶ According to (Švedova 1980):
  - ▶ -ec exhibits limited productivity and new compounds are only formed by analogy with existing compounds (e.g. *vertoletonosec* > *bronenosec*)
  - ▶ the suffixless construction is highly productive both in technical terminology and in everyday language
- ▶ But what are the criteria for determining productivity?
- ▶ Is it simply native speakers' intuition?
- ▶ This is a bit impressionistic, we want to be more precise and provide some numerical approximations

# Defining productivity

- ▶ The ability of an affix or of a morphological process to be actively used to produce new words, cf. (Bauer 1983: 18), (Plag 2003: 44), (Booij 2005: 68), (Haspelmath & Sims 2010: 114)
- ▶ (Corbin 1987): qualitative vs. quantitative approaches
  - ▶ *disponibilité* ‘availability’: productivity as a yes/no question; a word-formation process is either available or not
  - ▶ *rentabilité* ‘profitability’: the extent, the degree to which a certain morphological process can be employed to create new words
- ▶ **Profitability**: more focused on the actual use of a certain morphological process, aims at determining the exact value corresponding to the productivity of a morphological process during a certain time span

# Quantitative approaches to productivity

- ▶ Strongly rely on the notions of **type** (V) and **token** frequencies (N)
- ▶ And on the notion of *hapax legomenon* (V<sub>1</sub>), i.e. a word occurring only once in a certain text or corpus
- ▶ The most widespread corpus-based quantitative approach to morphological productivity is represented by the works of Baayen and associates ([Baayen & Lieber 1991](#)), ([Baayen 1992](#)), ([Baayen 1993](#)), ([Baayen 2001](#))
- ▶ Different measures of productivity proposed in the literature

# Type frequency

- ▶ Type frequency (V): the number of types formed with a given affix in a corpus of N tokens
- ▶ Can only indicate that a given morphological process was productive in the past
- ▶ Does not tell us anything about the synchronic availability/profitability of that process in the language

# Potential productivity

- ▶ Calculated by dividing the number of *hapax legomena* ( $V_1$ ) with a given affix by the number of tokens ( $N$ ) with that affix
- ▶ “Estimates the probability of coming across new, unobserved types, given that the size of the sample of relevant observed types equals  $N$ ” (Baayen & Lieber 1991: 809–810)
- ▶ Subject to several criticisms because entails some problematic aspects that must be dealt with

# Potential productivity

- ▶ Inadequate to compare differently-sized corpora: the value of  $P$  is highly dependent on the sample size ( $P$  is a function of  $N$ )
- ▶ However, this problem can be addressed in two ways:
  - ▶ dividing corpora into comparable pieces
  - ▶ resorting to parametric statistical models of frequency distribution known as LNRE (Large Number of Rare Events) models, cf. (Štichauer 2009), (Štichauer 2015), (Efthymiou, Fragaki & Markos 2012), (Naccarato 2016)
- ▶ A more crucial problem: not suitable to all types of data, i.e. what do we do if our data do not contain *hapax legomena*? (Back to this question later)

# Global productivity

- ▶ A function of P and V, represented through a two-dimensional graph with P on the horizontal axis and V on the vertical axis
- ▶ Productive affixes will show large values for P and V, while unproductive affixes will show low values for P and V
- ▶ As admitted by Baayen himself ([Baayen 1992](#): 124), it is difficult to rank different affixes in terms of global productivity due to their disparate positions on the graph



## Hapax-conditioned degree of productivity

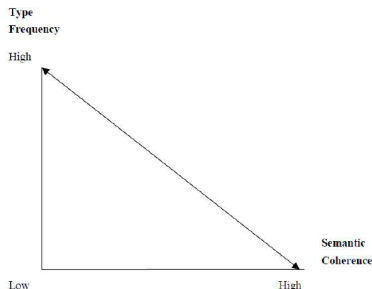
- ▶ Calculated by dividing the number of *hapax legomena* with a given affix by the total number of *hapax legomena* in the corpus
- ▶ “This measure asks ‘What proportion of new coinages use affix A?’ rather than asking ‘What proportion of words using affix A are new coinages?’ It is this latter which seems a more relevant question to ask” (Bauer 2001: 155)

## A side-note on *hapax legomena*

- ▶ Objections: *hapax legomena* are not necessarily new formations
- ▶ But for corpora of many tenths of millions tokens, most hapaxes indeed turn out to be unestablished words (Gaeta & Ricca 2015: 847)
- ▶ If the corpus data are manually checked, *hapax legomena* can be considered as a reliable indicator of productivity (Plag, Dalton-Puffer & Baayen 1999), (Haspelmath & Sims 2010: 130)

## Barðdal (2008)

- Within a constructionist approach to argument structure
- The function of type frequency and semantic coherence and the inverse correlation between the two (Barðdal 2008: 34)



- But how to determine what is semantically coherent?
- Plus, the same problems as for global productivity

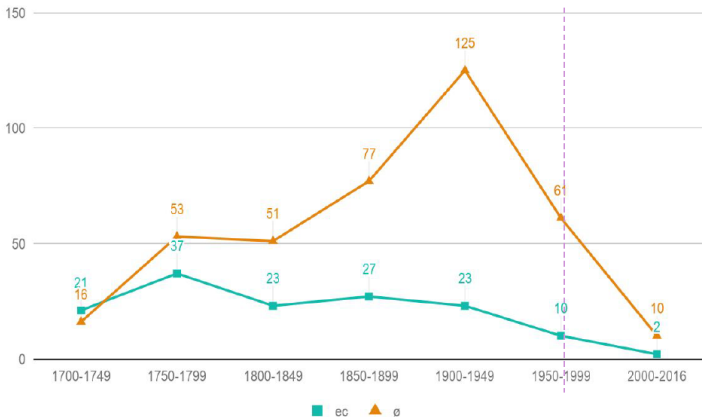
# Diachronic productivity

- The diachronic productivity of word-formation processes can be determined by looking at the number of neologisms (or first occurrences in a corpus) attested in a good historical dictionary or in large corpora that allow for diachronic investigations (Haspelmath & Sims 2010: 130)

## -ec vs. suffixless

- ▶ Russian National Corpus ([ruscorpora.ru/new/](https://ruscorpora.ru/new/)): an adequate source for such investigations, as it includes Russian texts from the 18<sup>th</sup> century to the present day and allows searching for a word and order the results by the creation date of the texts in which this word occurs
- ▶ For each compound I looked for the date of first occurrence in the RNC
- ▶ Then I counted the number of first occurrences for each construction in different time spans (of fifty years each)

## -ec vs. suffixless



## -ec vs. suffixless

- ▶ Opposite trends in terms of diachronic productivity for the two constructions:
  - ▶ almost constant drop for -ec
  - ▶ almost constant rise for the suffixless construction
- ▶ Compounds with -ec: frequently associated with religious concepts, e.g. *bogomolec* ‘pilgrim, devotee’, *ikonoborec* ‘iconoclast’, *psalmopevec* ‘psalmist’, etc.); religious texts make up a large portion of the earlier texts included in the RNC
- ▶ Peak of new occurrences belonging to the suffixless construction in the first half of the 20<sup>th</sup> century: maybe new naming needs during the post-revolutionary epoch?
  - ▶ e.g. *gruppovod* ‘the leader of a group in any work or activity (in the first years of the Soviet authority)’ (Mokienko & Nikitina 1998: 140)

# What about synchrony?

- ▶ The results for diachronic productivity could give us an idea of the current productivity trends
- ▶ But we cannot draw any reliable conclusion on the synchronic productivity of the two constructions based on this type of analysis
- ▶ We might go back to the productivity measures proposed by Baayen and associates
- ▶ But this also entails some problems



# Inadequate data

- ▶ Let's say we want to use the measure of **potential productivity**  
 $P=V_1/N$  to compare *-ec* and the suffixless construction
- ▶ We would have some troubles due to the nature of the data at our disposal
- ▶ Compounds in the RNC word-formation database were not extracted following a bottom-up approach (RNC is not yet provided with a thorough annotation of compounds)
- ▶ They actually are **dictionary data** – when a word is included in the dictionary, this is sign of its institutionalization
- ▶ Highly unlikely that they would be hapaxes in the corpus

## Inadequate data

- ▶ Some hapaxes were actually found for both constructions
- ▶ If we wanted to measure P, we would get something like this:

Table 2: *-ec* vs. suffixless

	Tokens	Hapaxes	Productivity
<i>-ec</i>	31,278	5	0.00016
suffixless	110,288	21	0.00019

- ▶ But a closer look at the data suggests that it might actually be the case that hapaxes are NOT (or not always) new formations, but archaic words which are not used anymore
- ▶ We cannot rely on these results, dictionary data can only prove the past productivity of a word-formation process

# Bottom-up

- ▶ The solution would be to retrieve the data by using a bottom-up approach
- ▶ But how can we retrieve compounds from a corpus if they are not tagged?

## Bottom-up

- ▶ The solution would be to retrieve the data by using a bottom-up approach
- ▶ But how can we retrieve compounds from a corpus if they are not tagged?
- ▶ This would be in principle possible (although highly time-consuming) for compounds in *-ec*: we could look for all words ending in *-ec* and then select compounds from the list
- ▶ But how to find compounds belonging to the suffixless construction?

## Bottom-up

- ▶ The solution would be to retrieve the data by using a bottom-up approach
- ▶ But how can we retrieve compounds from a corpus if they are not tagged?
- ▶ This would be in principle possible (although highly time-consuming) for compounds in *-ec*: we could look for all words ending in *-ec* and then select compounds from the list
- ▶ But how to find compounds belonging to the suffixless construction?
- ▶ Apparently, there is no easy way to do this

## A case study

- ▶ Partially specified constructions (*-ljubec* vs. *-ljub*)
- ▶ To test our hypothesis:
  - ▶ diachronic investigations demonstrate that the productivity of the construction in *-ec* is decreasing over time, which is not the case for the suffixless construction
  - ▶ this suggests that the two constructions might be different in terms of their synchronic productivity
  - ▶ in contexts in which both constructions are possible (e.g. where morphological constraints do not impose the choice of one construction instead of the other), it might be the case one construction is ousting the other in contemporary Russian

# Dictionary vs. bottom-up

Table 3: *-ljubec* vs. *-ljub*

		Types	Tokens	Hapaxes	Productivity
RNC	<i>-ljubec</i>	18	1,194	0	0
database	<i>-ljub</i>	18	565	3	0.005
RNC	<i>-ljubec</i>	83	1,518	40	0.026
bottom-up	<i>-ljub</i>	82	1,280	44	0.034

# Comparing corpora

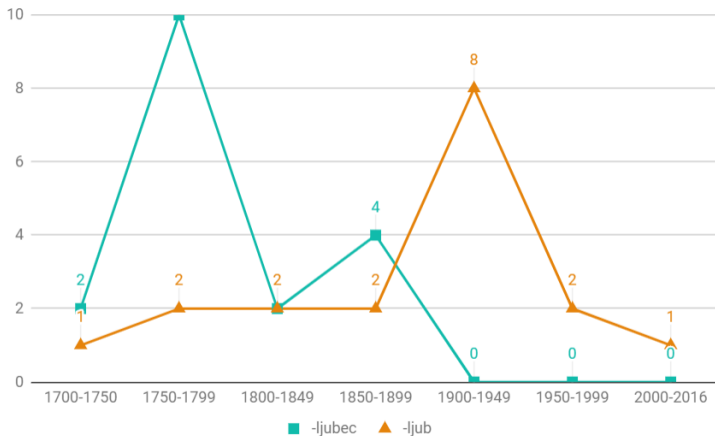
ruTenTen: ~15 billion words; cf. RNC ~600 million words

Table 4: RNC vs. ruTenTen

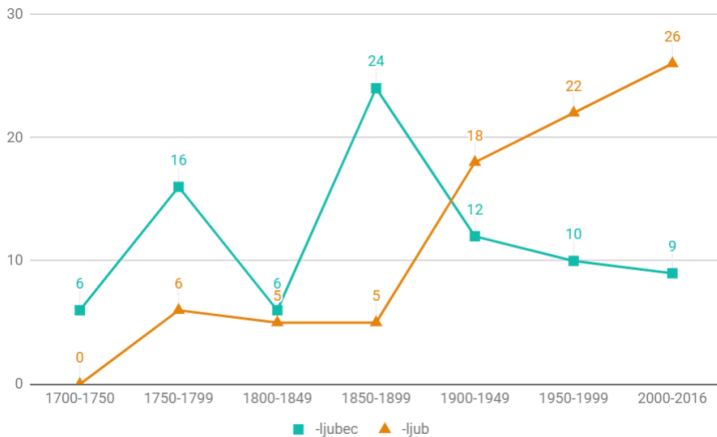
		Types	Tokens	Hapaxes	Productivity
RNC	- <i>ljubec</i>	18	1,194	0	0
database	- <i>ljub</i>	18	565	3	0.005
RNC	- <i>ljubec</i>	83	1,518	40	0.026
bottom-up	- <i>ljub</i>	82	1,280	44	0.034
ruTenTen	- <i>ljubec</i>	116	12,663	44	0.003
	- <i>ljub</i>	428	15,785	236	0.015



## Diachrony (RNC, database)



## Diachrony (RNC, bottom-up)



# Corpus-based approaches to productivity

- ▶ Calculations of productivity strongly rely on the process of data retrieval and on the size of the corpus employed
- ▶ Dictionary data can give us an idea of how productivity evolves over time, but they cannot be used to determine the synchronic productivity of word-formation constructions
- ▶ Measures based on token frequency and *hapax legomena* can be helpful if they are applied to data from big corpora retrieved following a bottom-up approach, otherwise they will give us misleading results

## -ec vs. suffixless

- ▶ The results of different corpus-based investigations of their productivity seem to go in the same direction
- ▶ The productivity of construction in *-ec* is decreasing over time, whereas the suffixless construction is more stable in such respects
- ▶ Synchronically the suffixless construction seems more productive than the construction in *-ec*
- ▶ It might be the case that the suffixless construction is ousting the construction in *-ec* in contexts in which both constructions are in principle possible
- ▶ But we could prove this only based on a case study comparing partially specified constructions

# An alternative method

- ▶ An approach to synchronic productivity which relies on the creativity of native speakers in the production of nonce formations (Fedorova & Naccarato Forth.)
- ▶ This will not give us exact numerical approximations for productivity rates, but will allow to compare different constructions among each other
- ▶ We can control the input, but, obviously, not the output, so more constructions are involved here
- ▶ We will still focus on the results obtained for *-ec* and the suffixless construction to compare corpus-based and experimental approaches

# The survey

- ▶ Spring 2018
- ▶ 60 students (RSUH and NRU HSE)
- ▶ 5 participants of Italian language courses (a control group to check whether knowledge of Italian would affect the easiness of carrying out the task)
- ▶ Task: translate Italian compounds into Russian in the form of (nonce) compounds

## The survey: An example

- ▶ Consider the Italian word *girasole* ‘sunflower’, which is a compound with the structure V+N: *girare* ‘turn’ and *sole* ‘sun’
- ▶ The Russian translation for this compound would be the non-compound *podsolnux*
- ▶ But what if we wanted to create a compound with the same meaning based on the Italian model?
- ▶ Take the corresponding verb and noun bases in Russian (*vertet*’, *povaračivat*’ ‘turn’ and *solnce* ‘sun’) or other suitable bases and build one!

## The survey: An example

- ▶ Consider the Italian word *girasole* ‘sunflower’, which is a compound with the structure V+N: *girare* ‘turn’ and *sole* ‘sun’
- ▶ The Russian translation for this compound would be the non-compound *podsolnux*
- ▶ But what if we wanted to create a compound with the same meaning based on the Italian model?
- ▶ Take the corresponding verb and noun bases in Russian (*vertet*’, *povaračivat*’ ‘turn’ and *solnce* ‘sun’) or other suitable bases and build one!
- ▶ e.g. *solncevert*, *vertisolnce*, *poverni-solnce*, *solncekrut*, *solncevertel’nik*, *solncekrutik*, etc.



# Input

10 Italian compounds: 5 agents + 5 instruments

Table 5: Agents

Compound	Verb		Noun		Translation
lustrascarpe	lustrare	<i>čistit'</i>	scarpe	<i>obuv'</i>	<i>čistil'ščik obuvi</i>
rompicollo	rompere	<i>slomat'</i>	collo	<i>šėja</i>	<i>smel'čak</i>
guardaboschi	guardare	<i>smotret'</i>	boschi	<i>lesa</i>	<i>lesnik</i>
mangiafuoco	mangiare	<i>est'</i>	fuoco	<i>ogon'</i>	<i>baxval</i>
piantagrane	piantare	<i>sažat'</i>	grane	<i>bedy</i>	<i>pridira</i>

Table 6: Instruments

Compound	Verb		Noun		Translation
portachiavi	portare	<i>nosit'</i>	chiavi	<i>ključi</i>	<i>brelok dlja ključej</i>
asciugacapelli	asciugare	<i>sušit'</i>	capelli	<i>volosy</i>	<i>fen</i>
asciugamani	asciugare	<i>sušit'</i>	mani	<i>ruki</i>	<i>polotence</i>
tritattutto	tritare	<i>izmel'čat'</i>	tutto	<i>vse</i>	<i>kuxonnyj kombajn</i>
scolapasta	scolare	<i>slivat'</i>	pasta	<i>pasta</i>	<i>duršlag</i>

# Reactions

- ▶ Knowledge of Italian did not seem to improve the ability to carry out the task
- ▶ Some respondents proposed more than one translation for a single compound
- ▶ Some respondents could not think of a translation for certain Italian compounds
- ▶ We got a total of 695 answers, from which we excluded existing compounds (e.g. *sorvigolova*), non-compounds (e.g. *izmel'čitel'*), and non-verbal constructions (e.g. *ključebrelók*)
- ▶ 661 selected answers: 332 agents and 329 instruments

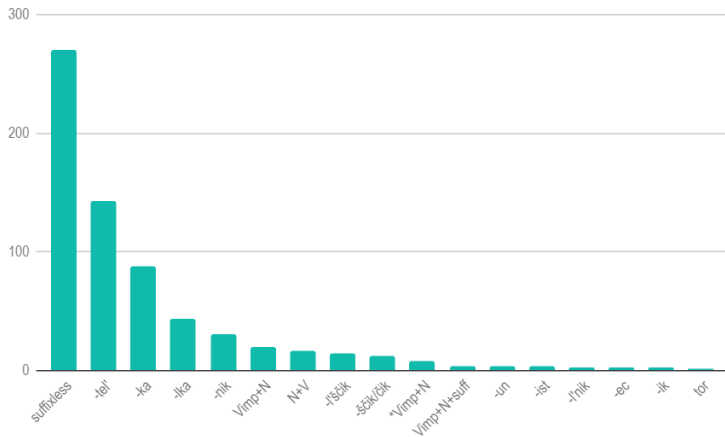
## Proposed translations: agents

- ▶ lustrascarpe: *tuflečist, bašmačist, tufleter, čistibot, ...*
- ▶ rompicollo: *šeelom, lomošej, slomišej, ...*
- ▶ guardaboschi: *lesoxranitel', lesosmotritel', leso(o)xrannik, ...*
- ▶ mangiafuoco: *ogneed, plameed, ognežor, ...*
- ▶ piantagrane: *bedosad, rastobed, bedorastitel', ...*

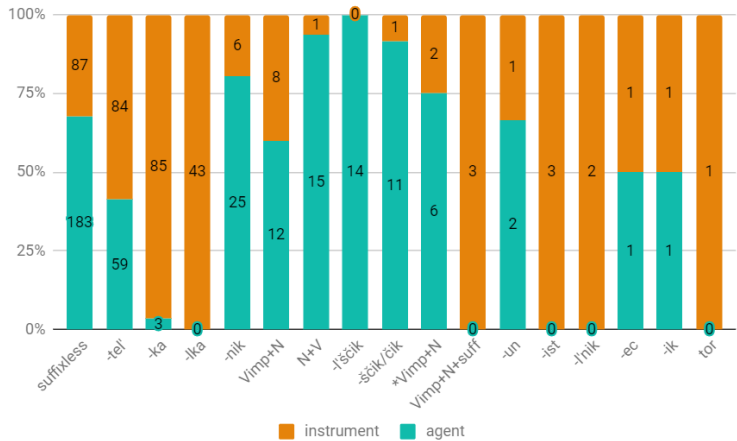
## Proposed translations: instruments

- ▶ portachiavi: *ključenoska, ključenos, ključenositel', ...*
- ▶ asciugacapelli: *volososuška, volososušitel', sušivolos, ...*
- ▶ asciugamani: *rukосуška, rukосуšitel', rukoter, ...*
- ▶ tritatutto: *vseizmel'čitel', vsekrošitel', krošivse, ...*
- ▶ scolapasta: *pastosliv, makaronoslivatel', slivovod, ...*

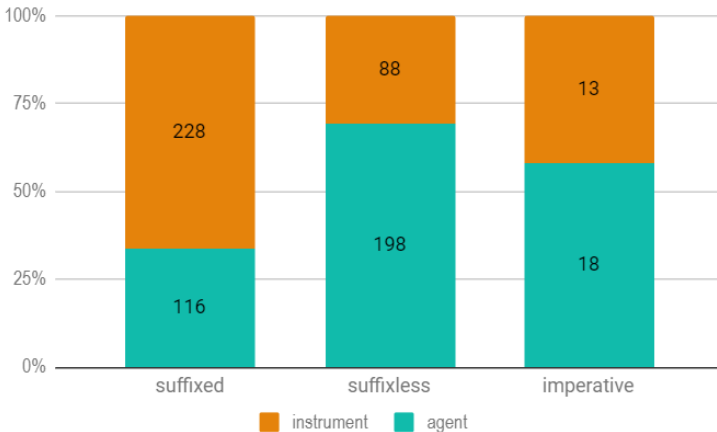
# Constructions



# Semantic distribution



# Semantic distribution





## Results: Summary

- ▶ The **suffixless construction** was chosen in almost 30% of the cases, which is quite remarkable, considering that as many as 17 constructions were used
- ▶ In 40% of the cases it was chosen to build agent nouns, whereas the most frequent suffixed constructions, i.e. *-tel'*, *-ka*, and *-lka*, were chosen more frequently to build instrument nouns
- ▶ The **construction in -ec** was chosen only twice (*ključnosec* and *bašmačistec*), which is sign of its extremely low degree of productivity in contemporary Russian

## A side note: The imperative construction

- ▶ Surprisingly, in 6.5% of the cases constructions with verb stems in first position were proposed
- ▶ In some cases, they show the typical structure of the ‘imperative’ construction of the type *sorvigolova* ‘daredevil’, e.g. *slomajšeja*
- ▶ In other cases, the first base in the compound is not in the imperative form, but a linking vowel is added to the verb stem (e.g. *slomošeja*); sometimes a suffix is also added (e.g. *suševoloska*)
- ▶ Attempts to use this construction are attested, which might suggest that this construction is not so unproductive as pointed out in the literature (?)

# Analogy

- ▶ Apparently, analogy plays a key role in the choice of a specific construction
- ▶ Nonce words are often formed based on existing models:
  - ▶ *tuflečist* > *trubočist*
  - ▶ *ogneed* > *ljudoed*, *mjasoed*, *miroed*, *bukvoed*, ...
  - ▶ *ogneglot* > *živoglot*
  - ▶ *ključenos* > *vodonos*, *medonos*
  - ▶ *bedovod* > *sadovod*, *ovcevod*, ...
  - ▶ *šeelom* > *kostolom*, *dubolom*
  - ▶ *rukoter* > *poloter*

# Analogy

- ▶ Nonce words including suffixes are often based on existing deverbals:
  - ▶ *leso(o)xrannik* > *oxrannik*
  - ▶ *lesoxranitel'* > *xranitel'*
  - ▶ *lesosmotritel'* > *smotritel'*
  - ▶ *vol(so)sušilka* > *sušilka*
  - ▶ *obuvečistil'sčik* > *čistil'sčik*
  - ▶ *tufleočistitel'* > *očistitel'*
  - ▶ *vseizmel'čitel'* > *izmel'čitel'*
  - ▶ *ključederžatel'* > *deržatel'*
  - ▶ *ključeperenosčik* > *perenosčik*
  - ▶ *bedonositel'* > *nositel'*

# Analogy or tendency to analyticity?

- ▶ Frequent use of *-a-* instead of *-o-* (e.g. *pastasliv*)
- ▶ Spelling mistake?
- ▶ Tendency to analyticity?
- ▶ Analogy with compounds formed with non-Slavic components (e.g. *akvapark*)?

## *-ec* vs. suffixless

- Corpus-based and experimental investigations give converging results

**Table 7:** Productivity: *-ec* vs. suffixless

	<i>-ec</i>	suffixless
<b>Diachronic productivity</b>	decreasing	increasing/stable
<b>Synchronic productivity (Corpus methods)</b>	lower	higher
<b>Synchronic productivity (Nonce formations)</b>	almost null	very high

# General observations







- ▶ Absolute synonymy is very rare
- ▶ This also holds for word formation
- ▶ Rival constructions always show some type of differentiation
  - ▶ semantic
  - ▶ diachronic
  - ▶ stylistic
- ▶ If they are very similar both semantically and stylistically, then they will probably be different in terms of their productivity, and one of them will slowly oust the other over time
- ▶ In the case of *-ec* vs. suffixless, several factors are at play: *-ec* being perceived as archaic and mostly belonging to the written register, becomes increasingly unproductive, which might favor the productivity of the functionally similar suffixless construction

## Methodological remarks


- ▶ The choice of a specific method for measuring productivity strongly depends on the nature of the data at our disposal
- ▶ **Dictionary data:** inappropriate for measurement of synchronic productivity; can be used to determine the diachronic productivity of rival constructions, provided that diachronic information is available
- ▶ **Corpus data:** in principle the most appropriate material to be used for measurements of synchronic productivity, but often useless, due to the lack of a thorough annotation for compound words (and, in general, of any type of information regarding word formation)
- ▶ **Nonce formations:** experiments relying on word-formation creativity as a valid alternative for the study of synchronic productivity of word-formation constructions



# References I

- 
- Baayen, Harald R. 1992. Quantitative aspects of morphological productivity. In G. Booij & J. van Marle (eds.), *Yearbook of morphology* 1991, 109–149. Dordrecht: Kluwer.
- 
- Baayen, Harald R. 1993. On frequency, transparency and productivity. In G. Booij & J. van Marle (eds.), *Yearbook of morphology* 1992, 181–208. Dordrecht: Kluwer.
- 
- Baayen, Harald R. 2001. *Word frequency distributions*. Dordrecht: Kluwer.
- 
- Baayen, Harald R. & Rochelle Lieber. 1991. Productivity and english derivation: A corpus-based study. *Linguistics* 29(5). 801–843.
- 
- Barðdal, Jóhanna. 2008. *Productivity. Evidence from case and argument structure in Icelandic*. Amsterdam/Philadelphia: John Benjamins.
- 
- Bauer, Laurie. 1983. *English word-formation*. Cambridge: Cambridge University Press.

## References II

- 
- Bauer, Laurie. 2001. *Morphological productivity*. Cambridge: Cambridge University Press.
- 
- Bloomfield, Leonard. 1933. *Language*. New York: Holt, Rinehart, & Winston.
- 
- Booij, Geert. 2005. *The grammar of words. An introduction to linguistic morphology*. Oxford/New York: Oxford University Press.
- 
- Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam/Philadelphia: John Benjamins.
- 
- Corbin, Danielle. 1987. *Morphologie dérivationnelle et structuration du lexique*. Tübingen: Niemeyer.
- 
- Cruse, Alan D. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.
- 
- Efthymiou, Angeliki, Georgia Fragaki & Angelos Markos. 2012. Productivity of verb-forming suffixes in Modern Greek: A corpus-based study. *Morphology* 22. 515–543.

## References III



Fedorova, Ljudmila & Chiara Naccarato. Forth. Okkazional'noe slovosloženie v russkom jazyke: produktivnost' i analogija [Nonce compounds in Russian: productivity and analogy]. In I. Krapova, S. Nistratova & L. Ruvoletto (eds.), *Studi italiani di linguistica slava: nuove prospettive e metodologie di ricerca*. Venezia: Edizioni Ca' Foscari.



Gaeta, Livio & Davide Ricca. 2015. Productivity. In P. O. Müller, I. Ohnheiser, S. Olsen & F. Rainer (eds.), *Word-formation. an international handbook of the languages of europe*, 842–858. Berlin: Mouton de Gruyter.



Goldberg, Adele. 1995. *Constructions: A construction grammar approach to argument structures*. Chicago: The University of Chicago Press.



Haspelmath, Martin & Andrea D. Sims. 2010. *Understanding morphology*. London: Hodder Education.

## References IV



Mokienko, Valerij M. & Tat'jana G. Nikitina. 1998. *Tolkovyj slovar' jazyka Sovpedii* [*Explanatory dictionary of the language of the Soviet epoch*]. Sankt-Peterburg: Folio-Press.



Naccarato, Chiara. 2016. A corpus-based quantitative approach to the study of morphological productivity in diachrony: The case of *samo*-compounds in Russian. In H. Christ, D. Klenovšak, L. Sönning & V. Werner (eds.), *A blend of MaLT: Selected contributions from the Methods and Linguistic Theories Symposium 2015*, 133–152. Bamberg: University of Bamberg Press.



Neset, Tore & Anastasia Makarova. 2018. The decade construction rivalry in russian: Using a corpus to study historical linguistics. *Diachronica* 35(1). 71–106.



Plag, Ingo. 2003. *Word-formation in English*. Cambridge: Cambridge University Press.

# References V



Plag, Ingo, Christiane Dalton-Puffer & Harald R. Baayen. 1999. Morphological productivity across speech and writing. *English Language and Linguistics* 3(2). 209–228.



Sichinava, Dmitry. 1999. Ograničenija na vtoruju glagol'nuju osnovu v russkix složnyx slovax s nulevym suffiksom [Restrictions on the second verb stem in Russian compounds with the zero suffix]. *Term paper*.



Štichauer, Pavel. 2009. Morphological productivity in diachrony: The case of deverbal nouns in *-mento*, *-zione* and *-gione* in Old Italian from the 13th to the 16th century. In F. Montermini, G. Boyé & J. Tseng (eds.), *Selected proceedings of the 6th Décembrettes*, 138–147. Somerville: Cascadilla Proceedings Project.

# References VI



Štichauer, Pavel. 2015. From emergent availability to full profitability: The diachronic development of the Italian suffix *-zione* from the 16th to the 20th century. In S. Augendre, G. Couasnon-Torlois, D. Lebon, C. Michard, G. Boyé & F. Montermini (eds.), *Proceedings of the Décembrettes. 8th International Conference on Morphology*, 319–326. Toulouse: Université de Toulouse.



Švedova, Natal'ja Ju. 1980. *Russkaja Grammatika* [*Russian Grammar*]. Moskva: Nauka.



Tagabileva, Maria. 2013. Composites denoting nomina agentis in the russian language: distinguishing competing models. *Wiener Slawistischer Almanach* 85. 196–208.