UNIVERSITY OF BERGAMO

School of Doctoral Studies

Doctoral Degree in Linguistic Sciences

XXX Cycle

SSD: L-LIN/01

Compound agent nouns in Russian: A comparison of rival word-formation constructions

Advisor

Chiar.ma Prof.ssa Silvia Luraghi

Doctoral Thesis

Maria Chiara NACCARATO

Student ID 1036191

Academic year 2016/17

Table of contents

List of figures

List of tables

List of abbreviations

A	adjective/adverb; adjectival/adverbial stem
AG	Agent
AN	animate
BIASP	biaspectual
COM	Comitative
HUM	human
INAN	inanimate
INSTR	Instrument
INTR	intransitive
IPFV	imperfective
LOC	Location
LV	linking vowel
MNR	Manner
N	noun; nominal stem
NUM	numeral; numeral stem
PART	participle
PFV	perfective
PRO	pronoun; pronominal stem
REC	Recipient
SUFF	suffix
TH/PT	Theme/Patient
TOP	Topic
TR	transitive
V	verb; verbal stem
V_{imp}	verb in the imperative form
X	word/stem belonging to any word class

List of languages

Bel.	Belarusian
Bul.	Bulgarian
Cze.	Czech
Gr.	Ancient Greek
Ita.	Italian
Lat.	Latin
Lit.	Lithuanian
Mod.Gr.	Modern Greek
OCS	Old Church Slavonic
PIE	Proto-Indo-European
Pol.	Polish
Rus.	Russian
Scr.	Serbo-Croatian
Skt.	Sanskrit
Slo.	Slovene
Spa.	Spanish
Srb.	Sorbian
Svk.	Slovak
Ukr.	Ukrainian

Acknowledgements

This dissertation is the result of a three-year PhD project that has involved, to a greater or lesser extent, many people, to whom I would like to express my gratitude.

First and foremost, I would like to thank my main supervisor, Silvia Luraghi, who has been present at every single stage of this PhD. She has constantly provided thorough readings, insightful comments, and valuable suggestions. But even more importantly, she has taught me what it really means to be a researcher and a teacher, and has been a constant model of integrity and work ethic.

I am also grateful to my other supervisors, Gianguido Manzelli and Nicola Grandi, for their interest to this work and useful discussion. In particular, Gianguido Manzelli has been a constant presence cheering up the working days in Pavia, and an endless source of scientific knowledge.

I could not have carried out this research without the help of Olga Lyashevskaya, who has provided me with the list of compounds from the RNC word-formation database that has constituted the very base of this study. Her help has also been fundamental for carrying out the quantitative analyses described in Chapters 4 and 5, of which she has followed (with great patience) every step.

My study periods at the School of Linguistics of the Higher School of Economics in Moscow have enormously contributed to the outcome of this dissertation. For this opportunity, I have to thank all the researchers and professors working at the HSE who have allowed me to spend, between 2015 and 2017, a total of twelve months in Moscow, in particular Nina Dobrushina and Michael Daniel.

I am deeply indebted to Ljudmila Fedorova, who has spent hours of her time looking at every single compound in my database and reading my chapters. The discussions we have had during these three years are among the most fruitful moments of my study periods in Russia.

I would also like to thank Laura A. Janda and Tore Nesset for having hosted me at the University of Tromsø for two intense weeks between October and November 2016. They have given me the opportunity to discuss my research with the CLEAR group, whose members have given me useful suggestions, in particular Svetlana Sokolova.

I am also grateful to the scholars who have shared and/or discussed their works on compounds with me: Maria Tagabileva, Vjačeslav Terkulov, Aleksandr Barulin, and Aleksej Bogdanov.

I want to thank my linguist colleagues and friends, who have brightened up these three intense years. Chiara Zanchi has shared with me every single step of this PhD. She has become an essential reference point and one of my dearest friends, whom I deeply admire, both personally and professionally. Her abiding passion for linguistics (and for anything she does) makes her a role model for anyone surrounding her. Erica Pinelli has shared with me her love for Russian (and Russia), and has been a constant and fundamental presence during these three years. The time we have spent together discussing both linguistic questions and life matters is invaluable. Samira Verhees has been by my side during all my study periods in Moscow. She has shared with me not only the room at the dorm, but also study and fun moments for almost one third of this PhD. Her curiosity and hunger for knowledge make her one of the most inspiring people I have met in a long time.

Finally, my warmest thanks to my family for having always been there for me and for having supported every choice I made.

Introduction

The aim of this dissertation is to analyze and compare rival agentive constructions within the domain of compounding in modern Russian.

The choice of compounding as the object of this study is due to the almost total lack of investigations on compounds in Russian. In studies on word-formation, compounding usually fades into the background and is seen as a less important (and less productive) process with respect to derivation. However, as I will show, compounding turns out to be a productive word-formation process in modern Russian, and deserves further consideration. This dissertation aims to partially fill this gap in the literature by focusing on one of the most productive group of compound constructions in Russian, i.e. synthetic agentive compounds (e.g. *basn-o-pis-ec* 'fable writer'). By the term "synthetic", I refer to verb-based compounds in which two word-formation processes (compounding and derivation/conversion) occur simultaneously. I use the term "agentive" because the processes of deverbal derivation/conversion embedded in such constructions mostly result in the formation of nouns denoting Agents. However, as will be discussed further (cf. particularly Chapter 3), the meanings emerging from such processes also extend to Non-Prototypical Agents (or Carriers of State) and Instruments.

The decision of restricting the analysis to the specific group of synthetic agentive compounds is due to the intent of providing a thorough analysis of the formal and semantic features of such compounds. Such a fine-grained analysis would have been impossible if the scope of the dissertation would not have been narrowed down to a limited number of compounds. Apart from the descriptive aim, this study also exploits quantitative corpus-based methods to investigate the distribution and productivity of rival constructions giving rise to synthetic agentive compounds, and thus contributes to demonstrating the importance of quantitative investigations in studies on word-formation. Starting from the observation that different constructions are employed with the same functions, I intend to determine what makes these constructions different from one another.

Although the study is restricted to compounds, its results contribute, on the one hand, to a general understanding of the rival constructions employed in modern Russian

to form agent nouns (both compound and non-compound). On the other hand, the analysis of compounds provides further insights into Russian agentive constructions that cannot emerge from studies that deal exclusively with derivation. For instance, as will be shown in Chapter 3, the complete polysemy schema of each construction is often acquired only once compounding is also considered, as some of the possible meanings brought about by agentive constructions might not be attested in derivation. Conversely, some of the meanings found in derived words do not seem to be attested in compounding, which confirms that there is variation between derivation and compounding and that it is worth paying more attention to the less investigated topic of compound agentive constructions.

Chapter 1 is constituted by a first part in which the phenomenon of compounding is discussed in general (Sections 1.1 to 1.4), and a second part that focuses on this word-formation process in Russian (Sections 1.5 and 1.6). First, I introduce the notion of compounding and discuss the main issues concerning the definition and delimitation of the concept of "compound" (Section 1.1). Second, I address the question of the classification of compounds, which has been a major matter of debate in studies on compounding. By resorting to syntactic and/or semantic criteria, different scholars have proposed more or less diverging classifications of compounds that allow grouping them in different types (Section 1.2). One of the most investigated compound types is constituted by synthetic compounds, i.e. verb-based compounds formed through a word-formation process that involves the simultaneous occurrence of compounding and derivation/conversion and gives rise to lexemes in which the non-verbal constituent fulfils a specific semantic role with respect to the verbal constituent. Synthetic compounds are the topic of Section 1.3, while Section 1.4 focuses on the notion of parasynthetic compounds, which could be considered as a subclass of the synthetic type, as the main difference consists in the fact that the latter notion also encompasses noun-based compounds, and thus includes all compounds formed through the simultaneous occurrence of compounding and derivation/conversion. Synthetic and parasynthetic compounds are among the most common compounding patterns in the Slavic languages (cf. Szymanek 2009). The discussion of such constructions appears particularly relevant because, as already mentioned, the compounds analyzed in this study are ascribable to the synthetic class, as they are the result of the simultaneous occurrence of two word-formation processes, i.e. compounding and derivation/conversion. The second part of the

chapter, which deals with the phenomenon of compounding in Russian, first summarizes the existing literature on the topic (Section 1.5), and then reviews all compound types in Russian (Section 1.6). Compounds are first grouped according to their output category (noun, adjective, adverb, and verb), and then according to their internal structure, i.e. the categories of the input elements.

Chapter 2 is devoted to a description of the theoretical framework adopted for this study, and discusses how compounds can be understood in a constructionist approach. First, I summarize the origin and development of the major constructionist theories (Section 2.1), and then I focus particularly on the fundamental tenets of Construction Morphology (Section 2.2), which is a morphological theory developed by Geert Booij that applies the constructionist principles at the word level, and is now one of the most popular theories in studies on word-formation. Within such an approach, the lexicon is organized and represented as a hierarchy of schemas with varying degrees of abstractness, which allows capturing similarities at a higher level, but also subregularities within more general schemas. The hierarchical lexicon also allows capturing the blurriness of the boundary between compounding and derivation by means of partially specified intermediate schemas that are particularly useful to represent compounds formed with affixoids (cf. Section 2.2.2). In Section 2.3, I address again the question of synthetic and parasynthetic compounding, and I show that such processes are satisfactorily described in a constructionist approach based on the notions of schema unification and embedded productivity. In Section 2.4, I propose a constructionist account of Russian synthetic compounds with agentive meanings based on the idea of a hierarchical lexicon.

Chapter 3 deals with the question of the polysemy of agentive constructions. The first part of the chapter focuses on the origin, etymology and use of the suffixes that are employed both in derivation and compounding to form agent nouns in Russian (Section 3.1). Next, I provide a brief cross-linguistic overview of the polysemy of agentive suffixes, and of how such polysemy has been accounted for within different theoretical frameworks, specifically within cognitive, constructionist and diachronic approaches (Sections 3.2 to 3.5). Agentive constructions in the world's languages commonly bring about polysemy patterns that most of the times include Non-Prototypical Agents (which I call "Carriers of State") and Instruments. Other meanings frequently connected to agentive constructions are Place, Event, and Object/Patient. In Section 3.6, I exploit the

analytical tools of Construction Morphology to describe the polysemy of the agentive suffixes that feature in the constructions analyzed in Chapters 4 and 5, i.e. *-ec*, *-lec*, *-tel'*, *-nik*, *-ščik/čik*, *-l'ščik*, *-ka*, and *-lka*. In Section 3.7, I consider the polysemy of another agentive construction that is the result of the combination of compounding and conversion, which I call "the suffixless construction". This construction shows a large polysemy pattern, which covers all the range of meanings covered by the other suffixed agentive constructions.

Chapter 4 illustrates the qualitative and quantitative analyses lying at the core of this research and aiming at the comparison of rival compound agentive constructions in Russian. After a brief description of the Russian National Corpus (henceforth RNC), which has been used to check the data and retrieve examples (Section 4.1), I explain the process of selection of the data employed for the analysis. As discussed in Section 4.2, I have used the RNC word-formation database to compile my own database of compounds, by selecting the agentive constructions that constitute the object of my analysis, for a total of 831 compounds. In Section 4.3, I analyze and compare rival compound agentive constructions based on a number of parameters regarding their formal and semantic features, i.e. the word class and semantic role of the compound's non-verbal element, the transitivity and aspect of the compound's verbal element, the animacy of the compound's referent and the semantics of the compound. The constructions examined show higher or lower variation and more or less similar behaviors among each other according to each of the parameters considered. Given that this comparison involves nine constructions and six parameters, I employ statistical methods to achieve a more detailed and easily interpretable picture of the analysis carried out. In Section 4.4, I show how logistic regression models and random forests can help understand the distribution of rival constructions based on different parameters, and determine what parameters are more relevant to distinguish them from one another. The data are also compared in terms of their distribution over time, and across genres and registers, by resorting to the metatextual information available in the RNC. This part of the analysis is described in Section 4.5, which shows that diachronic and stylistic features are also significant to understand the distribution of the rival constructions analyzed. The analysis of 29 doublets (i.e. pairs of compounds sharing the same lexical bases, but belonging to different constructions), which is shown in Section 4.6, supports the results of the

preceding analyses, and brings out the special type of rivalry existing between the construction in *-ec* and the suffixless construction, which emerges at various stages of this study.

Chapter 5 addresses the question of productivity, which is a fundamental issue in studies on word-formation. Qualitative and quantitative definitions of productivity are discussed in Section 5.1, while Section 5.2 is mainly focused on the possible problems arising when measuring productivity. A fundamental issue to consider when assessing the productivity of certain constructions concerns the nature of the data, that is the source from which the data are extracted, and the characteristics of the corpus employed, as I show in Section 5.3. Diachrony, genres, and registers are also significant factors that need to be taken into consideration when measuring productivity (Sections 5.4 and 5.5). In Section 5.6, I show how the productivity of certain constructions seems to be subject to certain restrictions related to the nature of their components, such as the availability of autonomous deverbals outside compounding and the type of verbal base included in compounds. In Section 5.7, I carry out two case studies comparing the productivity of two pairs of partially specified schemas, which demonstrates once again how the nature of the data and of the corpora employed significantly affects calculations of productivity. Finally, in Section 5.8, I briefly discuss the relation between productivity and polysemy.

This dissertation includes three appendices. Appendix 1 contains the full analysis of the 831 compounds selected, in which each compound is classified according to the six parameters discussed in Chapter 4 (cf. Section 4.3). Appendix 2 includes the information retrieved from the RNC regarding the number of tokens and the first occurrence of each compound in the RNC. Appendix 3 comprises the list of base verbs embedded in compounds belonging to the suffixless construction, which is relevant to the discussion in Chapter 5 (cf. Sections 5.4 and 5.6.2).

1. Compounds in Russian

The present chapter introduces some fundamental notions about compounding in general, and subsequently provides an overview of compound words in Russian. In Section 1.1, I discuss the notion of "compound" in linguistics. In Section 1.2, I deal with the problem of the classification of compounds and discuss the main approaches to this question. Sections 1.3 and 1.4 focus on some particular compound types, i.e. synthetic and parasynthetic compounds, which are relevant to the whole discussion carried out in this study. In Section 1.5, I account for the existing literature on the phenomenon of compounding in Russian. Finally, in Section 1.6, I provide an overview of the different types of Russian compounds.

1.1. What is a compound?

The status and definition of compound words and of the process of compounding have been a matter of debate in many linguistic studies. In different theoretical frameworks, the question of defining what a compound is has been addressed in various and sometimes contrasting ways. In the present section, I review some of the definitions of compounds that have been proposed in the literature.

A first question to deal with when formulating a definition of "compound" consists in defining the status of its constituting elements. The definitions of "compound" that have been proposed in the literature are based on different assumptions and make use of the terms "word", "stem" or "lexeme" to refer to the building elements of compounds, which partly depends on the language under examination.

Marchand (1960: 11) claims that "[w]hen two or more words are combined into a morphological unit, we speak of a compound". Similarly, Fabb (1998: 66) defines a compound as "a word which consists of two or more words". These definitions, holding on the notion of "word", are probably adequate to account for compounds in English (in which words and stems often coincide), but they become problematic when one considers

languages such as Russian, in which compounding is almost never based on full words, but rather on stems.

A safer way to define compounding without risking of excluding languages that behave differently from English is either to specify that the constituents of a compound can be free forms or stems (Olsen 2000: 897), or to resort to the notion of "lexeme", which can be intended as encompassing both words and stems. Bauer (2001b: 695) defines a compound as "a lexical unit made up of two or more elements, each of which can function as a lexeme independent of the other(s) in other contexts, and which shows some phonological and/or grammatical isolation from normal syntactic usage". In 2003, Bauer defines compounding as "the formation of a new lexeme by adjoining two or more lexemes" (Bauer 2003: 40). Similarly, Haspelmath (2002: 85) describes a compound as "a complex lexeme that can be thought of as consisting of two or more base lexemes". Booij (2005a: 75) claims that the defining property of compounding is that "it consists of the combination of lexemes into larger words". Dressler (2006: 24) argues that compounds can be defined as "grammatical combinations of words, that is of lexical items or lexemes, to form new words".

As Lieber & Štekauer (2009: 4–5) acknowledge, two types of issue must be addressed when delimiting the notion of "compound". First, it is not always easy to draw a distinction between compounds and derivatives (the "micro question", in Lieber & Štekauer's terms). Second, in some languages, distinguishing between compounds and phrasal forms can also be problematic (the so-called "macro question").

The first question mainly concerns compounds in which one or both constituents are bound roots, also known as affixoids,[1] i.e. words that "have become similar to affixes in having a specialized meaning when embedded in compounds" (Booij 2010: 57), such as the Dutch suffixoid *-man* 'man' (cf. Booij 2005b: 115) in words such as *bladenman* 'lit. magazines man, magazine seller'. Given the special status of these elements, scholars do not always agree on how to consider words formed with these units, i.e. derivatives or compounds. By adopting a Construction Morphology approach, in which the lexicon is conceived as a hierarchy of schemas and subschemas with different degrees of generalization (cf. Chapter 2, Section 2.2), one can capture the similarity between

[1] Such elements have also been called "semi-affixes" (Marchand 1960, 1967) or "semi-words" (Scalise 1984).

compounding and affixal derivation (which becomes particularly evident when considering compounds based on affixoids) without necessarily drawing a strict line between the two word-formation processes (cf. Booij 2005b, 2009, 2010). Ralli (2010) investigates the question of delimiting compounding and derivation, and discusses the different theoretical approaches to the issue that have been proposed in the literature. The author concludes that the two processes show no clear demarcation and that they should be treated within the same grammatical component, i.e. morphology.

The second question concerns instead the distinction between compounds and phrases, which is not always easy to draw, especially in languages such as English. Different types of criteria (i.e. phonological, morphological, and syntactic criteria) are usually employed when attempting to delimit the phenomenon of compounding from syntactic processes (cf. Lieber & Štekauer 2009: 8–14), but it has been pointed out that "there are hardly any universally accepted criteria for determining what a compound is" (Lieber & Štekauer 2009: 6). For instance, if spelling cannot be taken as a criterion for compoundhood in English, where compounds are sometimes written as single words and sometimes as two words (with or without a hyphen), in Slavic languages compounds are normally spelled as single words. Thus, in the Slavic tradition, spelling is often seen as an important indicator of the word's status (cf. Lieber & Štekauer 2009: 7). However, Lieber & Štekauer note that, "if we acknowledge that the spoken language is primary, and the writing system only an artificial system designed to capture the spoken word, there must clearly be some criteria which lead writers to write a sequence as one word rather than two" (Lieber & Štekauer 2009: 7–8), and spelling cannot be considered as a criterion of compoundhood. At the same time, the unequivocal spelling convention in the Slavic languages makes the "macro question" discussed above less problematic in such languages. Benigni & Masini (2009: 179) list some other important properties that unequivocally delimit compounds from phrases in Russian, i.e. the "bound" status of the compound's constituents, the absence of relational information in compounds (that is, inflection and conjunctions), the presence of linking elements in compounds (usually the linking vowels *-o-* and *-e-*; cf. Section 1.6), and the reversed order of the constituents in compounds (cf., for instance, the phrase *vladelec doma*.GEN 'owner of the house' and the compound *dom**o**vladelec* 'house owner').

1.2. Classifying compounds

A second matter of debate in studies about compounds concerns their classification. In the literature about compounding, indeed, various classifications have been proposed, each one differing from the other, to a greater or lesser extent, in terms of the terminology and of the criteria employed (cf. Scalise & Bisetto 2009: 35).

Most classifications proposed in the literature are based, more or less explicitly, on the traditional Sanskrit classification of compounds. Bloomfield (1933: 235–237) makes a first distinction between copulative compounds (or *dvandva*),[2] such as *bittersweet*, and determinative (subordinative or attributive) compounds (or *tatpuruṣa*), such as *blackbird*. The author further specifies that compounds can either be endocentric, when one of the elements can be considered as the head of the compound (as both *bittersweet* and *blackbird*), or exocentric (or *bahuvrīhi*),[3] when the head is absent from the compound, as in the compound *turnkey*, in which the head element is a verb, but the compound as a whole is a noun.

Fabb (1998) distinguishes: a) endocentric compounds (i.e. compounds that include a head constituent), such as *sneak-thief*, where *thief* is the head; b) exocentric (or *bahuvrīhi*) compounds (i.e. compounds that do not include a head constituent), such as *greenhouse* (a *greenhouse* is not a type of *house*); and c) co-ordinate (or *dvandva*) compounds, in which both constituents share head-like characteristics, such as *student-prince*.

Olsen (2001) adopts the distinction in determinative (e.g. *coffee cup*), possessive (e.g. *greybeard*), and copulative compounds, which broadly corresponds to the classification proposed by Fabb (1998). The author, then, focuses particularly on copulative compounds and, following Fanselow (1981), distinguishes between traditional dvandvas (e.g. Skt. *simhāvyāghnill* 'the lion and the leopard'), which denote two referents, and pseudo-dvandvas (e.g. *lion-leopard*), which denote one referent only. The latter type is sometimes defined "appositive" (cf. Haspelmath 2002 and Booij 2005a).

[2] A Sanskrit term referring to those compounds in which neither element is subordinated to the other.
[3] A type of compound in the Sanskrit grammar, in which the referent is denoted by specifying a certain property that he/she/it possesses. The term *bahuvrīhi* literally means 'much rice', but denotes a rich man. This term is usually employed to refer to exocentric possessive compounds of the type *redhead*, which denote an entity by specifying one of its properties (such as that of having a red head, i.e. red hair), but in some cases the term is used to encompass all exocentric compounds, and not only possessive.

Bauer (2001b) also follows the traditional Sanskrit classification in: a) *tatpuruṣa* (also known as determinative/endocentric) compounds (also including the subtype known as *karmadhāraya*, i.e. attributive/appositive compounds), such as *blackbird*; b) *dvandva* (also known as copulative/aggregative/coordinative) compounds, such as *Alsace-Lorraine*; and c) *bahuvrīhi* (also known as possessive/exocentric) compounds, such as *egghead*. The author also adds the class of synthetic (also known as verbal/verbal-nexus/deverbal/secondary) compounds, such as *truck driver*, which will be discussed more in detail in Section 1.3.

Haspelmath (2002) classifies compounds in: a) endocentric, such as *lipstick*; b) exocentric, such as *redhead*; c) affix compounds, such as *green-eyed*; d) coordinative compounds, such as the Korean compound *elun-ai* 'adult and child'; and e) appositive compounds, such as Spa. *poeta-pintor* 'poet-painter'.

A similar classification is given by Booij (2005a), who classifies compounds in: a) endocentric, such as Ita. *capostazione* 'station master'; b) exocentric, such as Ita. *portalettere* 'postman, lit. carry letters'; c) *bahuvrīhi*, such as *baldhead*; d) *dvandva*, such as Skt. *candrādityāu* 'the moon and the sun'; and e) appositive compounds, such as *blue-green*.

Grandi (2006) distinguishes between coordinate compounds (*composti di coordinazione*) and hierarchical compounds (*composti gerarchici*), the latter including both subordinate and attributive/appositive compounds. The author also further distinguishes between cases in which the relation between the two members of the compound is marked, and whether it is marked morphologically or syntactically. So, for example, in Mod.Gr. *xartofakela* 'stationery, lit. notepaper envelopes', the coordination is marked morphologically by the linking element *-o-*, while in Ita. *punto e virgola* 'lit. dot and comma, semicolon' the coordination is marked syntactically by the conjunction *e* 'and'.

Scalise & Bisetto (2009) refine their previous classification of compounds (cf. Bisetto & Scalise 2005) and introduce a new proposal including different levels. At a first level, compounds are classified in subordinate, attributive/appositive and coordinate. Subordinate compounds are then classified in ground compounds, in which there is no verbal root, such as *windmill* (endocentric) and Ita. *sottoscala* 'basement, lit. under stairs' (exocentric), and verbal-nexus compounds, such as *bookseller* (endocentric) and

pickpocket (exocentric). Attributive compounds can also be either endocentric, such as *blue-eyed*, or exocentric, such as *redskin*. Appositive compounds are mainly endocentric, such as *swordfish*. Finally, coordinate compounds can be both endocentric, such as Spa. *poeta-pintor* 'poet-painter', or exocentric, such as *mother-child*.

As I will show in Chapter 2 (cf. Section 2.2), by applying a constructionist approach, compounds can be classified in the form of a hierarchical lexicon, which allows taking into consideration both the form and the semantics of compounds, accounting for high-level and low-level regularities, and grouping compounds according to such regularities. In my view, this type of classification provides a more rigorous organization of the data, and allows avoiding the problem of forcibly placing a compound in a certain group, which might be too narrow or too broad.

1.3. Synthetic compounds

A particular class of compounds which is often kept separate from the others is constituted by so-called synthetic compounds. In the present section, I discuss the most important accounts of synthetic compounds that have been proposed in the literature.

Bloomfield (1933: 231) defines synthetic compounds as "[c]ompounds with special features of word-formation". The author exemplifies his idea of a synthetic compound by showing the compound *bootblack*, in which we find the agentive suffixless form *-black* (from the verb *to black*) appearing only in compounds, whereas outside compounding we only find the agent noun *blacker* (e.g. *blacker of boots*). Compounds of the type *meat-eater* are instead considered as "semi-synthetic" (Bloomfield 1933: 232) because the deverbal *eater* also exists outside compounding, but in compounding the suffix *-er* is supposed to be added to the whole phrase *eat meat*, i.e. compounding and derivation take place simultaneously.

Within formal theoretical frameworks, these compounds are usually referred to as "verbal" (cf. Roeper & Siegel 1978; Selkirk 1982), "deverbal" (cf. Di Sciullo 1992, 2005), or "secondary" (cf. Scalise 1994).

Roeper & Siegel (1978) describe the process giving rise to verbal compounds (that is, from the authors' perspective, those compounds formed with a verbal affix, such as -

er, *-ing*, and *-ed*) as a process of lexical transformation based on a verb and its subcategorized complements. Verbal compounds are distinct from root compounds in that root compounds (i.e. words such as *crystal-clear*, *deep-fry*, *air-condition*, etc.) do not include derivational affixes and are not necessarily based on a verbal element (Roeper & Siegel 1978: 206). Moreover, while the meaning of root compounds can be completely unpredictable, verbal compounds are described as predictable and compositional in meaning, and extremely productive (Roeper & Siegel 1978: 206–207). According to the First Sister (FS) Principle proposed by the authors, "[a]ll verbal compounds are formed by incorporation of a word in first sister position of the verb" (Roeper & Siegel 1978: 208), that is, the non-head of the verbal compound must be a word which can appear immediately after the verb in a corresponding verb phrase, which means that it must be its direct object (cf., for instance, **peace-thinking* vs. *peacemaking*).

Selkirk (1982) defines verbal compounds as "endocentric adjective or noun compounds whose head adjective or noun (respectively) is morphologically complex, having been derived from a verb, and whose nonhead constituent is interpreted as an argument of the head adjective or noun" (Selkirk 1982: 23), as in *cake baker* or *housecleaning*.

Di Sciullo (1992, 2005) regards deverbal compounds as compounds that include a predicate-argument relation. In particular, Di Sciullo (1992) argues that argument saturation inside compounds follows the Thematic Hierarchy proposed by Grimshaw (1990: 8), which is represented in (1):

(1) (Agent (Experiencer (Goal/Source/Location (Theme))))

In other words, argument selection in deverbal compounds must follow this hierarchy (from right to left), which means that the argument saturated inside the compound must be lower in the hierarchy compared to any other argument found outside the compound. This would explain why in most deverbal compounds the non-head element has the role of Theme, which is lower in the Thematic Hierarchy. However, while Grimshaw (1990) claims that the first (subject) argument cannot be saturated within a compound, Di Sciullo (1992) argues that both English and Italian show cases in which the first argument is saturated inside the compound, such as in *taxi-driver* and in Ita. *tagliacarte* 'lit. cut paper,

paper cutter'. While in English compounds such as *taxi-driver* the first argument is supposed to be saturated by the agentive suffix *-er*, the absence of an overt suffix with this role in Italian leads the author to propose the presence of *pro*, a nominal head attributing the syntactic category to the compound (Di Sciullo 1992: 67).

Kiefer (1993) observes that argument selection in synthetic compounds cannot be based on the hierarchy proposed by Grimshaw and bases his observations on Hungarian data, showing in particular that "relevance rather than thematic hierarchy seems to be the main organizing principle of argument selection" (Kiefer 1993: 46) and that the Thematic Hierarchy somehow reflects the scale of relevance. Let us consider, for instance, the Russian synthetic compound *očevidec* 'eye witness' (2). Although the base verb *videt'* 'see' in principle would allow the incorporation of an element with the role of Theme, which would refer to the event seen by the Agent, in this case it is the Instrument (i.e. the eyes) that has been incorporated as first constituent of the compound, which is probably due to the fact that what is relevant here is that the event has been seen directly, i.e. with one's own eyes.

(2) *oč-e-vid-ec* 'eye witness'
eye-*e*-see-*ec*

Scalise (1994: 145) attributes the following properties to synthetic compounds: a) they have a nominal or adjectival head which derives from a verb; b) the non-head is an argument of the head; c) the thematic role of the non-head is that of Theme or Patient; and d) they show transparent semantics.

In their classification of compounds, both Fabb (1998) and Bauer (2001b) reserve a special position to synthetic compounds, which are described as compounds whose head elements are derived from verbs through affixation. Haspelmath (2002) defines a synthetic compound as a "nominal compound whose dependent noun fills an argument position in the head's valence" (Haspelmath 2002: 223). In Booij (2005a, 2009), synthetic compounding is described as a word-formation process that involves the simultaneous occurrence of compounding and derivation and gives rise to compounds in which the non-head constituent fulfils a specific semantic role with respect to the verbal constituent.

As I will show in Chapter 4, the accounts of synthetic compounds that have been summarized in the present section are not satisfying when one analyzes this type of compounds in a language like Russian. First, the generalization put forward by many scholars that such compounds show nominal first constituents playing the semantic role of Theme or Patient with respect to the verbal base can be confirmed as a major tendency, but not as the only possibility. Indeed, there is great variation in terms of the word class and semantic role played by the first constituent in Russian synthetic compounds, as the data presented in Chapter 4 (cf. Section 4.3.2) will demonstrate. Moreover, most scholars only refer to synthetic compounds as the result of the combination of compounding and derivation, whereas cases of conversion are not taken into consideration. Remarkably, the combination of compounding and conversion results in one of the most productive synthetic constructions in Russian (cf. Chapter 3, Section 3.7), and deserves further investigations. Finally, in my view, the notion of head in such compounds is problematic because we do not always find an autonomous noun resulting from deverbal derivation/conversion outside compounding (e.g. *očevidec* 'eye witness' vs. **videc* 'one who sees'). As I will argue in Chapter 2 (cf. Section 2.3), a constructionist approach to the analysis of synthetic compounds allows accounting for such constructions in terms of schema unification, without necessarily resorting to the notion of head.

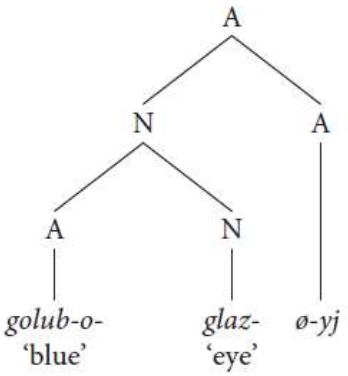
1.4. Parasynthetic compounds

Bisetto & Melloni (2008) and Melloni & Bisetto (2010) identify a separate class of compounds, which they term "parasynthetic", following Serrano Dolader (1995). By this term, the authors refer to compounds resulting from "a word-formation process consisting in the merger of two lexical stems (forming a non-attested compound) with a derivational affix" (Bisetto & Melloni 2008: 233). It is important to point out that, differently from the notion of "synthetic compound" as it has been proposed by various scholars (cf. Section 1.3), the notion of "parasynthetic compound" as intended by Bisetto & Melloni (2008) and Melloni & Bisetto (2010) is not limited to cases of verb-based compounding, i.e. the second constituent of the compound is not necessarily verbal, but it can also be nominal. Parasynthetic compounding gives rise to compounds, such as *able-bodied*, in

which neither the second constituent plus the suffix form an autonomous word outside compounding (**bodied*), nor the two constituents without the suffix form an existing compound (**able-body*), and thus it seems that the two word-formation processes take place simultaneously. The authors support their analysis with Slavic data, as in Slavic languages this seems to represent the core pattern of compounding. Even in cases in which the second constituent plus the suffix form an autonomous form outside compounding, as in the compound adjective *železnodorožnyj* 'railway-related' (in which *dorožnyj* is an autonomous adjective meaning 'related to the road'), the suffix seems to take scope over the complex base *železnaja doroga* 'lit. iron road, railway' (Bisetto & Melloni 2008: 239; cf. also Benigni & Masini 2009: 176).

Bisetto & Melloni (2008) point out that, while cases such as *železnodorožnyj* can be nicely accounted for also in a constructionist framework (cf. Booij 2005b, 2009, 2010; Gaeta 2006, 2010), within which these compounds are analyzed as the result of schema unification (i.e. the unification of the compounding schema and the derivation schema, cf. Chapter 2, Section 2.3), other Slavic cases pose a challenge for the constructionist analysis. In particular, the authors refer to cases such as the Russian compound adjective *goluboglazyj* 'blue-eyed'. In compounds of this type, there is no overt derivational suffix, that is, the second constituent *glaz-* 'eye' is directly attached to the adjectival inflection *-yj*[4] (cf. also Szymanek (2009: 475) for the same compound types in Polish).

(3)



[4] The ending *-yj* is the nominative singular masculine form for adjectives ending with a hard consonant. The corresponding ending when the preceding consonant is soft is *-ij*. The feminine hard and soft endings are *-aja* and *-jaja*, while the neuter hard and soft endings are *-oe* and *-ee*.

The sequence **glazyj* 'eyed', then, is not only inexistent, but also impossible, because Russian denominal adjectives cannot be formed without the addition of a suffix. To explain such cases, Bisetto & Melloni (2008: 248) postulate the existence of an unexpressed suffix deriving an adjective from a complex base, as shown in (3).

According to Bisetto & Melloni (2008: 248), this pattern would arise in cases in which "the nominal stem refers to a salient inalienably possessed body part of humans and animals". Chovanová & Štichauer (2014), by contrast, argue that the constraint at play is not that of inalienable possession, but rather the "Principle of Integrated Meronymy", which is formulated as follows: "A part of an individual allows for an assertion holding for the whole referent (in its entirety) on the basis of the Principle of Integrated Meronymy. What enables the part to characterize the whole is the fact that the relevant properties must be equally salient or valid for the whole as well" (Chovanová & Štichauer 2014: 150). For further discussion on this topic (i.e. inalienable possession vs. integrated meronymy), see Chapter 2 (Section 2.3).

The other problematic case for a constructionist analysis according to Bisetto & Melloni (2008) is constituted by compounds such as the Russian compound noun *glubokomyslie* 'profundity of thought', which they analyze as a sequence of an adjectival stem (*glubok-* 'profund'), a linking element (*-o-*), and a nominal stem (*mysli-* 'thought) followed by the neuter inflection (*-e*). Such cases are seen as problematic because, according to the authors, these compounds also lack an overt suffix, and thus the inflectional marker itself would be responsible for the gender change, which makes this noun neuter. Thus, they postulate "the presence of a nominalizing suffix associated to the neuter gender and taking scope over the pre-modified N (or V) stem" (Bisetto & Melloni 2008: 242). However, this analysis cannot be considered correct, as these formations do include a suffix, i.e. the suffix *-i(j)-e*, which is employed in parasynthetic formations, both non-compound, such as *bessmertie* 'immortality', from *bez smerti* 'without death' (cf. Townsend 1980: 151), and compound, such as *glubokomyslie* 'profundity of thought', from *glubokaja mysl'* 'profound thought' (cf. Townsend 1980: 202).

As I will show in Chapter 2 (cf. Section 2.3), such cases are not really problematic to fit in a constructionist approach, and the question of the inexistence of certain building blocks of such compounds outside compounding is overcome by the concept of "schema unification". In Section 2.3, I will further discuss synthetic and parasynthetic compounds,

the issue of the subsequentiality/simultaneity of the derivation and compounding processes in such compounds, and I will show how such phenomena can be better accounted for in a constructionist approach.

1.5. State of the art of the study of compounds in Russian

The present section reviews the existing literature on Russian compounds. A first general observation in this respect is that studies dealing with word-formation in Russian usually pay little attention to compounding, which often fades into the background if compared to other word-formation processes such as prefixation and suffixation. Indeed, Russian shows an extremely rich derivational system, and prefixation is highly productive in verbal word-formation, while suffixation is particularly productive in nominal word-formation. Within this picture, compounding is often seen as a marginal phenomenon in Russian word-formation, and has been scarcely investigated. However, as I will show throughout this study (cf., particularly, Chapter 5), compounding turns out to be a productive word-formation process in contemporary Russian, and would deserve a more central place in studies about Russian word-formation.

The Academic Grammar of Russian (henceforth AG-80)[5] includes a consistent chapter devoted to word-formation (§§ 191–1110), whose authors are Vladimir Vladimirovič Lopatin and Igor' Stepanovič Uluchanov. The word-formation processes described are organized according to the output category of the words formed: a) nouns (§§ 208–608); b) adjectives (§§ 609–790); c) verbs (§§ 791–975); d) adverbs (§§ 976–1031), and e) other parts of speech (§§ 1032–1110). Each of these sections includes a subsection devoted to compounds: a) compound nouns (§§ 550–587); b) compound adjectives (§§ 754–775); c) compound verbs (§§ 961–963); and d) compound adverbs (§§ 1026–1029). The description of compound words provided by Vladimir Vladimirovič Lopatin within the word-formation section of the AG-80 is probably still one of the most comprehensive accounts of Russian compounds, and includes comments both on formal and semantic aspects of compounding in Russian, and on its productivity.

[5] AG-80 is the common abbreviation used to refer to Švedova, Natal'ja Ju. (ed.) 1980. *Russkaja Grammatika*. Moskva: Nauka.

Townsend (1980) provides an overview of word-formation processes in Russian, and his study is mainly devoted to affixation processes. The author observes that "[t]hough less important than suffixation, combination plays an important role in nominal word-formation, and *compound words* (the result of combination) comprise a number of numerous and productive types" (Townsend 1980: 201). Nevertheless, only a small part of the description provided by Townsend (1980) is dedicated to compounds, and the discussion is mainly concerned with formal aspects, i.e. the types of constituent which take part in compounding processes. This account is by no means exhaustive and systematic, and it is almost not concerned at all with the semantic aspects of compounding.

Barulin & Ajchenval'd (1988) adopt a morpheme-based approach to the analysis of complex words, within which compounds are also included. The main goal of this paper consists in a classification proposal of complex words according to the type of syntactic/semantic relation existing between their constituents, i.e. between the two stems in compounds, between the stem and the affix in derived words. However, the theoretical aim of this study prevents a thorough analysis of compound types.

Janko-Trinickaja (2001)[6] analyzes the features of synchronic word-formation in Russian and devotes a small section (pp. 342–375) of her study to compounds. The account provided by the author is more a theoretical discussion regarding the problem of delimiting compounds from derived words and phrases (cf. Section 1.1), and is more concerned with formal properties of compound words rather than with their semantic features.

Zemskaja (2005, 2011) investigates the functions and mechanisms of Russian word-formation. Compound words are analyzed by Zemskaja (2005: 44–50) mainly based on data taken from dictionaries of neologisms produced in the Eighties. This report, which is not intended to provide an exhaustive account of compounding processes in Russian, aims instead at giving an idea of the most productive compounding patterns based on contemporary data. In Zemskaja (2011), the account of Russian compounds mainly follows the AG-80, and compounds belonging to different word classes are discussed separately. However, this account is by no means as comprehensive as the one

[6] The book was written in 1972, but published only in 2001.

included in the AG-80, and it just reports on some of the compounding patterns available in contemporary Russian.

In the works on Russian word-formation by Uluchanov (2008, 2015), compounding is not treated systematically, but it is just sporadically discussed as one of the productive mechanisms of contemporary word-formation within the more general discussion aiming at providing a theoretical explanation of the functioning of the word-formation system in Russian.

Šanskij (2010) devotes a small section of his work on Russian word-formation to compounding (pp. 269–275), and discusses some of the most common compounding patterns, without providing a systematic account of all the possible types of compound in Russian.

To my knowledge, there is just a small number of studies which are entirely devoted to the phenomenon of compounding in Russian. One such case is constituted by the works of Vasilevskaja (1962, 1968). These studies, however, mainly discuss the theoretical status of compound words and their relationship with phrases, whereas the discussion of specific types of compounds is organized based on some of the most frequent compound constituents, which cannot provide a complete picture of compounding in Russian.

Molinsky (1973) examines compound noun formations in Russian and, particularly, their relationship with the phrases from which they are supposed to originate. Although the major part of this work deals with so-called *složnosokraščennye slova* 'abbreviated compounds', the second part of the book dedicated to thematic[7] compound formations (pp. 109 ff.) provides a systematic account of different types of compounds in Russian, though limited to compound nouns, i.e. excluding compound adjectives, adverbs and verbs.

An innovative description of compounding in Russian is contained in the study carried out by Terkulov (2008a), in which compounds are analyzed and classified by adopting an onomasiological perspective. Following this approach, compounds are seen as linguistic structures that represent knowledge, and the word-formation model is seen as a formula of regular transposition of knowledge about objects, processes, and

[7] The author employs the term "thematic" to refer to compounding patterns in which the two constituents of the compound are conjoined by a compound marker, i.e. a linking element (Molinsky 1973: 17).

characteristics into words. The focus here is on the transformation process from the base unit on which the compound is based to the compound itself. Accordingly, compounds are classified in different types according to their relationship with the base unit (a phrase or a word). Although the model proposed by Terkulov (2008a) is intended to embrace all the possible types of compounds in Russian, the main object of investigation is represented by compounds deriving from a transformation process that the author names *univerbalizacija*, i.e. a process by which the compound formed shows a total identity of meaning with its base unit, as in *belobokij* 'having white hips' (from the phrase *s belymi bokami* 'with white hips'), differently from cases such as *belobiletnik* 'white card holder' (from the phrase *belyj bilet* 'white card'), in which the semantics of the compound is the result of the semantics of the base unit plus the semantics of the formant, i.e. the agentive suffix *-nik*.

Benigni & Masini (2009) offer a descriptive survey of compounding in Russian. Their paper deals with theoretical questions, such as the delimitation of compounds with respect to derived words and phrasal lexemes (cf. Section 1.1.), and includes a classification proposal of Russian compounds based on the output category (i.e. nouns, adjectives, adverbs, and verbs), and on the type of syntactic relationship between the compound's constituents (i.e. subordinate, attributive, and coordinate compounds), following the classification system proposed by Bisetto & Scalise (2005).

A number of recent studies investigate some specific types of compounds in Russian. Terkulov (2008b) examines *bahuvrīhi* compounds, i.e. adjectival possessive compounds such as *goluboglazyj* 'blue-eyed', while Terkulov (2013) focuses on compounds deriving from a process defined by the author as *psevdouniverbalizacija*, i.e. a process by which compounds are created based on existing models rather than from a specific phrase, such as the compound *vertolet* 'helicopter', which was created based on the compound *samolet* 'airplane'.

Bogdanov (2011) investigates the semantics and syntax of deverbal adjectives, and devotes a consistent part of his work to compound adjectives of the type *derevoobrabatyvajuščij* 'wood processing', *social'noznačimyj* 'socially significant', and so on. The syntactic and semantic properties of such compounds are analyzed based on data from the RNC and from the Russian web.

Fedorova (2004, 2014a) addresses the question of the interpretation of the semantics of Russian compounds, focusing on a number of archaic compounds, whose interpretation is not straightforward (e.g. *rotozej* 'scatterbrain). Fedorova (2007, 2008, 2012) examines compounds showing the structure $[V_{imp}\text{-}N]_N$, such as *sorvigolova* 'daredevil', in which we find the imperative form of the verb *sorvat'* 'tear off' followed by the noun *golova* 'head', which is not inflected. The author discusses both the form of such compounds (concluding that the interpretation of the first verbal constituent as an imperative form is the preferred option) and their semantics, and compares such constructions to similar constructions in other languages, such as English, some Romance languages (i.e. French, Spanish, and Italian) and other Slavic languages (i.e. Polish and Bulgarian). Fedorova (2010, 2013) focuses on the semantics of coordinate compounds (e.g. *severo-zapad* 'North-west') and paired words (e.g. *gusi-lebedi* 'geese-swans'). Fedorova (2014b) examines compounds including a first-position numeral constituent (e.g. *desjatièтažnyj* 'ten-floor'), and discusses both the form and the semantics of such compounds. Fedorova (2015) investigates Russian adjectival compounds expressing inalienable possession (cf. also Sections 1.4 and 2.3), i.e. lexemes such as *belolicyj* 'having a white face', which describe a certain characteristic of a human being or an animal by referring to one of his/her/its body parts having a certain property.

Kapatsinski & Vakareliyska (2013) investigate Russian right-headed $[N\text{-}N]_N$ compounds containing recently borrowed English or German stems. By carrying out quantitative investigations based on corpus data, the authors conclude that the productivity of such constructions in Russian is restricted to the occurrence of specific head nouns.

Tagabileva (2013) compares two competing compounding constructions giving rise to agent nouns in Russian, i.e. the construction formed with the suffix *-ec*, e.g. *morechodec* 'seaman', and the suffixless construction, e.g. *morechod* 'seaman'. In her paper (based on data from the RNC), the author shows that, although the suffixless construction seems to be slowly ousting the older construction in *-ec*, morphonological constraints on the suffixless construction prevent the complete extrusion of the *-ec* construction. Moreover, the archaic forms in *-ec* are still used in written texts and show high-register features that the suffixless construction does not have, which also contributes to the survival of the construction in *-ec* (cf. Chapter 4, Section 4.6).

To my knowledge, quantitative investigations of compounds in Russian are almost completely absent. Kapatsinski & Vakareliyska (2013) and Tagabileva (2013) represent some of the rare examples of quantitative studies on compounding in Russian.

Although the current stage of the annotation of corpora of contemporary Russian does not allow carrying out specific corpus-based investigations of compounds (cf. Sections 4.2 and 5.2), the big data that are currently available could be employed to carry out quantitative analyses which could give us a more comprehensive picture of the phenomenon of compounding in Russian. As I will show in Chapters 4 and 5, in this study, I propose such type of quantitative analyses to compare rival constructions giving rise to compound agent nouns in Russian.

1.6. Types of compounds in Russian

The present section provides an overview of compounds in Russian. The description of the different compound types is organized based on the category of the output lexemes, i.e. compound nouns (1.6.1), compound adjectives (1.6.2), compound adverbs (1.6.3), and compound verbs (1.6.4). This account is based on data from the literature on Russian word-formation listed so far (and particularly on the AG-80), as well as from the word-formation database of the RNC employed for the analyses carried out in Chapters 4 and 5.

Compounds in Russian are formed through the combination of two elements which can be either full words (4), or stems (5). Šanskij (2010: 269) distinguishes the two situations by employing the notions of *slovosloženie* 'word combination' and *osnovosloženie* 'stem combination'. The word-combination process gives rise to what Šanskij (2010: 269) calls *složnye sostavnye slova*, which are spelled with a hyphen, whereas the stem-combination process gives rise to *složnye slitnye slova*, in which the two constituents are connected through a linking element.

(4) *divan-krovat'* 'sofa-bed' (from *divan* 'sofa' and *krovat'* 'bed')

(5) *knig**o**ljub* 'bibliophile' (from *kniga* 'book' and *ljubit'* 'love')

In some cases, the compound's constituents are affixoids (cf. Section 1.1), such as compounds formed with the constituent *sam-o-* 'self' (6) (cf. Chapter 2, Section 2.2.2.).

(6) *sam**o**let* 'airplane' (from *sam* 'self' and *letat'* 'fly')

In so-called "abbreviated compounds" (cf. Molinsky 1973: 17), an abbreviated form is attached to a full word (7) or to another abbreviated form (8).

(7) *zarplata* 'wage' (from *zarabotnaja plata*)

(8) *chimfak* 'chemistry department' (from *chimičeskij fakul'tet*)

Leaving aside abbreviated compounds, in what follows I will only focus on compounds *stricto sensu*.

As mentioned above, when compounds are not spelled through a hyphen, a linking element is added after the first stem. This linking element[8] is usually either *-o-* (as in examples (5) and (6) or *-e-*. The linking element *-e-* appears after palatalized and unpaired consonants (AG-80 § 585), as in (9), although some exceptions are also found (see, for instance example (16), in which we find the linking element *-o-* even though the preceding consonant is palatal).

(9) *ogn**e**tušitel'* 'fire extinguisher' (from *ogon'* 'fire' and *tušit'* 'extinguish')

More rarely, other linking elements are employed, as in examples (10) to (14).

(10) *dv**u**člen* 'binomial' (from *dva* 'two' and *člen* 'member')

(11) *dv**uch**letnyj* 'two-years old' (from *dva* 'two' and *let*[9] 'years')

(12) *tr**ech**mernyj* 'three-dimensional' (from *tri* 'three' and *mera* 'measure')

(13) *pjat**i**borec* 'pentathlete' (from *pjat'* 'five' and *borot'sja* 'fight')

[8] These linking elements are defined "interfixes" in the AG-80 (cf. § 585). Barulin (2017) prefers to interpret such elements as suffixes belonging to the first constituent of compound words.
[9] The singular form of the word "year" is *god*. The form *let* is the genitive plural.

(14) *seb**ja**ljubec* ‘self-lover’ (from *sebja*[10] ‘oneself’ and *ljubit’* ‘love’)

While the AG-80 considers the highlighted elements as other possible interfixes appearing in compounds (cf. AG-80 § 585), Benigni & Masini (2009: 174) argue that in such cases there is no linking element, but the first constituent is inflected, i.e. it is a wordform. Thus, examples (11) to (13) would include the genitive inflection of the numeral, while the form *dvu-* in example (10) is considered as an allomorph of *dvuch-* (Benigni & Masini 2009: 174). Finally, the form *sebja-* in (14) would be considered as the genitive or accusative form of the pronoun.

In the classifications and analyses carried out in the rest of this study, I will follow the approach adopted by the AG-80, and will consider such elements as interfixes. Throughout this work, I will use the abbreviation LV, which stands for “linking vowel”, because compounds showing a vowel as linking element represent by far the most common situation. However, as we have just seen, there are cases in which the linking element is not constituted by a vowel alone, as in (11) and (12).

The compounding process can be either pure or combined with suffixation. In the Russian literature on word-formation, the two processes are usually referred to as *čistoe složenie* ‘pure compounding’ (15) and *složenie v sočetanii s suffiksaciej* ‘compounding combined with suffixation’ (16) (cf. AG-80 § 203; Zemskaja 2005: 44–50, 2011: 283–286, 303–304; Uluchanov 2008: 29, 62; Šanskij 2010: 269–275).

(15) *les**o**park* ‘forest park’ (from *les* ‘forest’ and *park* ‘park’)

(16) *basn**o**pis**ec*** ‘fable writer’ (from *basnja* ‘fable’ and *pisat’* ‘write’)

Thus, while in the case of pure compounding the first word or stem is attached (through a hyphen or a linking vowel) to a second autonomous word, and the syntactic relationship between the two constituents can either be coordinate or subordinate/attributive, in the case of compounding combined with suffixation, two stems are attached through a linking vowel, the second stem is followed by a suffix, and the syntactic relationship between the two stems can only be subordinate/attributive, but never coordinate. The latter situation

[10] *Sebja* is the genitive and accusative form of the reflexive pronoun “oneself”, which lacks the nominative form.

(i.e. compounding combined with suffixation) is usually considered as including only cases in which the second stem plus the suffix do not form an autonomous derivative, or in which, if such autonomous derivative exists, the suffix takes scope over the complex base rather than over the second base alone. Thus, according to the AG-80, compounds such as *dačevladelec* 'cottage owner' (from *dača* 'cottage' and *vladet'* 'own, possess') are considered as examples of pure compounding because *vladelec* 'owner' is an existing word outside compounding, and the compound is considered as the result of the combination of the word *dača* and the word *vladelec*, i.e. corresponding to the phrase *vladelec dači* 'owner of the cottage' (cf. AG-80, § 550). On the other hand, compounds of the type *ognetušitel'* 'fire extinguisher' (from *ogon'* 'fire' and *tušit'* 'extinguish') are considered as the result of a process of compounding combined with suffixation because the verbal base plus the suffix do not form an existing derivative outside compounding (**tušitel'*) (cf. AG-80, § 557). However, considering that it is not always easy to determine whether the compounding process is pure or not for compounds including a suffix (cf. Chapter 2, Section 2.3), in what follows I will always consider such cases as the result of a compounding process combined with suffixation. Furthermore, I include in this group of non-pure compounds those constructions in which compounding is combined with conversion (cases of zero suffixation, according to the AG-80, cf. §§ 579–584).

1.6.1. Compound nouns

In the present section, I classify Russian compound nouns based on their input elements. As already mentioned, the classification proposed is based on data from the literature on Russian word-formation, as well as from the word-formation database of the RNC employed for the analyses conducted throughout this study. Abbreviated compounds such as the examples in (7) and (8) will be excluded from the classification, as they are not compounds *stricto sensu*.

1.6.1.1. Noun-based compound nouns

Noun-based compound nouns show a noun or a nominal stem[11] as the main element of the compound. Within this type of compound nouns, several subtypes can be identified, depending on the word class of the other element of the compound and on the presence/absence of a suffix, as shown below.

[N-N]$_N$

The [N-N]$_N$ type is the only type of noun-based compounds made up of two nouns separated by a hyphen and showing the head element on the left. The noun in second position usually specifies a property of the head noun, as in *gorod-geroj* 'hero city' (17).

(17) *gorod-geroj* 'hero city'
city-hero

[N-LV-N]$_N$

In the [N-LV-N]$_N$ type, the first element is a nominal stem followed by a linking vowel and a noun. The second noun is the head constituent of the compound, while the first element constitutes a further specification of the head element, as in *maslozavod* 'oil/dairy factory' (i.e. a specific type of factory) (18), or *ovcebyk* 'musk ox' (i.e. a specific type of ox) (19).

(18) *masl-o-zavod* 'oil/dairy factory'
oil/butter-*o*-factory

(19) *ovc-e-byk* 'musk ox'
sheep-*e*-ox

[11] In the classification proposed here, and throughout this study, I will use the abbreviation N to refer both to nouns and nominal stems, without specifying every time whether we are dealing with a full word or a stem. However, the status of such elements is easily determinable by the phonological context that follows: whenever N is followed by a linking vowel (LV) or by suffix (SUFF), we are dealing with a stem, whereas when N is followed by a hyphen or by nothing, we are dealing with a full word. The same applies to the other possible parts of speech (A, PRO, NUM, V).

[N-LV-N-SUFF]$_N$

The [N-LV-N-SUFF]$_N$ type shows the same pattern as the preceding type, but the second constituent of the compound is not a full word, but a nominal stem, followed by a suffix. The suffixes that can be embedded in such compounds are numerous and confer a wide range of different meanings to the compounds formed. For example, the compound *sudoremontnik* 'worker repairing ships' in (20) is formed with the agentive suffix *-nik*, whereas the compound *narodovlastie* in (21) is formed with the suffix *-i(j)-e*, which gives rise to an abstract noun meaning 'democracy'.

(20) *sud-o-remont-nik* 'worker repairing ships'
ship-*o*-renovation-*nik*

(21) *narod-o-vlast-ie* 'democracy'
people-*o*-power-*ie*

[A-LV-N]$_N$

Noun-based compound nouns can also show an adjectival stem as first member. The adjectival stem defines a property of the head noun in second position. An example of the [A-LV-N]$_N$ type is constituted by the compound *suchofrukty* 'dried fruit' shown in (22).

(22) *such-o-frukty*[12] 'dried fruit'
dried-*o*-fruit

[A-LV-N-SUFF]$_N$

In the [A-LV-N-SUFF]$_N$ type, the pattern is identical to the previous one, but the second constituent of the compound is a nominal stem followed by a suffix, as exemplified by the compound *krasnoderevščik* 'carpenter using fine wood' shown in (23).

(23) *krasn-o-derev-ščik* 'carpenter using fine wood'
red-*o*-wood-*ščik*

[12] *Frukty* is the plural form of *frukt*.

[PRO-LV-N]$_N$

The [PRO-LV-N]$_N$ type can be exemplified by the compound *samozaščita* 'self-defense' (24), in which we see the reflexive pronoun *sam-* 'self' in first position specifying the noun *zaščita* 'defense'.

(24) *sam-o-zaščita* 'self-defense'
self-*o*-defense

[PRO-LV-N-SUFF]$_N$

The same pattern with the addition of a suffix is instantiated by the compound *samočuvstvie* 'state of health' (25), formed with the suffix *-i(j)-e*.

(25) *sam-o-čuvstv-ie* 'state of health'
self-*o*-feeling-*ie*

[NUM-LV-N]$_N$

Numeral stems are also found in first position, modifying the noun in second position, as in the compound *dvuokis'* 'dioxide' (26).

(26) *dv-u-okis'* 'dioxide'
two-*u*-oxide

[NUM-LV-N-SUFF]$_N$

The compound *pjatiugol'nik* 'pentagon' shown in (27) exemplifies the [NUM-LV-N-SUFF]$_N$ pattern, which is similar to the preceding one, but shows a nominal stem in second position followed by a suffix.

(27) *pjat-i-ugol'-nik* 'pentagon'
five-*i*-angle-*nik*

1.6.1.2. Verb-based compound nouns

Compound nouns can also be verb-based, i.e. the main element of the compound is a verbal stem. In most cases, the verbal stem is found in second position, followed by a nominalizing suffix. Such compounds belong to the class of synthetic compounds, which have been discussed in Section 1.3, and show a relationship of argumental nature between the verbal stem and the non-head constituent, i.e. the non-head constituent is an argument of the base verb.

[N-LV-V-SUFF]$_N$

The [N-LV-V-SUFF]$_N$ type is by far the most common pattern of verb-based compounding. As for noun-based compounds, the range of suffixes employed is quite varied. An example of this pattern is constituted by the compound *basnopisec* 'fable writer' (28), in which the stem of the noun *basnja* 'fable' constitutes the object of the verb *pisat'* 'write', and the suffix *-ec* confers the agentive meaning to the compound. Another example of such pattern is the compound *rybolovstvo* 'fishing' (29), in which the stem of the noun *ryba* 'fish' is the object of the verb *lovit'* 'catch', and the suffix *-stvo* gives rise to an activity noun.

(28) *basn-o-pis-ec* 'fable writer'
fable-*o*-write-*ec*

(29) *ryb-o-lov-stvo* 'fishing'
fish-*o*-catch-*stvo*

[A-LV-V-SUFF]$_N$

Adjectival/adverbial bases are also common in first position when compounds are verb-based. As will be discussed in Chapter 4 (see Section 4.3.1 for further discussion), in such compounds it is not always easy to determine whether the first element should be treated as an adjective or as an adverb, which is why the two categories are kept together in this classification. This pattern can be illustrated by the compounds *dolgožitel'* 'long liver' (30) and *čistopisanie* 'calligraphy' (31).

(30) *dolg-o-ži-tel'* 'long liver'
long-*o*-live-*tel'*

(31) *čist-o-pisa-nie* 'calligraphy'
clear-*o*-write-*nie*

[PRO-LV-V-SUFF]$_N$

The [PRO-LV-V-SUFF]$_N$ type can be exemplified by the compound *vsederžitel'* 'the almighty' (32), in which the first element is constituted by the stem of the pronoun[13] *ves'* 'all'. The verbal base of the compound is constituted by the verb *deržat'* 'hold', followed by the agentive suffix *-tel'*.

(32) *vs-e-derži-tel'* 'the almighty'
all-*e*-hold-*tel'*

[NUM-LV-V-SUFF]$_N$

Though more rarely, numeral stems are also found as first constituents in verb-based compounds, as illustrated by the compound *odnoljub* 'one-woman man' (33).

(33) *odn-o-ljub-ø*[14] 'one-woman man'
one-*o*-love-*ø*

[V$_{imp}$-N]$_N$

A particular case of verb-based compound nouns is constituted by compounds showing a verb in the imperative form as first element and a non-inflected noun as second element, such as the compound *sorvigolova* 'daredevil' (34), in which the constituent *sorvi-* is the imperative form of the verb *sorvat'* 'take off'.

[13] I use the term "pronoun" to refer to all those items that are defined as *mestoimenie* 'pronoun' in the Russian tradition. The label *mestoimenie* 'pronoun' is used to refer to personal pronouns, e.g. *ja* 'I', *ty* 'you', *on* 'he', etc. (which are included under the label *mestonimenija-suščestvitel'nye* 'pronouns-nouns'), but also to possessive pronouns and adjectives, e.g. *moj* 'my/mine', *tvoj* 'your(s)', etc., demonstrative pronouns and adjectives, e.g. *ètot* 'this', *tot* 'that', etc., indefinite pronouns and adjectives, e.g. *ves'* 'all', *kto-to* 'someone', etc. (so-called *mestoimenija-prilagatel'nye* 'pronouns-adjectives').

[14] Cases of conversion are represented with the symbol *-ø* for practical needs and reasons of graphical representation, as discussed in Chapter 3 (cf. Section 3.7).

(34) *sorvi-golova* ‘daredevil’
take_off-head

[V_{imp}-N-SUFF]$_N$

The same pattern can also embed a suffix, as in *vertišejka* ‘wryneck’ (35), in which the suffix *-ka* is added after the base of the noun *šeja* ‘neck’.

(35) *verti-šej-ka* ‘wryneck’
spin-neck-*ka*

[V-LV-N-Ø]$_N$ / [N-LV-V-Ø]$_N$

Another particular case is constituted by “reversible” compounds of the type *lizobljud*/*bljudoliz* ‘lickspittle’, in which the main verbal element can be found either on the left (36) or on the right (37).

(36) *liz-o-bljud-ø* ‘lickspittle’
lick-*o*-plate-*ø*

(37) *bljud-o-liz-ø* ‘lickspittle’
plate-*o*-lick-*ø*

1.6.2. Compound adjectives

The present section provides an overview of compound adjectives in Russian. Compounding in Russian is very productive for the creation of adjectives (cf. Zemskaja 2005: 44, 2011: 303), which is also confirmed by the figures corresponding to the word-formation database of the RNC, which includes almost 17,000 compounds, more than half of which are adjectives (almost 12,000).

1.6.2.1. Adjective-based compound adjectives

A first subclass of compound adjectives is constituted by compounds that are the result of a pure compounding process, i.e. the second constituent of the compound is an autonomous adjective or participle. In such cases, the first constituent of the compound can either be an adjectival or a nominal stem.

[A-LV-A]$_A$

The [A-LV-A]$_A$ type is exemplified by the compound *černo-belyj* 'black and white' (38), in which the first adjectival stem *čern-* 'black' is followed by a linking vowel and attached (through hyphenation) to the adjective *belyj* 'white'. The compound *temno-sinij* 'dark blue' shown in (39) belongs to the same pattern. However, while *černo-belyj* 'black and white' shows a relationship of coordination between the two constituents of the compound, *temno-sinij* 'dark blue' shows an attributive relationship, as the first-position adjective *temnyj* 'dark' modifies the second-position adjective *sinij* 'blue'.

(38) *čern-o-belyj* 'black and white'
black-*o*-white

(39) *temn-o-sinij* 'dark blue'
dark-*o*-blue

[N-LV-A]$_A$

When the first element of the compound is constituted by a nominal stem, we have compounds of the type [N-LV-A]$_A$, such as *trudosposobnyj* 'able-bodied' (40).

(40) *trud-o-sposobnyj* 'able-bodied'
work-*o*-able

[A-LV-PART]$_A$

In the [A-LV-PART]$_A$ type, the second constituent of the compound is a participle modified by an adjectival/adverbial stem, as in *glubokouvažaemyj* 'honorable' (41).

(41) *glubok-o-uvažaemyj* ‘honorable’
profound-*o*-respectable

[N-LV-PART]$_A$

Alternatively, the head participle can be modified by a nominal stem, giving rise to the [N-LV-PART]$_A$ pattern, as illustrated by compounds such as *vodoottalkivajuščij* ‘water-repellent’ (42).

(42) *vod-o-ottalkivajuščij* ‘water-repellent’
water-*o*-repellent

1.6.2.2. Noun-based compound adjectives

Another consistent group of compound adjectives is represented by those compounds whose structure is constituted by a nominal stem in second position followed by an adjectivalizing suffix. The first-position stem of the compound in such cases can be adjectival, nominal, or numeral.

[A-LV-N-SUFF]$_A$

The [A-LV-N-SUFF]$_A$ type can be exemplified by compounds such as *prostodušnyj* ‘simple-minded’ (43), in which the nominal stem in second position is followed by the adjectivalizing suffix *-n-* (plus the inflectional ending *-yj*), which gives rise to a possessive compound meaning ‘having a simple mind’. In some cases, the adjectivalizing suffix can be absent, as in *goluboglazyj* ‘blue-eyed’ (44), in which the nominal stem *glaz-* ‘eye’ is directly attached to the adjectival inflectional ending *-yj* (cf. Sections 1.4 and 2.3).

(43) *prost-o-duš-n-yj* ‘simple-minded’
simple-*o*-soul-*n-yj*

(44) *golub-o-glaz-ø-yj* ‘blue-eyed’
blue-*o*-eye-*ø-yj*

[N-LV-N-SUFF]$_A$

We find a first-position nominal stem specifying the second nominal stem in compounds such as *zmeevidnyj* 'snake-like' (45). Here too, the adjectivalizing suffix can be absent, as in *lastonogij* 'pinniped' (46).

(45) *zme-e-vid-n-yj* 'snake-like'
snake-*e*-aspect-*n-yj*

(46) *last-o-nog-ø-ij* 'pinniped'
fin-*o*-leg-*ø-ij*

[NUM-LV-N-SUFF]$_A$

Finally, we can find numeral stems modifying the main nominal constituent, as in *trechètažnyj* 'three-floor' (47) and *dvuglavyj* 'two-headed' (48).

(47) *tr-ech-ètaž-n-yj* 'three-floor'
three-*ech*-floor-*n-yj*

(48) *dv-u-glav-ø-yj* 'two-headed'
two-*u*-head-*ø-yj*

1.6.2.3. Verb-based compound adjectives

Verb-based compound adjectives show a second-position verbal stem followed by an adjectivalizing suffix. The first stem of the compound in such cases can be nominal, adjectival/adverbial, pronominal, and numeral.

[N-LV-V-SUFF]$_A$

The [N-LV-V-SUFF]$_A$ type can be illustrated by the compound *smertonosnyj* 'deadly, lethal' (49), in which the stem of the noun *smert'* 'death' is the object of the verb *nosit'* 'carry'. The suffix *-n-* (plus the inflectional ending *-yi*) gives rise to an adjective meaning 'carrying death'.

(49) *smert-o-nos-n-yj* ‘lethal’
death-*o*-carry-*n-yj*

[A-LV-V-SUFF]$_A$

The compound *bystrochodnyj* ‘high-speed’ exemplifies the pattern [A-LV-V-SUFF]$_A$, in which the first constituent is an adjectival/adverbial stem. In some cases, the adjectivalizing suffix can be absent, as in *mimoezžij* ‘passing by’ (51).

(50) *bystr-o-chod-n-yj* ‘high-speed’
fast-*o*-go-*n-yj*

(51) *mim-o-ezž-ø-ij* ‘passing by’
near-*o*-go-*ø-ij*

[PRO-LV-V-SUFF]$_A$

The pattern [PRO-LV-V-SUFF]$_A$ is instantiated by the compound *samodel’nyj* ‘self-made’ (52), in which we see the pronominal stem *sam-* ‘self’.

(52) *sam-o-del’-n-yj* ‘self-made’
self-*o*-make-*n-yj*

[NUM-LV-V-SUFF]$_A$

Finally, verb-based compound adjectives can show a numeral stem in first position, as in *odnoznačnyj* ‘unequivocal’ (53).

(53) *odn-o-znač-n-yj* ‘unequivocal’
one-*o*-mean-*n-yj*

1.6.3. Compound adverbs

Compound adverbs do not seem particularly productive in Russian (cf. Zemskaja 2005: 49, 2011: 319–320; Benigni & Masini 2009: 188), or at least their productivity is restricted to a small number of patterns, such as the reduplicative pattern of the type *edva-edva* 'hardly' (54), the pattern formed with the prefixoid *polu-* 'half' and a gerundive form, as in *polusidja* 'semisitting' (55), and the pattern in which the main verbal constituent is followed by the suffix *-om*, as in *mimochodom* 'in passing' (56) and *samoplavom* 'at drift' (57).

(54) *edva-edva* 'hardly'
hardly-hardly

(55) *pol-u-sidja* 'semisitting'
half-*u*-sitting

(56) *mim-o-chod-om* 'in passing'
near-*o*-go-*om*

(57) *sam-o-plav-om* 'at drift'
self-*o*-navigate-*om*

1.6.4. Compound verbs

Compound verbs are also not typical of Russian word-formation (cf. Zemskaja 2005: 49, 2011: 313; Benigni & Masini 2009: 188; Šanskij 2010: 274). Šanskij (2010: 274) argues that the few compound verbs existing in Russian are the result of a calquing process from Ancient Greek, such as *blagodarit'* 'thank' (cf. Gr. *eukharistéō*) (58) and *zloslovit'* 'speak evil' (cf. Gr. *kakologéō*) (59).

(58) *blag-o-darit'* 'thank'
good-*o*-give

(59) *zl-o-slovit'* 'speak evil'
bad-*o*-speak

Janko-Trinickaja (2001: 367–368) notes that, although compounding is not a common word-formation process for the creation of verbs, in contemporary Russian compound verbs are not so infrequent.

In general, scholars agree that the formation of compound verbs is almost only restricted to patterns formed with affixoids such as *polu-* 'half' and *sam-o-* 'self', such as *poluzakryt'* 'half-close' (60), and *samoustranit'sja* 'withdraw' (61) (cf. AG-80 § 961; Janko-Trinickaja 2001: 368–369; Zemskaja 2005: 49, 2011: 313; Benigni & Masini 2009: 188; Pazel'skaja 2013: 582).

[*polu*-V]$_V$

(60) *pol-u-zakryt'* 'half-close'
half-close

[*samo*-V]$_V$

(61) *sam-o-ustranit'sja* 'withdraw'
self-distance

In most cases, when the first constituent of the compound verb is not an affixoid as *polu-* or *sam-o-*, but a noun, the word-formation process at play is not compounding, but derivation from a compound noun (cf. Janko-Trinickaja 2001: 369–370; Pazel'skaja 2013: 583), as in *pylesosit'* 'vacuum' (from *pylesos* 'vacuum cleaner') (62) and *senokosit'* 'make the hay' (from *senokos* 'haymaking') (63).

(62) *pylesos-it'* 'vacuum'
vacuum_cleaner-*it'*[15]

(63) *senokos-it'* 'make the hay'
haymaking-*it'*

[15] The suffix *-t'* forms the infinitive of the verb and is preceded by the thematic vowel *-i-*.

The other possible situation according to Pazel'skaja (2013: 387) is that the compound verb is the result of a calquing process, as already pointed out by Šanskij (2010: 274). Pazel'skaja (2013: 385, 387) adduces the example of the verb *vidoizmenjat'* 'modify', which would derive from the noun *vidoizmenenie* 'modification' (*vid* 'aspect', *izmenenie* 'change'), which in turn constitutes a calque from Lat. *modificatio*.

1.7. Summary

The present chapter has discussed the main questions regarding the notion of "compound", i.e. the status of the constituent elements of compound words and the problem of delimiting compounds from derivatives and phrases (Section 1.1). In Section 1.2, I have accounted for the compounds' classification proposals that have been put forward in the literature. In Sections 1.3 and 1.4, I have focused on the notions of "synthetic" and "parasynthetic" compounding, which will be further discussed in Chapter 2 (cf. Section 2.3). In Section 1.5, I have reviewed the existing literature on the phenomenon of compounding in Russian, as it is described both in broader studies on Russian word-formation, and in surveys of specific compound types. Finally, in Section 1.6, I have provided an account of the different compound types in contemporary Russian, and I have focused particularly on compound nouns and adjectives, which are the most productive compound types in Russian.

2. A constructionist approach to compounding

The present chapter describes the theoretical framework that provides the basis for the analyses carried out in this study, i.e. *Construction Morphology* (henceforth CxM). Section 2.1 deals with the origin and development of *Construction Grammar* (henceforth CxG), within which CxM emerged. Section 2.2 discusses the fundamental notions and applications of CxM and focuses on the concept of hierarchical lexicon, in particular as applied to compounding and word-formation processes lying at the boundary between compounding and derivation. Section 2.3 shows how synthetic and parasynthetic compounds can be accounted for within a CxM framework, and introduces the notions of "schema unification" and "embedded productivity". In Section 2.4, I provide a constructionist account of Russian synthetic compounds with agentive meanings, which will be the object of the analyses carried out in Chapters 4 and 5.

2.1. The origin and development of Construction Grammar

In the present section, I briefly summarize the origin and development of the major constructionist theories, mainly based on the accounts provided by Hoffman & Trousdale (2013) and Masini (2016).

As recalled by Masini (2016: 15), the origin of CxG as a theoretical framework in linguistics reaches back to the paper by Fillmore, Kay and O'Connor "Regularity and Idiomaticity in Grammatical Constructions: The Case of *Let Alone*" (1988), in which the authors overtly take a stance against the traditional generative approach and explicitly use the notion of "construction": "The overarching claim is that the proper units of a grammar are more similar to the notion of construction in traditional and pedagogical grammars than to that of rule in most versions of generative grammar" (Fillmore, Kay & O'Connor 1988: 501). While traditional generative approaches only focused "on the core system, putting aside phenomena that result from historical accident, dialect mixture, personal idiosyncrasies, and the like" (Chomsky 1995: 20), CxG was concerned from the very beginning with phenomena usually considered as the "periphery" of the language. By

analyzing the pragmatic, semantic and syntactic features of the English construction *let alone*, Fillmore, Kay and O'Connor (1988) demonstrate that the realm of idiomaticity is far from being a collection of irregular expressions to be considered as exceptions to the rules of grammar. Rather, idioms are productive and highly structured (they are *constructions*), and deserve serious grammatical investigations.

Works by Charles Fillmore and Paul Kay (Fillmore, Kay & O'Connor 1988; Kay 1997; Kay & Fillmore 1999) build up what is known as *Berkeley Construction Grammar* (henceforth BCG, cf. Fillmore 2013: 112), a framework within which the grammar of a language is seen as "the set of its grammatical constructions, the rules that unite formal and semantic information into various kinds of linguistic objects, together with the principles that constrain and connect them" (Fillmore 2013: 112). The "construction", then, is the basic unit of linguistic analysis, which has to be considered as a conventionalized association of a certain form and a certain meaning. In Kay & Fillmore (1999), the authors examine another English construction, the *What's X Doing Y?* construction (e.g. *What's that fly doing in my soup?*), and show that constructions form hierarchical structures, in which specific constructions inherit their structure from more abstract constructions. The idea of a hierarchical structure allows accounting for general regularities, but also highlighting the particular features of more specific constructions.

Goldberg (1995) applies the constructionist approach to the study of the argument structure of verbs (so-called *Argument Structure Constructions*, such as the ditransitive construction, the caused-motion construction, and others), thus extending the notion of "construction" to more abstract mechanisms of language (cf. Masini 2016: 24). Her approach is usually called *Cognitive Construction Grammar* (CCxG), and is influenced by Fillmore's *Frame Semantics* (Fillmore 1982; Fillmore & Baker 2010), which entails the attribution of frames to lexical units, and by Lakoff (1987). Within this approach, constructions (i.e. form and meaning pairings) are seen as the building blocks of language. As the author claims, "[a]ny linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency" (Goldberg 2006: 5). The works by Goldberg have also inspired analyses of various grammatical constructions in other languages (cf., for instance, Michaelis &

Ruppenhofer 2001 and Hilpert 2008 on German; Barðdal 2008 on Icelandic), contrastive studies (such as Boas (2010), in which English constructions are compared to their counterparts in other languages), research in first language acquisition (e.g. Tomasello 2003), second language acquisition (e.g. Gries & Wulff 2005), and language change (e.g. Traugott 2008).

Croft (2001) develops the so-called *Radical Construction Grammar* model (henceforth RCxG), in which constructions are seen as the "primitive units of syntactic representation" (Croft 2001: 46), in place of grammatical and lexical categories. RCxG, indeed, "rejects grammatical categories independent of the constructions that define them" (Croft 2013: 218). This approach is particularly useful for typological analyses, as the right way to compare typologically different languages is to start from functions rather than from forms, and it is exactly the typological diversity of languages that leads to the hypothesis that all grammatical categories are language-specific and construction-specific. RCxG offers an innovative approach to grammatical categories and universals, and focuses both on language-internal and cross-linguistic variation.

Booij (2002a, 2002b, 2005b, 2007, 2009, 2010, 2013, 2015) elaborates a constructionist morphological theory (*Construction Morphology*, CxM), in which words are seen as constructions (i.e. conventionalized associations of form and meaning) at the morphological level. The CxM approach, which so far has been mainly applied to the study of word-formation and the lexicon, implies the idea of a hierarchical lexicon, in which constructions are organized in hierarchies in which more specific constructions are dominated by more abstract constructions. This way of representing the lexicon allows accounting for regularities and subregularities, but also to highlight the properties of more specific constructions. This theoretical framework, which is central to the investigations carried out throughout this study, will be discussed more in detail in Section 2.2.

As Masini (2016: 31) mentions, the most recent constructionist approach is the so-called *Sign-based Construction Grammar* (SBCG), developed by Boas and Sag (Boas & Sag 2012) based on BCG and *Head-Driven Phrase-Structure Grammar* (HPSG; cf. Pollard & Sag 1994). This theory can be seen as more formalist with respect to the other constructionist approaches, and aims at developing a formalizing method which could be employed by constructionists to develop and assess their hypotheses (Masini 2016: 31–

32). Within this theory, constructions are intended as "the means by which simpler signs are combined into more complex signs" (Michaelis 2013: 133).

Other theoretical frameworks, which are not usually defined "constructionist", are nonetheless quite similar and compatible with CxG. As Bybee (2013: 51) points out, "[c]onstructions, with their direct pairing of form to meaning without intermediate structures, are particularly appropriate for usage-based models" and, from the perspective of her usage-based theory, "constructions can be viewed as processing units or chunks – sequences of words (or morphemes) that have been used often enough to be accessed together" (Bybee 2013: 51). Bybee (2006, 2013) stresses the importance of type and token frequencies for the consolidation of constructions, i.e. different frequencies lead to different results (e.g. conventionalization, lexicalization, grammaticalization). Thus, within this perspective, language change is deeply influenced by usage.

Simone (2006, 2007) develops the so-called *Grammatica di Costruzioni e Categorie* 'Construction and Category Grammar' (GCC, cf. Masini 2016: 35). Differently from the RCG approach, within this model, the role of grammatical and lexical categories is still recognized as fundamentally important.

Cognitive approaches (Croft & Cruse 2004; Langacker 1987, 1991, 2008) are often very close to constructionist approaches. As Croft & Cruse (2004: 225) claim, "[t]he cognitive linguistic approach to syntax goes under the name of construction grammar". Broccias (2013) discusses Langacker's Cognitive Grammar (CG) approach and argues that it shares many assumptions with constructionist approaches. Within this approach, "language is viewed as a manifestation of the human mind (…) and is grounded in language-independent cognitive processes" (Broccias 2013: 192). Grammar is assumed to be inherently meaningful, and to consist of "assemblies of symbolic structures" (Broccias 2013: 192), i.e. form-meaning pairings that in constructionist approaches are called "constructions".

The *Parallel Architecture* model by Jackendoff (1997, 2002, 2013), though being a formal model of grammar, has come increasingly closer to constructionist approaches. The *Parallel Architecture* theory presupposes that phonology, syntax and semantics are autonomous combinatorial components that are linked by interface components. Within this picture, constructions are seen as particular interface rules. The compatibility and similarity of *Parallel Architecture* and CxG is highlighted by Jackendoff himself, who

admits that "[b]oth approaches emphasize the contribution of meaning to syntactic form; both approaches argue for the continuity between words, idioms, morphological affixes, and meaningful syntactic constructions; both approaches conceive of the lexicon as organized in terms of inheritance hierarchies" (Jackendoff 2013: 90).

The constructionist approach has also been applied to diachronic studies, giving rise to what is known as *Diachronic Construction Grammar* (cf. Barðdal et al. 2015), focusing particularly on the origin and development of constructions, i.e. on questions of language change. Traugott (2003) emphasizes the importance of constructions to explain phenomena of language change. Traugott & Trousdale (2010, 2014) propose a construction-based theory of language change, integrating CxG and studies on grammaticalization (cf. Masini 2016: 99).

As Masini (2016: 103–105) points out, the constructionist framework has also been adopted in studies on linguistic variation, both in a contrastive perspective (cf. Boas 2010), and in typological studies (cf. Ohori 2005; Croft et al. 2010). As regards sociolinguistic variation, the employment of constructionist approaches is still at an early stage, as noted by Hilpert & Östman (2014).

Constructionist approaches have also been adopted in psycholinguistic (cf. Bencini 2013) and neurolinguistic studies (cf. Pulvermüller, Cappelle & Shtyrov 2013), and in studies about language acquisition (cf. Tomasello 2003 and Diessel 2013 on first language acquisition; Gries & Wulff 2005 and Ellis 2013 on second language acquisition). Finally, other constructionist approaches, such as *Fluid Construction Grammar* (FCxG, cf. Steels 2013) and *Embodied Construction Grammar* (ECxG, cf. Bergen & Chang 2013) focus on the computational implementation of CxG.

2.2. Construction Morphology and the hierarchical lexicon

The present section focuses on the main tenets of CxM, the theoretical framework I adopt for the analyses carried out in this study. As already mentioned in Section 2.1, this morphological theory was developed by Geert Booij in a number of studies (2002a, 2002b, 2005b, 2007, 2009, 2010, 2013, 2015) aiming at "a better understanding of the relation between syntax, morphology, and the lexicon, and at providing a framework in

which both the differences and the commonalities of word level constructs and phrase level constructs can be accounted for" (Booij 2010: 1).

Within this framework, the basic unit of morphological analysis is the word. Thus, the approach adopted is not morpheme-based (e.g. Halle 1973; Siegel 1974; Kiparsky 1982), but word-based (e.g. Aronoff 1976; Booij 1977; Scalise 1984). The word is the minimal construction, "a pairing of form and meaning" (Booij 2010: 5), in which three types of information are correlated, i.e. the phonological form of the word (PHON), its morphosyntactic properties (SYN), and its semantic features (SEM), as represented in Figure 1.

Figure 1. The lexical representation of *baker* (Booij 2010: 7)

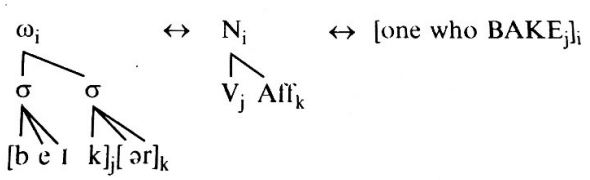


Figure 1 shows the lexical representation of the word *baker*. On the left, we see the formal properties, i.e. the phonological form of the word (ω_i), which is made of two syllables (each indicated by σ). The morphosyntactic information shows that the noun (N_i) is the result of deverbal derivation (V_j-Aff_k). Finally, on the right, we see the semantic information telling us that the meaning of the word *baker* is 'one who bakes'. The three blocks of information are associated through the symbol ↔ and the relation between the different blocks is specified through co-indexation. A more general schema that would be suitable for all deverbal agent nouns formed with the suffix *-er* in English is represented in Figure 2.

Figure 2. The schema for deverbal *-er* (Booij 2010: 8)

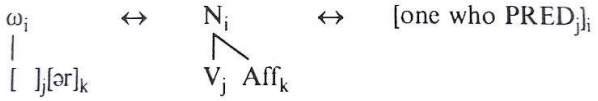


Figure 2 shows the generalized schema for deverbal *-er* in English, in which information related to a specific word is omitted, i.e. the phonological form of what comes before the suffix and the semantics of the predicate, which is replaced by the more general label PRED. Thus, we can see that constructions are organized into inheritance hierarchies, so that the more specific schema represented in Figure 1 is dominated by the more general schema in Figure 2. All the individual words ascribable to a schema are called "instantiations" (Booij 2010: 4) or "constructs" (Booij 2010: 12). An important aspect of the hierarchical system proposed in CxM is the concept of "default inheritance" (cf. Briscoe et al. 1993; Evan & Gazdar 1996; Killbury et al. 2006; Lascarides & Copestake 1999), i.e. the property of a word to inherit the features of its dominating node, unless it shows other specifications for that feature (Booij 2010: 27).

In what follows, I will discuss the principles of CxM, and particularly the concept of hierarchical lexicon, as related to compounding and phenomena at the boundary between compounding and derivation.

2.2.1. Compounds and Construction Morphology

In a CxM approach, compounds can be intended as constructions at the morphological level. Among the most investigated topics in CxM studies (e.g., among others, particle verbs, phrasal nouns, affixoids), compounding certainly occupies an important place. Booij (2009), for instance, develops "a constructional theory of compounding that makes use of some basic ideas of Construction Grammar, in particular constructional schemas, and the idea of a hierarchical lexicon (with multiple linking between words, and intermediate nodes between the most abstract schemas and the individual lexical items in order to express intermediate levels of generalization)" (Booij 2009: 202). Booij (2010: 17) shows the general schema for nominal compounds in Germanic languages (1):

(1) $[[a]_{Xk}\ [b]_{Ni}]_{Nj} \leftrightarrow [SEM_i \text{ with relation R to } SEM_k]_j$

The schema in (1) shows that, in Germanic languages, nominal compounds are formed by a nominal head on the right and a non-head element on the left which can belong to

one of the major lexical categories (i.e. noun, verb, adjective, pronoun), which is indicated by X. The schema specifies that the constituent in second position is not only the formal head, but also the semantic head, i.e. the compound denotes a certain type of $[b]_{Ni}$. The variables *a* and *b* stand for arbitrary sound sequences. The nature of the relation R is determined based on the meanings of the compound's constituents (Booij 2010: 17).

This abstract schema dominates a series of more specific subschemas, all ascribable to the general schema in (1). So, for instance, the schema in (2) is a subschema of (1) which comprises all English nominal compounds in which the head noun is modified by another noun $[a]_{Nk}$. The schema in (3) is a further specification of (2) and represents all English compounds in which the head element is the noun *shop*, modified by a noun. This partially specified construction could be instantiated by compounds such as *bookshop* and *woodshop*. Finally, in (4), we see the representation of a fully specified construction, which is dominated by all the preceding schemas.

(2) $[[a]_{Nk}\ [b]_{Ni}]_{Nj} \leftrightarrow [\text{SEM}_i \text{ with relation R to } \text{SEM}_k]_j$

(3) $[[a]_{Nk}\ \textit{shop}]_{Nj} \leftrightarrow [\textit{shop} \text{ with relation R to } \text{SEM}_k]_j$

(4) $[\textit{bookshop}]_{Nj} \leftrightarrow [\textit{shop} \text{ selling } \textit{books}]_j$

This representation of the lexicon as a hierarchy enables to account for high-level rules, but also for subregularities, and for the peculiarities of specific constructions.

2.2.2. Between compounding and derivation

The need for a hierarchical representation of the lexicon and for intermediate levels of generalization is particularly evident when one considers phenomena lying at the boundary between compounding and derivation, such as compounds formed with so-called affixoids (cf. Section 1.1).

First, the postulation of "partially lexically specified productive patterns" (Booij 2009: 209) within the hierarchical lexicon described above allows accounting for such intermediate cases without necessarily drawing a sharp boundary between compounding and derivation. The similarity between these two word-formation processes is captured

by intermediate schemas, which are dominated by the more abstract schema for compounding (since we have two lexical stems), but are similar to derivation schemas because one element is lexically specified.

(5) $[[a]_X\ [b]_Y]_Y$

(6) $[[a]_X\ b]_Y$

(7) $[a\ [b]_Y]_Y$

The schema in (5) is a general representation of right-headed compounding. The two phonological strings *a* and *b* are assigned a lexical category (X and Y). The fact that the lexical category of the compound coincides with the lexical category of the second constituent shows that this word-formation process gives rise to right-headed compounds. The schema in (6) represents suffixation. Here we see that the second element is not assigned a lexical category because it stands for the suffix. By contrast, in (7), we see the opposite situation, i.e. that the first element is not assigned a lexical category, as it represents the prefix. Here we can also observe that the category of the derived word coincides with the category of the second lexical element: prefixation, indeed, does not change the lexical class of the words it produces.

Let us now consider the case of affixoids and, as an example, the case of the prefixoid *sam-o-* 'self' in Russian, which I have discussed in a previous study (Naccarato 2016). The constituent *sam-o-* is formed by the reflexive pronoun *sam* 'self' followed by the linking vowel *-o-*, and is used as first constituent in Russian compounds in which the second head constituent can be a noun, an adjective, an adverb, or a verb. In compounds, we see that this element can have a pronominal function with the meaning of 'self' (e.g. *samozaščita* 'self-defense', *samouverennyj* 'self-confident, etc.) and an adverbial function with the meaning of 'autonomous/autonomously, by oneself' (e.g. *samolečenie* 'autonomous treatment', *samoventilacija* 'automatic ventilation', etc.). The two meanings can be represented as in (8) and (9).

(8) $[sam\text{-}o\text{-}[b]_Y]_Y \leftrightarrow$ [self *b*]

(9) $[sam\text{-}o\text{-}[b]_Y]_Y \leftrightarrow$ [autonomous/automatic *b*]

The two schemas in (8) and (9) can be ascribed to the more general schema of pure compounding (cf. Section 1.6) in Russian, which can be represented as in (5), and can be considered as partially specified subschemas. At the same time, the behavior of the constituent *sam-o-* in compounding, which shows a restricted number of functions and high regularity, makes it similar to prefixes (7), the only difference lying in its lexical (unbound) nature, which does not characterize prefixes.

The second advantage of a constructionist approach consists in the fact that the subregularities of such (often polysemous) items can be represented through constructional schemas. Thus, the two schemas in (8) and (9) allow accounting for the polysemy of the affixoid *sam-o-* in compounding, showing the two available subregular patterns.

2.3. Synthetic and parasynthetic compounds: schema unification and embedded productivity

As already discussed in Section 1.3, synthetic compounds constitute a particular class of compounds that deserves special investigations. In the present section, I will show how such compounds can be accounted for within a CxM framework by resorting to the notions of "schema unification" and "embedded productivity".

Booij (2005a: 90) defines synthetic compounding as that type of word-formation in which compounding and derivation take place simultaneously, as in *heart-breaker*, where the compounding process (i.e. *heart* + *break*) goes together with a process of deverbal derivation (i.e. *break* + *-er*) expressed by the addition of the agentive suffix *-er*. In such compounds, the left non-head constituent plays a specific semantic role with respect to the verbal base (Booij 2005a: 91) (cf. Chapter 1, Section 1.3). Booij (2005a: 91) points out that the conflation of compounding and derivation is also found in other compound types, such as *blue-eyed*, which are not usually included in the class of synthetic compounds[16] because they do not show a verbal base and, consequently, a relation of argumental nature between the verbal base and the non-head constituent. Such compounds are, instead, included by some scholars under the label of "parasynthetic

[16] Booij (2005b: 128), however, does include in the "synthetic" class also cases such as *blue-eyed*.

compounds" (Bisetto & Melloni 2008; Melloni & Bisetto 2010) (cf. Chapter 1, Section 1.4), which comprises both verb-based and noun-based compounds, whose main property is that they are the result of the simultaneous occurrence of compounding and derivation.

Studies on synthetic compounds have often been concerned with defining the order in which compounding and derivation occur in their formation, thus excluding the idea that they might be simultaneous processes. The two possible situations are represented in (10) and (11). In the first case, derivation is supposed to take place before compounding, i.e. a deverbal noun is attached to another noun and we get a process of $[N\text{-}N]_N$ compounding, e.g. [*heart*-[*break-er*]]. By contrast, in the second case, suffixation is supposed to take scope over the compound base, i.e. we first have a process of $[N\text{-}V]_V$ compounding giving rise to a compound verb, and subsequently we get a noun through suffixation of the complex base, e.g. [[*heart-break*]-*er*].

(10) $[N\text{-}[V\text{-}SUFF]_N]_N$

(11) $[[N\text{-}V]_V\text{-}SUFF]_N$

The first interpretation (10) is supported by Selkirk (1982), Di Sciullo & Williams (1987), and Booij (1988). Referring to synthetic compounds in Dutch, Booij (1988) claims that the deverbal noun is created first, and then it combines with another noun to form a compound. As the author argues, "[t]he argument interpretation of the left constituents of such compounds is made possible by the fact that the deverbal noun inherits the argument structure of its verbal base" (Booij 1988: 57).

The opposite interpretation (11) is supported by Lieber (1983), who claims that nouns do not have argument structure and synthetic compounds are the result of derivation of a complex verbal base. If this interpretation can be attractive because it accounts for the fact that the noun in first position is an argument of the verb (without resorting to the concept of inheritance of the argument structure), it also has some drawbacks, because the $[N\text{-}V]_V$ compounding process is not productive in English, as recognized by Lieber herself in a later study (Lieber 2004: 48).

Molinsky (1973: 114) distinguishes the two situations in (10) and (11) by resorting to the notions of "autonomous" and "bound nominalizations". The two patterns of compounding in Russian are represented as in (12), where the second verbal stem plus

the nominalizing suffix form an autonomous deverbal noun, and (13), where we see the nominalization of the two stems as a composite unit (Molinsky 1973: 114).

(12) [$Stem_1$-o-($Stem_2$-nom)]

(13) [($Stem_1$-o-$Stem_2$)-nom]

Autonomous nominalizations are instantiated by the author through compounds such as *zolotoiskatel'* 'gold seeker' (14), while bound nominalizations can be exemplified by compounds such as *pis'monosec* 'postman' (15).

(14) *zolot-o-iska-tel'* 'gold seeker'
gold-*o*-seek-*tel'*

(15) *pis'm-o-nos-ec* 'postman'
letter-*o*-bring-*ec*

While in (14) the verbal base plus the suffix form the autonomous deverbal *iskatel'* 'seeker', in (15) the deverbal **nosec* 'one who brings' does not exist outside compounding, which is why the author postulates that nominalization takes scope over a composite unit, which, however, is also inexistent outside compounding: **pis'monos(it')* 'bring letters'.

In Russian (and in Slavic languages in general, cf. Szymanek 2009), patterns of compounding combined with suffixation are very common and not only restricted to cases of synthetic compounding *stricto sensu*, but they also include cases of noun-based compounding, as will be discussed below. The Russian data clearly demonstrate that determining whether processes of conflation of compounding and derivation patterns as the result of the schema in (10) or of the schema in (11) is not always possible or desirable. If we consider, for instance, the compound *zubočistka* 'toothpick' (16), we see that the verbal base plus the suffix form the autonomous deverbal *čistka* 'cleaning'.

(16) *zub-o-čist-ka* 'toothpick'
tooth-*o*-clean-*ka*

However, the compound *zubočistka* 'toothpick' is unlikely to be the result of a process of derivation followed by compounding, as represented in (10). Indeed, the deverbal noun *čistka* has a processual meaning, i.e. 'cleaning process', whereas the compound does not denote a type of cleaning process, but an instrument. At the same time, we cannot say that the suffix is attached to a compound base, because the two stems do not form an autonomous compound verb as **zubočist(it')* 'clean teeth'. Such cases are better explained if we adopt a constructionist approach, which employs the notions of "schema unification" and "embedded productivity".

Booij (2005b, 2009, 2010) describes synthetic compounds as the result of the "conflation" (Booij 2005b: 128) of N-V compounding and suffixation, i.e. compounding and derivation are simultaneous processes, and "the systematic co-occurrence of the two word formation processes is expressed by template unification" (Booij 2009: 214). The conceptualization of such compounds as deriving from schema unification allows overcoming the question of the unproductivity of one of the two word-formation patterns, because "schema unification may result in a productive schema even though one of its building blocks is unproductive" (Booij 2010: 47). This phenomenon, i.e. "that a word formation process is normally unproductive, but is productive when it co-occurs with another word formation process", is defined "embedded productivity" (Booij 2010: 47). Thus, although the $[N\text{-}V]_V$ schema is unproductive in Russian (as it is in English and in Dutch), we can assume that it becomes productive once it is embedded in the unified schema in which it is conflated with suffixation (cf. also Gaeta 2006, 2010).

As mentioned above, Russian (and Slavic languages in general) is rich in compounds arising from the combination of compounding and derivation, and this is not only restricted to verb-based compounds. Noun-based nominal (17) and adjectival (18) compounds combining compounding and derivation are also quite frequent (cf. also Chapter 1, Section 1.6).

(17) *perv-o-klass-nik* 'first grader'
first-*o*-class-*nik*

(18) *dolg-o-sroč-n-yj* 'long-term'
long-*o*-term-*n-yj*

The compounds in (17) and (18) can also be described in terms of schema unification, as they are the result of the conflation of a $[A\text{-}N]_N$ compounding schema and a $[N\text{-}SUFF]_A$ derivation schema.

2.3.1. Presumed weaknesses of a Construction Morphology approach to (para)synthetic compounding

As discussed in Chapter 1 (cf. Section 1.4), Bisetto & Melloni (2008) have pointed out that some Slavic compounds pose a challenge for the constructionist approach. The authors refer to cases such as *goluboglazyj* 'blue-eyed' (19) and *glubokomyslie* 'profundity of thought' (20), in which the derivation schemas are considered to represent not only non-attested, but also impossible words (Bisetto & Melloni 2008: 241).

(19) *golub-o-glaz-ø-yj* 'blue-eyed'
blue-*o*-eye-*ø*-*yj*

(20) *glubok-o-mysl-ie* 'profundity of thought'
profound-*o*-thought-*ie*

As for the first case (19), the authors note that sequences such as **glazyj* 'eyed' do not give rise to well-formed words in Russian outside compounding, because adjectives are never formed through the addition of the adjectival inflection directly to a nominal base, without an overt suffix. To explain such cases, Bisetto & Melloni (2008: 248) postulate the existence of an unexpressed suffix deriving an adjective from a complex base, as shown in Chapter 1 (cf. Section 1.4). The zero suffix would emerge in cases of inalienable possession, i.e. when the nominal stem in second position denotes an entity that is inalienably possessed by the referent, usually human or animal body parts (cf. also Fedorova 2015). This would be in line with the more general, cross-linguistic tendency to avoid "morpho-phonological bulkiness with body-part expressions" (Bisetto & Melloni 2008: 249; cf. also Haiman 1983; Chappell & McGregor 1996).

As I have discussed in Chapter 1 (cf. Section 1.4), Chovanová & Štichauer (2014) propose a different view of such cases, and argue that the principle lying at the basis of

such constructions is not that of inalienable possession, since this principle "is sometimes too strong in the sense that it predicts (or overgenerates) cases which are not acceptable" (Chovanová & Štichauer 2014: 148), as in Cz. **tvrd-o-játr-ý alkoholik* 'hard-livered alcoholic', which is not a well-formed compound because it does not denote salient and visible body parts. By contrast, in other cases it seems too weak, as it also includes cases in which both the nominal component of the compound and the head noun modified by the adjectival compound refer to an artifact (e.g. Cz. *tenk-o-zub-ý hřeben* 'thin-toothed comb'), whereas inalienable possession is usually intended as affecting animate referents. Thus, the authors propose that the principle underlying such cases is rather that of "Integrated Meronymy", i.e. a salient part of a certain referent is employed to access the whole referent.

I follow Chovanová & Štichauer's approach to such compounds, and argue that they instantiate cases of meronymy rather than inalienable possession. In my view, however, the arguments adduced by the authors are not convincing. First, the body parts in such compounds are not necessarily "visible": in Russian, we do find compounds such as *dobr-o-serd-yj* 'kind-hearted'. Second, it is not the presence of compounds referring to artifacts rather than to animate entities that excludes the principle of inalienable possession, as inalienable possession is usually extended to include inherent parts of inanimate items, e.g. *branch* or *handle* (Heine 1997: 85). The principle of inalienable possession must be restricted to meronymy because such compounds do not cover all cases that are usually ascribed to inalienable possession. For instance, we never find kinship terms in such compounds, whereas kinship terms are usually treated as belonging to the domain of inalienable possession (Heine 1997: 85).

In any case, regardless of whether we want to consider these constructions as cases of inalienable possession or integrated meronymy, I do not see them as problematic for a constructionist approach. In my view, such examples can also be accounted for in terms of schema unification and embedded productivity. Not only unproductive, but also non-existing schemas can enter into use once they are embedded in a larger schema: "a language may acquire new morphological schemas as an effect of schema unification and subsequent simplification" (Booij 2010: 47).

The second case (20) is considered problematic by Bisetto & Melloni (2008: 242) because it is also seen as lacking an overt suffix, and the gender shift occurring in the

word-formation process (from feminine/masculine to neuter) cannot be brought about by the inflectional marker alone, which is why the authors postulate "the presence of a nominalizing suffix associated to the neuter gender and taking scope over the pre-modified N (or V) stem" (Bisetto & Melloni 2008: 242). However, as already discussed in Chapter 1 (cf. Section 1.4), I do not agree with this analysis, as these formations do include a suffix, i.e. the suffix *-i(j)-e*, which is used in parasynthetic formations, not necessarily compound, as in *bessmertie* 'immortality', from *bez smerti* 'without death' (cf. Townsend 1980: 151). Therefore, this construction can also be accounted for in terms of schema unification, in which the $[A\text{-}N]_N$ compounding schema conflates with suffixation through the addition of the suffix *-i(j)-e*.

In both cases, we are dealing with schemas that become productive in the language only once they are embedded in larger schemas, and the constructionist approach appears as the most appropriate framework to provide a satisfactory explanation of such phenomena.

2.4. A constructionist account of Russian compound agent nouns

In the present section, I provide a constructionist account of Russian synthetic compounds with agentive meanings that will be the object of the investigations carried out in Chapters 4 and 5.

The compounds that will be further analyzed can be ascribed, in the first place, to the general schema representing suffixed compound nouns in Russian, which is shown in (21). The structure of such compounds consists of the combination of two stems separated by a linking vowel and followed by a suffix. The nouns created denote a concrete or abstract entity showing a relation R with the second stem (which is the main constituent of the compound). The second stem is modified or specified by the first stem. The type of noun created depends on the suffix employed.

(21) $[X_k\text{-LV-}Y_i\text{-SUFF}]_{Nj} \leftrightarrow$ [entity with relation R to SEM_i; SEM_i is specified by SEM_k]

Verb-based suffixed compounds are represented through the schema in (22), which is a more specific schema, dominated by the more abstract schema in (21). The greater specificity of this schema is given by the substitution of the generic stem Y with the specified stem V, indicating that the second stem of the compound is verbal.

(22) $[X_k\text{-LV-}V_i\text{-SUFF}]_{Nj} \leftrightarrow$ [entity with relation R to SEM_i; SEM_i is specified by SEM_k]

The construction in (22) represents processes of synthetic compounding in Russian and is the result of schema unification. The $[X\text{-LV-}V]_V$ schema, which is not productive in Russian, becomes productive once it is embedded in a larger schema including the addition of a deverbal suffix.

The schema for synthetic compounds in (22) can be further specified, in the first place, by indicating the suffix involved in the compounding process. A first distinction to be made is between constructions forming nouns denoting the participant of an event, i.e. Agents, Instruments, Objects/Patients, such as the construction in (23), formed with the suffix *-ec*, and constructions giving rise to nouns denoting the Event itself, such as the construction in (24), formed with the suffix *-stvo*.

(23) $[X_k\text{-LV-}V_i\text{-}ec]_{Nj} \leftrightarrow$ [entity involved in SEM_i; SEM_i is specified by SEM_k]

(24) $[X_k\text{-LV-}V_i\text{-}stvo]_{Nj} \leftrightarrow$ [process of SEM_i; SEM_i is specified by SEM_k]

The two cases are exemplified by the compounds in (25) and (26).

(25) *basn-o-pis-ec* 'fable writer' (entity carrying out the action of writing fables)
fable-*o*-write-*ec*

(26) *slov-o-proizvod-stvo* 'word-formation' (process of producing words)
word-*o*-produce-*stvo*

If we focus on constructions forming nouns denoting the participant of an event, exemplified by the construction with the suffix *-ec* (23), we can make further

specifications and indicate the categorial status of the first stem of the compound, which can belong to one of the following lexical categories: nouns (27), adjectives/adverbs (28) (see Chapter 4, Section 4.3.1, for further discussion on the distinction between adjectival and adverbial stems in compounds), pronouns (29), and numerals (30).

(27) $[N_k\text{-LV-}V_i\text{-}ec]_{Nj} \leftrightarrow$ [entity involved in SEM_i; SEM_i is specified by SEM_k]

(28) $[A_k\text{-LV-}V_i\text{-}ec]_{Nj} \leftrightarrow$ [entity involved in SEM_i; SEM_i is specified by SEM_k]

(29) $[PRO_k\text{-LV-}V_i\text{-}ec]_{Nj} \leftrightarrow$ [entity involved in SEM_i; SEM_i is specified by SEM_k]

(30) $[NUM_k\text{-LV-}V_i\text{-}ec]_{Nj} \leftrightarrow$ [entity involved in SEM_i; SEM_i is specified by SEM_k]

In examples (31) to (34), I show one instantiation for each of the subschemas represented in (27) to (30).

(31) *basn-o-pis-ec* 'fable writer' (entity carrying out the action of writing fables)
fable-*o*-write-*ec*

(32) *vol'n-o-dum-ec* 'free thinker' (entity carrying out the action of thinking freely)
free-*o*-think-*ec*

(33) *vs-e-vid-ec* 'all-seeing' (entity carrying out the action of seeing everything)
all-*e*-see-*ec*

(34) *pjat-i-bor-ec* 'pentathlete' (entity carrying out the action of fighting in five sports)
five-*i*-fight-*ec*

At a lower level, by specifying one of the two stems embedded in a construction, we can identify other subschemas showing a certain regularity. Thus, a subschema of the construction in (27) is represented in (35), where the verbal base is specified.

(35) [N-LV-*ljub-ec*] 'person loving N'

The subschema in (35) accounts for a subregularity within the more abstract schema in (27), and specifies that, whenever the verbal stem in this construction is constituted by

the verb *ljubit'* 'love', the compound produced has the meaning of 'person loving N', as in the compounds shown in (36) to (38).

(36) *žizn-e-ljub-ec* 'life lover'
life-*e*-love-*ec*

(37) *pravd-o-ljub-ec* 'truth lover'
truth-*o*-love-*ec*

(38) *trud-o-ljub-ec* 'hard worker'
work-*o*-love-*ec*

Subregularities can also be found on the left side of the constructions considered. The subschema in (39) represents a subregularity within the construction in (29), where the pronominal stem is constituted by the affixoid *sam-o-* 'self' (cf. also Section 2.2.2).

(39) [*sam-o*-V-*ec*] 'self V-*ec*'

The subschema in (39) is instantiated by the compounds in (40) to (42).

(40) *sam-o-ljub-ec* 'self-lover'
self-*o*-love-*ec*

(41) *sam-o-pis-ec* 'self-recorder'
self-*o*-write-*ec*

(42) *sam-o-deržec* 'autocrat'
self-*o*-hold-*ec*

The compound in (42) instantiates a further subregularity within the [*sam-o*-V-*ec*] subschema, and is representative of those *sam-o*-compounds having the meaning of 'autonomous/autonomously, by oneself' (cf. Section 2.2.2).

To summarize what has been discussed so far, in (43) I show how compound agentive constructions with the suffix *-ec* can be accounted for in the form of a hierarchical lexicon.

(43) [X-LV-Y-SUFF]$_N$

[X-LV-V-SUFF]$_N$

[X-LV-V-*ec*]$_N$

[N-LV-V-*ec*]$_N$

[N-LV-*ljub*-*ec*]$_N$

žizneljubec ‘life lover’
pravdoljubec ‘truth lover’
trudoljubec ‘hard worker’

[…]

[…]

[PRO-LV-V-*ec*]$_N$

[*sam*-*o*-V-*ec*]$_N$

samoljubec ‘self-lover’
samopisec ‘self-recorder’
samoderžec ‘autocrat’

[…]

[…]

[A-LV-V-*ec*]$_N$

[…]

[NUM-LV-V-*ec*]$_N$

[…]

For reasons of space, the schemas and subschemas shown in (43) do not include the semantic information corresponding to each construction, which, however, can be recovered from the previous description. In addition, for obvious reasons, the hierarchy in (43) does not specify all the possible subschemas at each level. So, for instance, for the schema [N-LV-V-*ec*]$_N$, only one subschema is shown with some of its possible

instantiations, but more subschemas could be of course included, e.g. [N-LV-*nos*-*ec*]$_N$ 'entity carrying N' (*vodonosec* 'water carrier', *kop'enosec* 'spare-bearer, etc.).

The main advantage of a representation such as the one shown in (43) is that it allows accounting for both high-level regularities, to which different constructions can be ascribed, and for low-level regularities, which guarantee the opportunity to identify subconstructions showing certain peculiarities.

2.5. Summary

In the present chapter, I have first provided an overview of the origins of CxG and of its main subtheories (Section 2.1). Then, I have discussed the main tenets of CxM, focusing particularly on the notion of hierarchical lexicon and its importance for the description of word-formation processes lying at the boundary between compounding and derivation (Section 2.2). In Section 2.3, I have shown how synthetic and parasynthetic compounds can be accounted for within a CxM framework based on the notions of schema unification and embedded productivity, and I have argued that the cases that have been defined as problematic to fit in a constructionist model (cf. Bisetto & Melloni 2008) can also be easily accounted for in such a framework if we assume that schema unification allows unproductive or inexistent schemas to actively enter into the language system. Finally, in Section 2.4, I have proposed an account of Russian synthetic compounds with agentive meanings (through an exemplification based on the construction in *-ec*) in the form of a hierarchical lexicon. This account has demonstrated the advantages of a constructionist approach to compounding, which makes it possible to represent regularities, subregularities and peculiarities of the schemas and subschemas embedded in the hierarchy.

3. Agent nouns and their polysemy

The present chapter is devoted to a description of agent noun constructions in Russian. Particular attention is paid to the agentive suffixes involved in the compound constructions that will be discussed in Chapter 4. Section 3.1 focuses on the origin, etymology and use of agent noun suffixes in Russian. Section 3.2 deals with the polysemy of agent noun suffixes not only in Russian, but also from a cross-linguistic perspective. Sections 3.3 to 3.5 discuss how the phenomenon of polysemy in word-formation is accounted for in different theoretical frameworks, i.e. cognitive, constructionist and diachronic approaches. In Section 3.6, I investigate the polysemy of Russian agent noun suffixes, while Section 3.7 is devoted to the suffixless agentive construction.

3.1. Agent noun suffixes in Russian: origin, etymology and use

Russian employs a wide range of suffixes to form agent nouns. Among the most common are the following: *-tel'* (e.g. *chranit'* 'keep' → *chrani**tel'*** 'keeper, guardian'), *-nik* (e.g. *rabotat'* 'work' → *rabot**nik*** 'worker'), *-ščik*/*čik*[17] (e.g. *tancevat'* 'dance' → *tancov**ščik*** 'ballet dancer'; *letat'* 'fly' → *let**čik*** 'pilot'), *-l'ščik* (e.g. *nosit'* 'carry' → *nosi**l'ščik*** 'porter'), *-ec* (e.g. *kupit'* 'buy' → *kup**ec*** 'merchant'), *-lec* (e.g. *žit'* 'live' → *ži**lec*** 'tenant'), *-un* (e.g. *molčit'* 'be silent' → *molč**un*** 'taciturn person'), *-ok* (e.g. *streljat'* 'shoot' → *strel**ok*** 'marksman'), *-ar'* (e.g. *peč'* 'bake' → *pek**ar'*** 'baker'), *-ator* (e.g. *organizovat'* 'organize' → *organiz**ator*** 'organizer'), *-ant*/*ent* (e.g. *èmigrirovat'* 'emigrate' → *èmigr**ant*** 'emigrant'; *abonirovat'* 'subscribe' → *abon**ent*** 'subscriber'), *-ka* (e.g. *zaznat'sja* 'take on air' → *zaznaj**ka*** 'conceited person'), *-lka* (e.g. *gadat'* 'tell fortunes' → *gada**lka*** 'fortune-teller'), and others.

Not all these suffixes, however, take part in the formation of compound agent nouns, or at least, not all of them appear in the compounds that I found in the RNC word-formation database. This section mainly focuses on the origin and etymology of the

[17] The allomorph *-čik* is found after the following consonants: d, t, z, s, and ž (cf. AG-80, §213).

agentive suffixes that are involved in the formation of compounds in contemporary Russian, as represented in this database. In addition, some remarks on the semantics and use of these suffixes in modern Russian are also provided, consistently with the descriptions found in the AG-80.

3.1.1. The suffix -tel'

The suffix *-tel'* is equivalent to the OCS agentive suffix *-tel'ĭ* (e.g. OCS *datel'ĭ* 'giver', from *dati* 'give'), and has cognates in most Slavic languages (cf. Bel. *-tel*, Ukr. *-tel'*, Bul. *-tel*, Mac. *-tel*, Scr. *-telj*, Slo. *-tel*, Cze. *-tel*, Svk. *-tel'*, Srb. *-ć/śel*, Pol. *-ciel*), where it retains the function of deverbal agentivization (e.g. Rus. *pisatel'* 'writer', from *pisat'* 'write').

Most scholars trace the origin of this suffix as arising from the PIE agentive suffix **-ter-* (cf. Skt. *-tar-*, Lat. *-tōr-*, Gr. *-tōr-* and *-tēr-*; see Meillet 1905: 311–312; Vaillant 1974: 315–316; Efimova 2006: 60; Luschützky 2011: 79).

However, Vaillant (1974: 315–316) questions the hypothesis of a process of liquid dissimilation taking place after *r*, and suggests that the most plausible hypothesis is that the suffix originated from the Balto-Slavic suffix *-el-*, which developed in contact with Germanic *-il-*, deriving from **-el-*, and in Slavic it combined with the rests of PIE **-ter-*.

Vaillant (1974: 314–315) adds that the suffix *-tel'ĭ* was very productive in OCS, while in modern Slavic languages it seems to have retained its productivity almost only in Russian, while in the other languages its use is residual or limited to certain forms (e.g. calques from OCS) and registers (e.g. written register).

In modern Russian, the suffix *-tel'* denotes animate or inanimate entities that perform a certain action or that are intended to perform a certain action, and is productive in different linguistic areas, particularly in technical terminology in the case of nouns denoting Instruments (AG-80, §§ 211, 558). The suffix *-tel'* is only deverbal and is attached to infinitive stems.

3.1.2. Suffixes -nik, -ščik/čik and -l'ščik

The suffix *-nik* originates from OCS *-ikŭ* (PIE **-eiko-*, cf. Vaillant 1974: 541) based on the participial form in *-n-* (Vaillant 1974: 306; Luschützky 2011: 80): OCS *učiti* 'to teach' → *učenŭ* 'taught' → *učenikŭ* 'disciple, pupil'; the derived noun *učenikŭ* is then reanalyzed as *uč-enikŭ* and the suffix *-nik* appears as the result of metanalysis (cf. Luschützky 2011: 80). As Vaillant (1974: 539) points out, the OCS suffix *-ikŭ* was only productive in the form of *-nikŭ*.

This suffix is highly productive in all Slavic languages (cf. Bel. *-nik*, Ukr. *-nik*, Bul. *-nik*, Mac. *-nik*, Scr. *-nik*, Slo. *-nik*, Cze. *-ník*, Svk. *-ník*, Pol. *-nik*). In OCS, this suffix is mostly productive in deadjectival and denominal word-formation, while deverbal derivation is less frequent (Luschützky 2011: 77, 80). In modern Slavic languages, the suffix shows multifunctionality, unspecificity of meaning and the ability to derive nouns from nouns (e.g. Rus. *vestnik* 'messenger' from *vest'* 'piece of news'), as well as from adjectives (Cze. *hanebník* 'rotter' from *hanebný* 'disgraceful') and verbs (Pol. *pracownik* 'worker' from *pracować* 'work'), and all these factors seem to enhance its productivity (Luschützky 2011: 91).

The suffix *-ščik* (and its allomorph *-čik*) is the result of amalgamation of *-ĭsk(o)-* (a suffix deriving denominal adjectives) and *-ikŭ*, and the suffix *-l'ščik* is the result of further amalgamation (Luschützky 2011: 93–94).

In modern Russian, in deverbal derivation, the suffix *-nik* denotes animate or inanimate entities that perform a certain action or that are intended to perform a certain action, and is particularly productive in technical terminology, especially for nouns denoting weapons and mechanisms (AG-80, §§ 212, 562). The suffixes *-ščik/čik* and *-l'ščik* give rise to nouns denoting (mostly animate) entities that perform a certain action and are particularly productive in professional and technical terminology (AG-80, §§ 213, 214, 560, 561). The suffixes *-nik* and *-ščik/čik* are also employed in deadjectival and denominal derivation (cf. AG-80, §§ 285, 286, 329–331) and compounding (cf. AG-80, §§ 570, 571), while the suffix *-l'ščik* is only deverbal and is attached to infinitive stems.

3.1.3. Suffixes -ec and -lec

The suffix *-ec* is equivalent to the OCS suffix *-ĭcĭ*, which arose from PIE *-*(i)ko*- (cf. Gr. *-ikós*, Lat. *-icus*, Skt. *-ikáḥ*, Lit. *-ìkas*; see Meillet 1905: 340; Vaillant 1974: 300–301; Luschützky 2011: 80). In OCS, the suffix *-ĭcĭ* shows different meanings (cf. Vaillant 1974: 295–299):

a) agentive, e.g. *borĭcĭ* 'fighter' (from *brati(sja)* 'fight');
b) diminutive, e.g. *domĭcĭ* 'little house' (from *domъ* 'house');
c) nominalizer of adjectives, e.g. *bradatĭcĭ* 'bearded man' (from *bradatyj* 'bearded');
d) singulative, e.g. *cvětĭcĭ* 'one flower' (from *cvětъ* 'flower').

The semantic vagueness of this suffix led to different outcomes in modern Slavic languages: its multifunctionality is preserved in Bul. *-ec*, Mac. *-ec*, Cze. *-ec*, Svk. *-ec*, Srb. *-c*, Pol. *-iec*, Bel. *-ec*, Ukr. *-ec*, and Rus. *-ec*, in which the suffix can be attached not only to verbal bases, but also to nouns and adjectives, giving rise to nouns with a wide range of meanings. By contrast, Scr. *-ac* and Slo. *-ec* mainly give rise to nouns with purely agentive meanings (cf. Luschützky 2011: 87–90). The suffix *-lec* is the result of metanalysis (Luschützky 2011: 90) on the basis of participial forms in *-l-*.

In modern Russian, the suffix *-ec* is often described as unproductive in deverbal derivation and is mainly used to form nouns denoting Human Agents (AG-80, § 216). In compounding, it is said to show low degrees of productivity, limited to cases in which the verbal base plus the suffix form an autonomous deverbal outside compounding (AG-80, § 560), but, as I will show in Chapter 5 (cf. Section 5.5.1), this claim does not seem well-grounded. The suffix *-lec* also denotes Human Agents and shows some productivity in colloquial and publicistic speech (AG-80, § 217). The suffix *-ec* is also used in deadjectival and denominal derivation (cf. AG-80, §§ 287, 288, 332) and compounding (cf. AG-80, §575), while the suffix *-lec* is only deverbal and is attached to infinitive stems.

3.1.4. Suffixes -ka and -lka

The suffix *-ka* corresponds to OCS *-ŭka*, which is the feminine form of *-ŭkŭ*, and thus represents *-*u-kā* (cf. Skt. *-ukā-*, Lit. *-ukė*; see Vaillant 1974: 355). The suffix was not

productive in OCS, but it gained a great productivity in modern Slavic languages, where it is employed with a wide range of meanings (cf. Vaillant 1974: 352–355):

a) feminine, e.g. Rus. *inostranka* 'foreigner (woman)' (cf. masc. *inostranec*), Pol. *aktorka* 'actress' (cf. masc. *aktor*), Cze. *sousedka* 'neighbor (woman)' (cf. masc. *soused*), Slo. *ciganka* 'gypsy (woman)' (cf. masc. *cigan*), Scr. *profesorka* 'professor (woman)' (cf. masc. *profesor*), Bul. *bălgarka* 'Bulgarian (woman)' (cf. masc. *bălgarin*);
b) diminutive, e.g. Rus. *nožka* and Pol. *nóžka* 'small foot/leg' (from *noga* 'leg'), Cze. *hlavička* 'small head' (from *hlava* 'head'), Slo. *ženka* 'small woman' (from *žena* 'woman'), Bul. *knižka* 'small book' (from *kniga* 'book');
c) nominalizer of adjectives, e.g. Rus. *višnevka* 'cherry liqueur' (from *vyšnevyj* 'cherry (adj.)'), Pol. *pustka* 'solitude' (from *pusty* 'empty'), Slo. *edinka* 'only child' (from *edini* 'only'), Scr. *dvojka* 'twin' (from *dvoje* 'two);
d) abstract nouns, e.g. Rus. *varka* 'boiling process' (from *varit'* 'boil'), Cze. *nabídka* 'offer' (from *nabídnout* 'offer'), Bul. *pridobivka* 'acquisition' (from *pridobivam* 'acquire').

Remarkably, Vaillant (1974) does not mention the agentive meaning of the suffix *-ka* in cases of deverbal word-formation (e.g. Rus. *taratorka* 'chatterbox', from *taratorit'* 'babble'; *samoučka* 'autodidact', from *sam* 'self' and *učit'* 'study'), which, as I will show in Section 3.6, encompasses a large polysemy pattern.

The suffix *-lka* is the result of further amalgamation with the agentive suffix *-la* of uncertain origin (Luschützky 2011: 94).

In modern Russian, the suffixes *-ka* and *-lka* are mainly productive in colloquial speech and, partially, in specialized terminology (AG-80, §§ 228, 231, 262, 563, 564). While the suffix *-lka* is only used in deverbal word-formation and is attached to infinitive stems, the suffix *-ka* is also employed in deadjectival and denominal derivation (cf. AG-80, §§ 381, 397, 415, 434) and compounding (cf. AG-80, § 572), in which it shows large multifunctionality.

3.2. The polysemy of agent noun suffixes

The polysemy of agent noun suffixes in the world's languages is a well-known phenomenon and has been the object of many linguistic investigations (see, among others, Booij 1986, 2007, 2010; Rosenberg 2007; Rainer 2011, 2014, 2015; Luschützky 2011; Luschützky & Rainer 2011), in which it has been shown that, in different languages, the same word-formation pattern is often employed to derive both agent and instrument nouns. Consider, for instance, the English suffix *-er*, which is used to derive both Agents (e.g. *hunt* → *hunt**er***) and Instruments (e.g. *cook* → *cook**er***). Similar patterns are very common cross-linguistically. Booij (1986, 2007, 2010) shows a similar picture for the Dutch suffix *-er*. Rainer (2011) investigates the polysemy of agentive suffixes in Latin and in the Romance languages. Luschützky (2011) deals with the same issue as regards the Slavic languages. Luschützky & Rainer (2011) take a broader perspective of this phenomenon and explore the cross-linguistic distribution of patterns of Agent-Instrument polysemy.

Depending on the semantics of the verb, agentive suffixes can form nouns that do not denote a Prototypical Agent,[18] but rather an entity that could be better described as a Beneficiary, an Experiencer, or an entity that is characterized by a certain state or condition (Luschützky & Rainer 2011: 290), such as the Russian deverbals *polučа**tel'*** 'recipient' (Beneficiary, from *polučat'* 'receive'), *ljubi**tel'*** 'lover, amateur' (Experiencer, from *ljubit'* 'love'), and *ži**tel'*** 'inhabitant' (Carrier of State, from *žit'* 'live') that I will all include under the label "Carriers of State".

The same suffixes giving rise to agent and instrument nouns are also often employed to denote place nouns. This can be exemplified by the Spanish suffix *-dor*, which, besides agent nouns (e.g. *mata**dor*** 'bullfighter', from *matar* 'kill') and instrument nouns (e.g. *tene**dor*** 'fork', from *tener* 'hold'), also forms place nouns (e.g. *come**dor*** 'dining room', from *comer* 'eat') (Rainer 2011: 10). Though more rarely, this semantic extension also takes place in Russian, and can be exemplified by the agentive suffix *-ka*, which is sometimes used to form place nouns (e.g. *ostanov**ka*** 'stop', from the verb *ostanovit'* 'stop').

[18] Traditional definitions of "Agent" usually include animacy, volitionality, and control (cf. Luraghi 1995).

Moreover, in Russian as well as in other languages, the same agentive suffixes are sometimes used in denominal or deadjectival word-formation, as instantiated by the Russian examples *gvardeec* 'member of the guard' (from the noun *gvardija* 'guard'), *lyžnik* 'skier' (from the noun *lyži* 'skis'), *glupec* 'stupid person' (from the adjective *glupyj* 'stupid'), and so on.

The main question is how to account for the multifunctionality of agentive suffixes, that is, to determine the type of relationship between the different functions carried out by the same suffixes. As pointed out by Booij (2007, 2010), there are three possible approaches to the question of polysemy in derivation:

a) the approach taken by separationist morphologists, who deny the presence of a systematic form-meaning correspondence in morphology (cf. Beard 1995) and account for these situations as cases of homonymy;
b) the monosemy approach;
c) the polysemy approach.

In the monosemy approach, a general and vague meaning is associated to the word-formation process, and the more specific interpretation of the derived word is given by the context, as well as by the general knowledge of the world. We can find examples of this approach in Booij's (1986) account of Dutch deverbal nouns in *-er* as "subject names", i.e. nouns denoting the subject of the event described by the base verb. By avoiding the term "agent nouns" and using instead the label "subject names", Booij also accounts for Non-prototypical Agents, such as the above-mentioned Carriers of State. A similar approach is adopted by Plag (2003), who considers the case of the English suffix *-er* and suggests that "the semantics of *-er* should be described as rather underspecified, simply meaning something like 'person or thing having to do with X'. The more specific interpretations of individual formations would then follow from an interaction of the meanings of base and suffix and further inferences on the basis of world knowledge" (Plag 2003: 89). However, as pointed out by Booij (2010: 78) and Rainer (2014: 342–343), in a monosemy approach, all possible interpretations and language-specific regularities cannot be accounted for and the risk is that of overgeneralization.

Booij gets to the conclusion that the best solution consists in adopting a "regular polysemy" approach (cf. Apresjan 1974), "in which a prototypical meaning forms the starting point for deriving other meanings through the semantic extension mechanisms of

metaphor and metonymy" (Booij 2010: 78). This approach lies at the basis of cognitivist and constructionist accounts of polysemy in word-formation, as will be shown in Sections 3.3 and 3.4.

3.3. Cognitive approaches to polysemy in word-formation

Panther & Thornburg (2003) propose an account of English *-er* nominals in which the suffix is seen as "a polysemous symbolic unit whose meanings are conceptually related through metaphoric and metonymic links" (Panther & Thornburg 2003: 279).

The authors posit a central sense of the suffix *-er*, that is "a human Agent who performs an action or engages in an activity to the degree that in doing so defines a primary occupation" (Panther & Thornburg 2003: 285), as in *teacher*. Then, nonhuman Agents, i.e. animals (e.g. *retriever*), plants (e.g. *creeper*), and inanimate objects (e.g. *skyscraper*) are related to the central sense of the suffix *-er* through a metaphoric extension, that is personification (Panther & Thornburg 2003: 291). The other *-er* nominals with object referents are related to the central agentive meaning through metonymic extensions. The instrumental meaning in *-er* nominals, as in *can opener*, is the result of a metonymic extension from the agentive meaning and the relation between the two senses is one of contiguity: an Instrument is an object employed by an Agent to perform an action. Clothes worn by Agents to perform certain actions (e.g. *sneakers*) are defined as "Quasi-instruments", as they are conceptually related to Instruments. Nominals denoting what Panther & Thornburg (2003: 294) call "Purpose-locations", such as *diner*, are seen as metonymic extensions of the Instrument category, as they denote objects designed for special purposes. "Purpose-patients", such as *poster*, are also intended as natural extensions of Instruments because they are designed for a special purpose. "Valued-patients" (e.g. *keeper* in the sense of an entity that is worthy of being kept) and "True-patients" (e.g. *scrambler*, i.e. scrambled egg dish) are in turn natural extensions of the "Purpose-patient" category (Panther & Thornburg 2003: 294–296).

The authors also point out that the contrast between noun-based and verb-based derivations is not as significant as assumed by other scholars (cf. Ryder 1999), and that the analytical tools available in Cognitive Linguistics allow for a unified account of all -

er nominals, independently of the nature of their base (Panther & Thornburg 2003: 284). Thus, nominals such as *Wall Streeter* are considered as conceptually similar to deverbal nominals, as a certain action or activity can still be accessed metonymically: a Wall Streeter is a person professionally employed on Wall Street (Panther & Thornburg 2003: 288–289).

The analysis of *-er* nominals denoting object referents is then summarized by the authors through a conceptual network (see Figure 3), in which darker solid arrows represent metaphoric relations, while lighter open arrows represent metonymic extensions.

Figure 3. Object-level metaphoric and metonymic extensions from the central sense of *-er* (Panther & Thornburg 2003: 297)

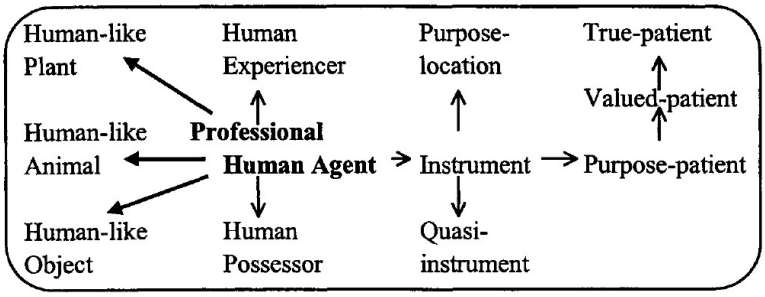
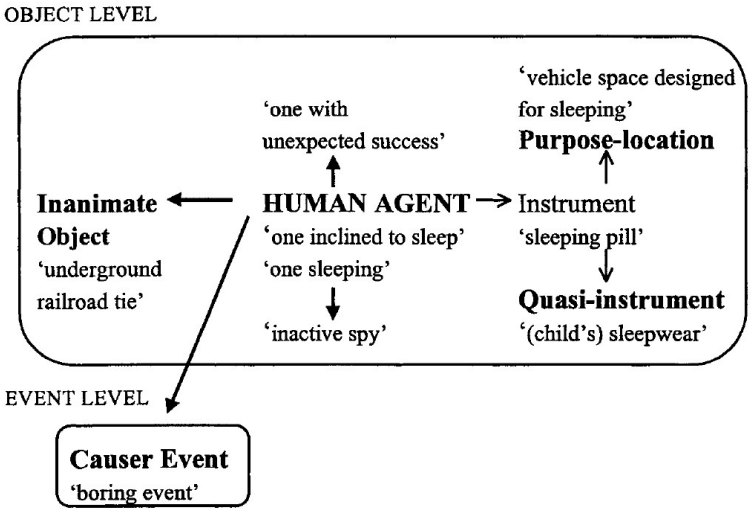


Figure 4. The polysemy of *-er* as manifest in *sleeper* (Panther & Thornburg 2003: 310)



The other group of *-er* nominals, i.e. those denoting Events, is related to the group denoting object referents through reification, that is through the EVENTS ARE OBJECTS metaphor, by which Events are seen as Agents (Panther & Thornburg 2003: 299). This category of nominals is added to the scheme just described at a different level. The two levels are represented jointly in the scheme reproduced in Figure 4, in which the polysemy of the suffix *-er* is exemplified through the nominal *sleeper*. The capability of the suffix *-er* to exploit the operations of metaphor and metonymy to produce nominals with such a wide range of different meanings is what lies at the basis of its productivity in English (Panther & Thornburg 2003: 315).

The paper by Panther & Thornburg (2003) is not the only example of an analysis of word-formation patterns in terms of metaphor and metonymy. The works by Koch (1999), Dirven (1999), Warren (1999), Basilio (2006), Janda (2010, 2011, 2014), and Nesset (2010) head in the same direction and focus on the role of metonymy in word-formation.

In particular, in Janda (2010, 2011, 2014), metonymy is employed to analyze polysemy patterns brought about by suffixation in three languages, i.e. Russian, Czech, and Norwegian. However, in this case, metonymy is not used to explain the relations between the different meanings brought about by a single suffix, but rather each of the different meanings that a suffix can create is itself explained as the result of a different metonymic shift. The suffix is the element responsible for the metonymic shift from a source (the base of the derivation process) to a target (the concept associated with the derived word), as in (1).

(1) *kupit'* 'buy' → *kupec* 'merchant'

The derived word *kupec* 'merchant' in (1) is the result of the metonymic shift ACTION FOR AGENT, which is signaled by the suffix *-ec*. For each suffix in each language, the author provides a list of the possible metonymic shifts, which are all contained in a database available at <https://opendata.uit.no/dataset.xhtml?persistentId=hdl:10037.1/10020>. Thus, for the Russian agentive suffix *-ec*, for example, the author identifies the metonymic shifts shown in Table 1.

Table 1. Metonymic shifts of the suffix *-ec* (adapted from the database by Janda 2011)

Source	Target	Source class	Target class	Source example	Target example
LEADER	ENTITY	noun	noun	*Lenin*	*leninec* 'Leninist'
LOCATION	LOCATED	noun	noun	*Kanada* 'Canada'	*kanadec* 'Canadian'
CHARACTERISTIC	ENTITY	qual adj	noun	*glupyj* 'stupid'	*glupec* 'stupid person'
ACTION	AGENT	verb	noun	*kupit'* 'buy'	*kupec* 'merchant'
GROUP	ENTITY	noun	noun	*gvardija* 'guard'	*gvardeec* 'guardsman'
EVENT	ENTITY	noun	noun	*poraženie* 'defeat'	*poraženec* 'defeatist'
ABSTRACTION	ENTITY	noun	noun	*soprotivlenie* 'opposition'	*soprotivlenec* 'oppositionist'

The metonymic shifts identified build the polysemy pattern of the suffix *-ec* in Russian, which does not only form nominals denoting Agents, but also Entities that are characterized by certain features, such as being located in a certain place, being part of a certain group, and so on. However, the author does not include compounds in the database, which prevents from getting a full polysemy schema.[19] In compounding, indeed, the suffix *-ec* is also used to form nominals denoting other types of referents, such as Instruments and Patients, as will be shown in Section 3.4.

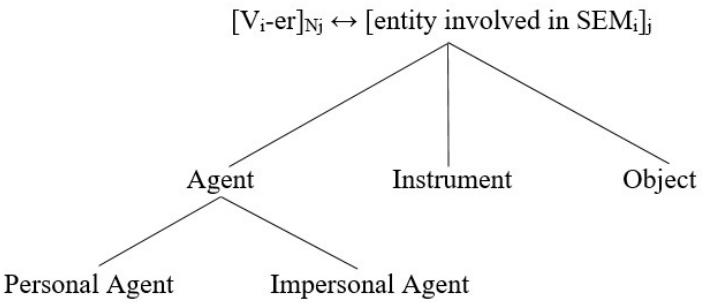
3.4. Polysemy and Construction Morphology

In a constructionist approach to word-formation, polysemy can be represented as a hierarchical schema with systematic relationships between form and meaning (cf. Chapter 2, Section 2.2). More precisely, polysemy in word-formation can itself "provide evidence for different levels of generalization and degrees of abstractness in a hierarchical lexicon" (Booij 2010: 76), since the different meanings brought about by a certain word-formation pattern can be conceived of as subschemas dominated by a single and more abstract schema. Thus, for example, the polysemy of Dutch deverbal *-er* nouns has been represented through the schema shown in Figure 5. The hierarchical representation of

[19] See Naccarato & Pinelli (Forth.) for an application of Janda's cognitive approach to the analysis of Russian compounds (both noun-based and verb-based) with agentive suffixes.

Dutch deverbal *-er* nouns specifies their different meanings and common properties by postulating different subschemas related to the most abstract one. At the highest level, we see that Dutch deverbal *-er* nouns denote some entity involved in the event described by the verb V. At a lower level, the schema specifies that this entity can act as an Agent performing the action V, an Instrument used to perform the action V, or an Object of the action V. Finally, the schema indicates that, when these nouns denote Agents, these can be either personal or impersonal.

Figure 5. The polysemy of Dutch deverbal *-er* nouns (adapted from Booij 2010: 80)



Although this schema does not indicate explicitly the type of semantic relation between the prototypical agentive meaning and the other possible interpretations, we should assume that the driving forces behind the polysemy pattern are to be found in the conceptual extensions that have been conventionalized in the lexicon of a certain language. Indeed, in previous works, the author suggests the extension schema reproduced in (2), according to which Personal Agent is the most prototypical type of Agent, since this meaning is often associated to human beings. The shift to Impersonal Agent is produced through metaphor: devices that perform certain actions can be conceived as Agents (e.g. a *printer* is something that prints). The rise of the Instrument meaning can instead be interpreted as a metonymic sense extension (in this case, we have conceptual contiguity): a *vaporizer* is an Instrument used by an Agent to vaporize.

(2) Personal Agent > Impersonal Agent > Instrument (Booij 1986: 509)

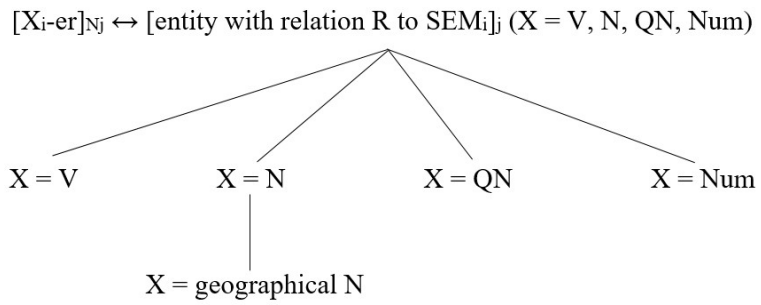
The hierarchical schema in Figure 5, instead, does not directly account for the semantic-conceptual mechanisms that exist among the different meanings produced by a single

suffix, but it is "a synchronic description of the relevant set of words, specifies the attested patterns, and the possibilities for coining new words of the different subtypes" (Booij 2010: 80).

What is not entirely convincing in the schemas proposed by Booij (1986, 2010), in my view, is the distinction between Impersonal Agent and Instrument. Indeed, it remains unclear how certain items would fall under the class of Impersonal Agents and others under that of Instruments. A *printer* is classified as an Impersonal Agent because it is "something that prints". However, for a printer to print, there has to be a Human Agent that activates the printer itself, exactly as a *vaporizer* is activated by a Human Agent to vaporize. For this reason, in the analysis carried out in the rest of this chapter, as well as in Chapter 4, I will not adopt this distinction, and I will classify both cases, i.e. *printer*-like and *vaporizer*-like items, as belonging to the class of Instruments.

One of the advantages of adopting a constructionist approach to represent polysemous word-formation constructions is that this approach enables to easily account for polysemy patterns that are not the result of cognitive extensions, but are due to historical coincidence, such as the case of the Dutch agentive suffix *-er*, which derives from the Latin suffix *-arius*, and the homophonous Dutch suffix *-er* used to form nouns denoting inhabitants of a place, which derives from the Germanic suffix *-warja* (Booij 2010: 82). Booij (2010) postulates a higher-level, more abstract schema dominating the more specific subschemas accounting for the different meanings of the Dutch suffix *-er*. The schema in Figure 6 shows the polysemy pattern of the Dutch suffix *-er*. The subschema X = V represents cases of deverbal derivation and stands for the schema shown in Figure 5, which I define as "polysemy in the narrow sense", as it includes the meanings that could be described in terms of semantic extension. This higher-level schema accounts for the other possible functions of the suffix *-er* in Dutch, which are not the result of semantic extensions. Thus, we see cases of denominal derivation (X = N), which includes the pattern forming nouns denoting inhabitants of a place that I mentioned above. In addition, we find cases in which the base of the word-formation process is constituted by a quantifier and a noun (X = QN), as in *tienponder* 'ten-pounder' (from *tien* 'ten' and *pond* 'pound'), or by a numeral (X = Num), as in *dertiger* 'person in his thirties' (from *dertig* 'thirty').

Figure 6. The polysemy of Dutch *-er* nouns (adapted from Booij 2010: 84)



Let us consider again the Russian agentive suffix *-ec*. Investigations on the semantics of this suffix based on the descriptions of the AG-80 (§§ 216, 287, 288, 332, 559, 575) and on the analysis of the word-formation database of the RNC (cf. Chapter 4) show that, when the word-formation process is deverbal, this suffix can have the following meanings:

a) Prototypical Agents, such as *kup**ec*** 'merchant' (from *kupit'* 'buy') or *basnopis**ec*** 'fable writer' (from *basnja* 'fable' and *pisat'* 'write');
b) Carriers of State, such as *čestoljub**ec*** 'ambitious person' (from *čest'* 'honor' and *ljubit'* 'love');
c) Instruments, such as *kitobo**ec*** 'whaler' (from *kit* 'whale' and *bit'* 'beat, kill');
d) Objects/Patients, such as *novobran**ec*** 'recruit' (from *novyj* 'new' and *brat'* 'take').[20]

One thing that has to be pointed out regarding the meaning that goes under the label of "Carriers of State" is that, in such cases, the difference with respect to Prototypical Agents only depends on the type of verb lying at the basis of the word-formation process. That is to say, the class of Carriers of State is kept separate from the class of Prototypical Agents based on different reasons with respect to Instruments and Objects/Patients. Whereas we know that *kitoboec* 'whaler' is an Instrument and not an Agent for reasons that depend on our knowledge of the world, and not because of the verb employed in this compound (and, in fact, we find the compound *bykoboec* 'bull fighter', based on the same verb *bit'*

[20] This meaning though is rare, and, in the compound *novobranec* 'recruit' (which is the only example of this kind found in the word-formation database of the RNC), the Object/Patient meaning seems to be acquired through the passive participial suffix *-n-*, which is found before the agentive suffix *-ec*.

'beat, kill', denoting an Agent), the fact that *čestoljubec* 'ambitious person' is defined as a Carrier of State and not as a Prototypical Agent merely depends on the semantics of the base verb. Indeed, whereas the verbs lying at the basis of nouns denoting Prototypical Agents or Instruments usually indicate a change of state (e.g. *kupit'* 'buy', *pisat'* 'write', *bit'* 'beat, kill'), this is not true for verbs lying at the basis of nouns denoting Carriers of State, which indicate a state (e.g. *ljubit'* 'love').

The suffix *-ec* is also productive in denominal and deadjectival derivation, giving rise to derivatives and compounds with a wide range of meanings that can denote:

a) a person with a certain geographical origin, such as *kanad**ec*** 'Canadian' (from *Kanada* 'Canada') or *novozeland**ec*** 'New Zealander' (from *Novaja Zelandija* 'New Zealand');
b) a member of a certain group, e.g. *krasnoarme**ec*** 'member of the Red Army' (from *Krasnaja Armija* 'Red Army');
c) an entity showing a relation of possession with the base, e.g. *dvoežen**ec*** 'bigamist' (from *dve ženy* 'two wives');
d) an agent performing a certain activity in a certain place (e.g. *ochotnorjad**ec*** 'person working at the *Ochotnyj Rjad*'),[21] or using a certain material (e.g. *krasnoderev**ec*** 'carpenter using fine wood', from *krasnoe derevo* 'fine wood');
e) a person with a certain characteristic, such as *glup**ec*** 'stupid person' (from *glupyj* 'stupid').

The semantic vagueness of this suffix has led some scholars to conclude that it would be more adequate to postulate the existence of "homonymous suffixes instead of assigning complex polysemy to one single morpheme" (Luschützky 2011: 87).

In a Construction Morphology approach, we can assume different subschemas for each meaning. Thus, we first identify a general schema to which all the derivatives and compounds formed with the suffix *-ec* can be traced back, as the one represented in (21), and then we identify the subschemas embedded within the more general schema, as shown in Table 2.

[21] *Ochotnyj Rjad* is the name of a street (and a mall) in the center of Moscow. The origin of this name dates back to the XVII century and indicates a market where it was allowed to trade game (*ochotnyj* is the adjective from *ochota* 'hunting').

(3) $[(X_k\text{-LV-})Y_i\text{-}ec]_{Nj} \leftrightarrow$ [entity with relation R to SEM_i; SEM_i is specified by SEM_k in compounding]

Table 2 shows the subschemas for the possible constructions with the suffix *-ec*. Remarkably, deverbal and denominal processes can give rise to both simple derivatives and compounds, while deadjectival processes only form derivatives.[22]

Table 2. Polysemy of the suffix *-ec*

Subschema	**Meaning**	**Examples**[23]
$[(X\text{-LV-})V\text{-}ec]_N$	[agent of V (X)]	*kup**ec*** 'merchant', *basnopis**ec*** 'fable writer'
	[carrier of the state V (X)]	*čestoljub**ec*** 'ambitious person'
	[instrument of V (X)]	*kitobo**ec*** 'whaler'
	[object/patient of V X]	*novobran**ec*** 'recruit'
$[(X\text{-LV-})N\text{-}ec]_N$	[inhabitant of (X) N]	*kanad**ec*** 'Canadian', *novozeland**ec*** 'New Zealander'
	[member of (X) N]	*gvarde**ec*** 'member of the guard', *krasnoarme**ec*** 'member of the Red Army'
	[entity possessing X N]	*dvoežen**ec*** 'bigamist'
	[agent with some relation to X N]	*ochotnorjad**ec*** 'person working at the *Ochotnyj Rjad*', *krasnoderev**ec*** 'carpenter using fine wood'
$[A\text{-}ec]_N$	[person with the characteristic A]	*glup**ec*** 'stupid person'

For certain subschemas, only compounds are found, i.e. [object/patient of V X], [entity possessing X N] and [agent with some relation to X N]. This way of representing the polysemy of the suffix *-ec* allows recognizing the presence of subgroups that share similar characteristics within the more abstract schema including all the constructions formed with this suffix. As Booij (2010: 84) points out, "we need morphological subschemas in order to account for the relevant subregularities. In a hierarchical lexicon, such subschemas can be specified without losing the possibility of expressing that they are instantiations of more general word formation schemas". Thus, the schema displayed in

[22] Adjectives can never be the main constituents of compounds in such constructions, but they can only modify the main constituents, e.g. *krasnoarmeec* 'member of the Red Army, where the adjective *krasnyj* 'red' modifies the nominal constituent *armija* 'army'.
[23] Examples are retrieved from the RNC word-formation database and from the AG-80.

Table 2 allows accounting for all the subregularities of the general schema [(X-LV-)Y-*ec*]$_N$, including both cases of polysemy "in the narrow sense", which are the result of deverbal word-formation, and cases of historical coincidence (i.e. cases of denominal and deadjectival word-formation), which result in a series of meanings that are hardly connectable to the meanings produced by deverbal word-formation.

3.5. Diachronic approaches

Other approaches to the study of polysemy in word-formation are more focused on the diachronic development of semantic extensions.

Simone (2000) describes the property of suffixes to give rise to a series of different meanings in terms of "lexical cycles". The author intends a "lexical cycle" as "an algorithm by which a lexical form (word, fixed phrase, etc.) having a base meaning M1 develops in diachrony a series of other meanings (M2, M3, …, Mn) that are not associated to it arbitrarily, but that are related to M1 in some way, though being semantically distinct from it, and that form a matrix of possible meanings"[24] (Simone 2000: 264). Within this approach, the different meanings produced by a suffix are seen as the result of diachronic developments. Thus, for instance, the author shows that the Latin word *ingressus* only had two meanings: it could indicate the action of entering or the beginning of an operation. In Italian, instead, the word *ingresso* developed other senses and it can refer to the process of entering into a place, the place that serves as a passage for those who enter, the instrument that legitimizes the entrance into a place (e.g. a ticket), and the person who enters a place with a ticket (Simone 2000: 262–263). These semantic developments are so frequent and so regular that they cannot be seen as the result of chance, but they must be understood as being part of a mechanism of sense extension (Simone 2000: 264).

Insacco (2014) bases her analysis on the theoretical framework developed by Simone (2000), and accounts for Italian nominalizations by resorting to the notion of

[24] "un algorithme par lequel une forme lexicale (mot, syntagme figé, etc.) douée d'un sens de départ M1, développe en diachronie une série d'autres sens (M2, M3, …, Mn), qui ne s'associent pas à elle de façon arbitraire, mais qui sont reliés à M1 d'une certain façon, tout en étant sémantiquement distinct de lui, et qui forment une matrice prédéfinie de sens possibles".

lexical cycle. In her work, the author analyzes the Italian suffixes *-ata*, *-mento* and *-zione*, and identifies the most common lexical cycles, which are summarized in Table 3.

Table 3. Most common lexical cycles for the Italian suffixes *-ata*, *-mento* and *-zione* (adapted from Insacco 2014: 123–124)

Lexical cycle	**Example**
Event → Abstract object	*traduzione* 'translation' (process and result)
Event → Concrete object	*costruzione* 'construction' (process and result)
Event → Manner	*camminata* 'walk' (process and way of walking)
Event → Place	*fermata* 'stop' (process and place)
Event → Human group	*protezione* 'protection' (process and human group, as in *Protezione Civile* 'Civil Defence')
Event → Time	*entrata* 'entrance' (process and time step)

The analysis proposed in Table 3 shows that the Italian suffixes considered, which were initially used with a processual meaning, developed in diachrony a series of semantic extensions, i.e. abstract or concrete object, manner, place, human group, and time.

Luschützky (2011), Luschützky & Rainer (2011, 2013) and Rainer (2011, 2015) question the polysemy approach proposed in constructionist and cognitive frameworks, and argue that only a diachronic approach can explain all cases of affixal multifunctionality, which cannot always be analyzed as the result of semantic extension. Luschützky (2011: 95) points out that "at least in languages with abundance of derivational affixes, the phenomenon has many more aspects, the often-cited extension from animate agentivity to inanimate instrumentality being only a minor case in point". Rainer (2015: 1308) critically observes that "[t]horough diachronic investigations could turn out to be a healthy antidote against cognitive analyses" and that "[i]t remains to be seen whether we need mechanisms of conceptual extension at all".

By adopting a diachronic approach, the authors suggest that the behavior of agentive formations in the world's languages cannot always be accounted for in terms of polysemy, but in some cases the multifunctionality of agentive suffixes is rather due to factors such as sound change, borrowing, and ellipsis. Rainer (2011) points out, for instance, that the outcomes of the Latin agentive suffix *-tor* and instrumental *-tōrium* have merged in some of the Romance languages by regular sound change, i.e. Romanian,

Provençal and Catalan. Thus, the multifunctionality of the Romanian suffix *-tor* (e.g. *organizator* 'organizer', *găuritor* 'drill', *dormitor* 'bedroom') is explained in terms of sound change rather than in terms of polysemy (Rainer 2011: 9–10; Luschützky & Rainer 2011: 328). Rainer (2011: 12) and Luschützky & Rainer (2011: 329) claim that the instrumental meaning of the French suffix *-eur* has been enhanced by loan translations from English (e.g. *ventilateur* < *ventilator*) starting from the second half of the 18th century. Luschützky & Rainer (2011: 328) report on cases of ellipsis by adducing the example of the French suffix *-ant*, which is the present participle marker reanalyzed as a deverbal agentive suffix as the result of ellipsis in noun-participle constructions (e.g. *nettoyant* < *produit nettoyant* 'cleaning agent').

The importance of diachronic investigations is recognized by Booij himself, who admits that in some cases ellipsis and homonymization are at play and that "patterns of polysemy do not necessarily follow from cognitive extension schemas but may also be due to historical coincidence" (Booij 2010: 81). However, although the author recognizes the importance of investigating the origin of polysemy, his approach is purely synchronic and mainly focuses on the way in which "the conceptual extension mechanisms have been conventionalized in the lexicon of a particular language" (Booij 2010: 80).

Arcodia (2014: 133) also points out that, although "[t]he constructionist approach can be very effective in representing the synchronic relationship between word formation patterns", "[…] it is inadequate to account for how derivational polysemy arises". After discussing the polysemy patterns of the Ewe suffix *-ví* and the Mandarin Chinese suffix *-bā*, the author concludes that, while in a Construction Morphology approach polysemy is intended as "a 'family' of specific subschemas dominated by a general schema" whose origin is not specified, in a diachronic approach we find an opposite picture, "as it is the existence of subschemas that 'broadens' the generic meaning of the pattern" (Arcodia 2014: 136–137).

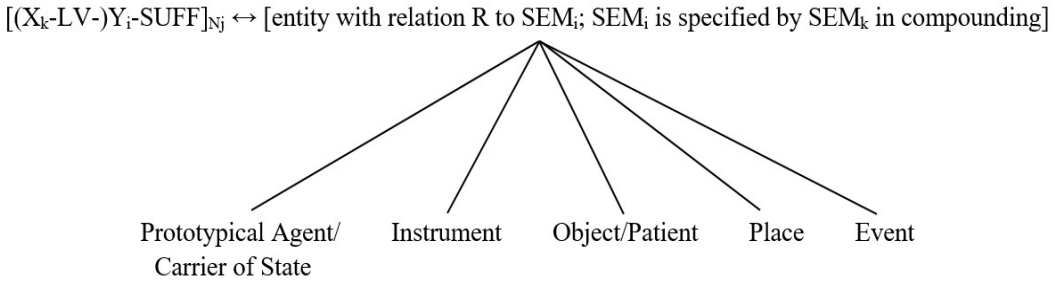
3.6. The polysemy of Russian deverbal agent noun suffixes

Sticking to the approach adopted throughout this study, I provide here a synchronic account of the polysemy of Russian agent noun suffixes by making use of the analytical

tools of CxM. From now on, only cases of deverbal word-formation are considered (i.e. only cases of polysemy "in the narrow sense"), consistently with the the analysis carried out in Chapter 4, whose object is constituted by Russian synthetic compounds with agentive meanings. Thus, although some of the suffixes considered (*-ec*, *-nik*, *-ščik/čik*, and *-ka*) are also productive in denominal and/or deadjectival derivation and compounding, these cases will not be taken into consideration.

When agentive suffixes are embedded in a deverbal word-formation process in Russian, the possible outcomes, apart from nouns denoting Prototypical Agents and Carriers of State, are nouns denoting Instruments, Objects/Patients, Places and Events. By adopting a constructionist approach, I represent the polysemy of Russian deverbal agent noun suffixes in the form of a hierarchical lexicon, similarly to Booij (2010: 80), as shown in Figure 7.

Figure 7. Polysemy of Russian deverbal agent noun suffixes



The schema shown in Figure 7 is a synchronic representation of the polysemy of Russian deverbal agentive suffixes, i.e. a representation of how the semantic extension mechanisms discussed above (cf. Sections 3.3 and 3.4) have been conventionalized in the Russian lexicon. This schema does not represent explicitly the semantic-conceptual mechanisms existing among the different meanings produced by agentive suffixes (which, as we have seen, must be intended as mechanisms of metonymic and metaphoric extension), but accounts for the synchronically possible coinages of new words of the different subtypes.

Note that not all the Russian suffixes considered are capable of giving rise to nouns with all the meanings represented in Figure 7. In fact, only the suffix *-ka* seems to

have such capability, while most suffixes only cover some of the possible meanings (see Table 4). In particular, all the suffixes considered share the agentive meaning, which, in some cases and depending on the semantics of the verbal bases, can be less prototypical and include cases of nouns denoting Carriers of State (cf. *-tel'*, *-nik*, *-ec*, and *-ka* in Table 4). Most suffixes can also bring about the instrumental meaning (cf. *-tel'*, *-nik*, *-ščik/čik*, *-ec*, *-ka*, and *-lka* in Table 4). Only few suffixes can also form nouns denoting Objects/Patients (i.e. *-nik*, *-ec*, and *-ka*), Places (i.e. *-ka* and *-lka*), and Events (only *-ka*).

The suffixes under examination form both derivatives and compounds, showing the structures represented in schemas (4) and (5).

(4) $[V_i\text{-SUFF}]_{Nj} \leftrightarrow$ [entity or process of $SEM_i]_j$

(5) $[X_k\text{-LV-}V_i\text{-SUFF}]_{Nj} \leftrightarrow$ [entity or process of SEM_i; SEM_i is specified by SEM_k]

In the case of compound formations (5), an element X (i.e. a bound stem belonging to one of the following parts of speech: noun, adjective/adverb, pronoun, numeral, and followed by a linking vowel) has some relation R with respect to the verbal base V. As will be shown in Chapter 4 (cf. Section 4.3.2), this relation is of argumental nature, i.e. the element X is an argument of the verb V. However, the semantic role of X with respect to V can vary and X can act as a Theme/Patient, Recipient, Instrument, Location, Path, Goal, etc. with respect to the verb V.

The polysemy subschemas for Russian deverbal agent noun suffixes are shown in Table 4. The examples, including both derivatives and compounds, are retrieved from the word-formation database of the RNC and from the AG-80.

Table 4. Polysemy of Russian deverbal agent noun suffixes

Subschema	**Meaning**	**Examples**
$[(X\text{-})V\text{-}tel']_N$	[agent of V (X)]	*učі**tel'*** 'teacher', *vzjatkoda**tel'*** 'briber'
	[carrier of the state V (X)]	*ljubi**tel'*** 'lover', *blagožela**tel'*** 'benevolent person'
	[instrument of V (X)]	*nakopi**tel'*** 'storage device', *ognetuši**tel'*** 'fire extinguisher'
$[(X\text{-})V\text{-}nik]_N$	[agent of V (X)]	*rabot**nik*** 'worker', *zakonoprestup**nik*** 'criminal'
	[carrier of the state V (X)]	*muče**nik*** 'martyr', *čelovekonenavist**nik*** 'misanthrope'

	[instrument of V (X)]	*okuč**nik*** ‘hilling machine’, *rukomoj**nik*** ‘washstand’
	[object/patient of V X]	*vol’nootpuščen**nik*** ‘released slave’
[(X-)V-*šč**ik/čik*]$_N$	[agent of V (X)]	*tancov**ščik*** ‘dancer’, *granatomet**čik*** ‘grenadier’
	[instrument of V (X)]	*bombardirov**ščik*** ‘bomber plane’, *toplivozaprav**ščik*** ‘refueller’
[(X-)V-*lščik*]$_N$	[agent of V (X)]	*nosi**l’ščik*** ‘porter’, *nalogoplate**l’ščik*** ‘tax payer’
[(X-)V-*ec*]$_N$	[agent of V (X)]	*kup**ec*** ‘merchant’, *basnopis**ec*** ‘fable writer’
	[carrier of the state V X]	*čestoljub**ec*** ‘ambitious person’
	[instrument of V (X)]	*rez**ec*** ‘cutter’, *kitobo**ec*** ‘whaler’
	[object/patient of V X]	*novobran**ec*** ‘recruit’
[(X-)V-*lec*]$_N$	[agent of V]	*kormi**lec*** ‘breadwinner’
	[carrier of the state V (X)]	*strada**lec*** ‘martyr’, *domovlade**lec*** ‘landlord’
[(X-)V-*ka*]$_N$	[agent of V (X)]	*tarator**ka*** ‘chatterbox’, *samouč**ka*** ‘autodidact’
	[carrier of the state V (X)]	*zaznaj**ka*** ‘conceited person’, *vseznaj**ka*** ‘know-all’
	[instrument of V (X)]	*ter**ka*** ‘grater’, *chleborez**ka*** ‘bread slicer’
	[object/patient of V (X)]	*zapis**ka*** ‘note’, *samokrut**ka*** ‘roll-up’
	[place related to V (X)]	*ostanov**ka*** ‘stop’, *lesopil**ka*** ‘saw-mill’
	[event related to V (X)]	*čist**ka*** ‘cleaning’, *gazoočist**ka*** ‘gas treatment’
[(X-)V-*lka*]$_N$	[agent of V]	*gada**lka*** ‘fortune-teller’
	[instrument of V (X)]	*zažiga**lka*** ‘lighter’, *gazonokosi**lka*** ‘lawnmower’
	[place related to V]	*kuri**lka*** ‘smoking room’

The suffix *-tel’* is usually indicated as being one of the major and most productive agentive suffixes in Russian. As can be observed in Table 4, it is used both in derivation and in compounding to form:

a) Prototypical Agents, e.g. *učitel’* ‘teacher’ (from *učit’* ‘teach’) and *vzjatkodatel’* ‘briber’ (from *vzjatka* ‘bribe’ and *dat’* ‘give’);
b) Carriers of State, e.g. *ljubitel’* ‘lover’ (from *ljubit’* ‘love) and *blagoželatel’* ‘benevolent person’ (from *blago* ‘good’ and *želat’* ‘wish’);
c) Instruments, e.g. *nakopitel’* ‘storage device’ (from *nakopit’* ‘accumulate’) and *ognetušitel’* ‘fire extinguisher’ (from *ogon’* ‘fire’ and *tušit’* ‘extinguish’). This suffix is never used to form nouns denoting Objects/Patients, Places, or Events.

The suffix *-nik* is often described as mainly productive in denominal and deadjectival word-formation. However, it also shows some productivity in deverbal derivation and compounding, where it gives rise to nouns denoting:

a) Prototypical Agents, e.g. *rabotnik* 'worker' (from *rabotat'* 'work') and *zakonoprestupnik* 'criminal' (from *zakon* 'law' and *prestupit'* 'breach');
b) Carriers of State, e.g. *mučenik* 'martyr' (from *mučit'* 'torture') and *čelovekonenavistnik* 'misanthrope' (from *čelovek* 'person' and *nenavidet'* 'hate');
c) Instruments, e.g. *okučnik* 'hilling machine' (from *okučit'* 'hill') and *rukomojnik* 'washstand' (from *ruka* 'hand' and *myt'* 'wash');
d) Objects/Patients, e.g. *vol'nootpuščennik*[25] 'released slave' (from *vol'nyj* 'free' and *otpustit'* 'let go').

The suffixes *-ščik/čik* and *-l'ščik* show a more restricted polysemy pattern with respect to *-nik*. In particular, the suffix *-ščik/čik* forms (both in derivation and in compounding) nouns denoting:

a) Prototypical Agents, e.g. *tancovščik* 'dancer' (from *tancevat'* 'dance') and *granatometčik* 'grenadier' (from *granata* 'grenade' and *metat'* 'throw');
b) Instruments, e.g. *bombardirovščik* 'bomber plane' (from *bombardirovat'* 'bomb') and *toplivozapravščik* 'refueller' (from *toplivo* 'fuel' and *zapravit'* 'fill').

The suffix *-l'ščik* is only used to form nouns denoting Agents (both in derivation and in compounding), e.g. *nosil'ščik* 'porter' (from *nosit'* 'carry') and *nalogoplatel'ščik* 'tax payer' (from *nalog* 'tax' and *platit'* 'pay').

The suffix *-ec* shows the same polysemy pattern as the suffix *-nik*, and builds nouns denoting:

a) Prototypical Agents, e.g. *kupec* 'merchant' (from *kupit'* 'buy') and *basnopisec* 'fable writer' (from *basnja* 'fable' and *pisat'* 'write');
b) Carriers of State (though, apparently, only in compounding), e.g. *čestoljubec* 'ambitious person' (from *čest'* 'honour' and *ljubit'* 'love');
c) Instruments, e.g. *rezec* 'cutter' (from *rezat'* 'cut') and *kitoboec* 'whaler' (from *kit* 'whale' and *bit'* 'kill');
d) Objects/Patients, e.g. *novobranec* 'recruit' (from *novyj* 'new' and *brat'* 'take').

As I have discussed above, I only found the latter meaning in the compound *novobranec* 'recruit', where the Object/Patient meaning seems to be acquired through the passive

[25] Note that this compound is based on the passive participial form of the verb *otpustit'* 'let go' (i.e. *otpuščennyj*), which confers the object/patient meaning to the word.

participial suffix *-n-*, which is found before the agentive suffix *-ec*. Its related suffix *-lec* shows a more restricted polysemy pattern, and is only employed to denote:

a) Prototypical Agents (though examples are only found in derivation), e.g. *kormilec* 'breadwinner' (from *kormit'* 'feed');
b) Carriers of State, e.g. *stradalec* 'martyr' (from *stradat'* 'suffer') and *domovladelec* 'landlord' (from *dom* 'house' and *vladet'* 'own').

As mentioned above, the suffix *-ka* shows the largest polysemy pattern and gives rise (both in derivation and in compounding) to nouns denoting:

a) Prototypical Agents, e.g. *taratorka* 'chatterbox' (from *taratorit'* 'babble') and *samoučka* 'autodidact' (from *sam* 'self' and *učit'* 'study, learn');
b) Carriers of State, e.g. *zaznajka* 'conceited person' (from *zaznavat'sja* 'be arrogant') and *vseznajka* 'know-all' (from *vse* 'all' and *znat'* 'know');
c) Instruments, e.g. *terka* 'grater' (from *teret'* 'grate') and *chleborezka* 'bread slicer' (from *chleb* 'bread' and *rezat'* 'cut');
d) Objects/Patients, e.g. *zapiska* 'note' (from *zapisat'* 'write down') and *samokrutka* 'roll-up' (from *sam* 'self' and *krutit'* 'roll');
e) Places, e.g. *ostanovka* 'stop' (from *ostanovit'sja* 'stop') and *lesopilka* 'saw-mill' (from *les* 'wood' and *pilit'* 'saw');
f) Events, e.g. *čistka* 'cleaning' (from *čistit'* 'clean') and *gazoočistka* 'gas treatment' (from *gaz* 'gas' and *očistit'* 'purify').

Its related suffix *-lka* shows a more restricted polysemy pattern and forms:

a) Prototypical Agents (only in derivation), e.g. *gadalka* 'fortune-teller' (from *gadat'* 'guess');
b) Instruments, e.g. *zažigalka* 'lighter' (from *zažigat'* 'light') and *gazonokosilka* 'lawnmower' (from *gazon* 'lawn' and *kosit'* 'mow');
c) Places (only in derivation), e.g. *kurilka* 'smoking room' (from *kurit'* 'smoke').

Interestingly, while some of the suffixes show the same polysemy patterns in derivation and in compounding (cf. suffixes *-tel'*, *-ščik/čik*, *-l'ščik*, and *-ka*), for other suffixes, instantiations of certain meanings were only found either in derivation or in compounding. So, the suffixes *-lec* and *-lka* show larger polysemy patterns in derivation, while the suffixes *-nik* and *-ec* show larger polysemy patterns in compounding (cf. also

Naccarato & Pinelli Forth. for a discussion on the different behaviors of agentive suffixes in derivation and compounding).

3.7. The suffixless construction and its polysemy

Deverbal agent nouns in Russian are also formed through conversion or zero derivation, as in *storož* 'guard', from the verb *storožit'* 'guard'.

Before discussing the polysemy of this construction in Russian, I will briefly address the controversial issue of zero derivation in word-formation. As is well known, word-formation processes such as the formation of the English verb "milk" (meaning "extract milk", from the noun "milk") are usually treated as cases of conversion or zero derivation, since no overt affix is added to the source word to derive the new word.

The question of how to define such cases, i.e. as cases of conversion or zero derivation, has been addressed by many scholars with regard to different languages.

Plag (2003), for example, investigates English verb-deriving affixes and concludes that there is no basis for assuming the existence of a zero suffix in that case, since there is no overt suffix in English which expresses exactly the same range of meanings as conversion. Thus, by applying the so-called "overt analogue criterion" (Plag 2003: 140), which states that the postulation of a zero suffix is justified only in those cases where there is also an overt suffix that expresses exactly the same meaning, to the case of English verb-deriving affixes, Plag concludes that it is more reasonable to talk of non-affixational conversion, rather than of zero derivation (Plag 2003: 140–143).

Booij (2005a: 39), referring again to verb-deriving affixes in English and other languages, claims that "(i)f one wants to treat conversion as a kind of affixation, one is forced to assume a zero-morpheme that is added to the input word. However, there is no independent evidence for such a zero-affix, and we do not even know if the zero-morpheme should be taken to be a prefix or a suffix".

Itkin (2013: 52–53) discusses the case of Russian deverbal nouns which do not show an overt suffix and summarizes the three approaches to this issue in the Russian tradition:

a) these nouns can be regarded as cases of zero derivation, as in the approach adopted by the AG-80 (specifically, by Lopatin and Uluchanov, who are responsible for the word-formation section of the AG-80) and by Worth (1972);
b) the word-formation process at play in these cases is conversion, as claimed by Mel'čuk (1975/1997);
c) the derivation process of these nouns is not pointed out in any particular way, but it is rather regarded as equivalent to words with no overt ending such as *sneg* 'snow', as in the approach adopted by Kuznecova & Efremova (1986).

However, the author points out that there does not seem to be a meaningful difference between the first two approaches, and that it is not hard to imagine a case in which the same phenomenon could be described either in terms of conversion or zero derivation with no significant differences (Itkin 2013: 54).

In most studies on Russian word-formation, it is a common habit to describe such cases as instances of zero derivation (cf. AG-80 §§ 446–467, 579–584; Molinsky 1973; Townsend 1980: 23–25; Zemskaja 2005: 49, 2011: 284).

Following Booij (2005a), I argue that conversion is a more appropriate term to describe such a phenomenon in word-formation, and in the rest of this study I will consistently employ the term "suffixless construction" to refer to such cases. However, for practical needs and reasons of graphical representation (especially regarding the figures and tables in Chapters 4 and 5), I represent this construction by using the symbol *-ø*.

The schemas representing this construction in derivation and compounding are shown in (6) and (7).

(6) $[V_i\text{-}Ø]_{Nj} \leftrightarrow$ [entity or process of SEM_i]$_j$

(7) $[X_k\text{-}LV\text{-}V_i\text{-}Ø]_{Nj} \leftrightarrow$ [entity or process of SEM_i; SEM_i is specified by SEM_k]$_j$

Similarly to the suffix *-ka*, the suffixless construction is able to cover all the meanings represented in Figure 7. As summarized in Table 5, this construction can form nouns denoting:

a) Prototypical Agents, e.g. *storož* 'guard' (from *storožit'* 'guard') and *èkskursovod* 'tour guide' (from *èkskursija* 'tour' and *vodit'* 'guide');
b) Carriers of State, e.g. *knigoljub* 'bibliophile' (from *kniga* 'book' and *ljubit'* 'love');
c) Instruments, e.g. *nasos* 'pump' (from *nasosat'* 'pump) and *granatomet* 'grenade launcher' (from *granata* 'grenade' and *metat'* 'throw');
d) Objects/Patients, e.g. *šov* 'seam' (from *šit'* 'sew') and *vetrogon* 'frivolous person' (from *veter* 'wind' and *gonjat'* 'drive away');
e) Places, e.g. *sklad* 'warehouse' (from *skladyvat'* 'stack');
f) Events, e.g. *pusk* 'starting' (from *puskat'* 'start') and *snegopad* 'snow fall' (from *sneg* 'snow' and *padat'* 'fall').

Table 5. Polysemy of the suffixless construction

Subschema	**Meaning**	**Examples**
$[(X\text{-})V\text{-}ø]_N$	[agent of V (X)]	*storož* 'guard', *èkskursovod* 'tour guide'
	[carrier of the state V X]	*knigoljub* 'bibliophile'
	[instrument of V (X)]	*nasos* 'pump', *granatomet* 'grenade launcher'
	[object/patient of V (X)]	*šov* 'seam', *vetrogon* 'frivolous person'
	[place related to V]	*sklad* 'warehouse'
	[event related to V (X)]	*pusk* 'starting', *snegopad* 'snow fall'

The role of the suffixless construction in the formation of deverbal agent nouns in Russian seems particularly interesting if we compare it to the other constructions available. As will be shown in Chapter 4 (cf. Section 4.6), where the rivalry of the different constructions that form compound agent nouns is compared at the word level, in 82.8% of the cases, i.e. in 24 out of 29 pairs of rival compounds found in the database, the opposition regards a suffixed construction (in most cases *-ec*) and the suffixless construction. The multifunctionality of this construction could be seen as a determining factor in these respects, in the sense that the suffixless construction could potentially substitute and oust the other rival constructions, most of which show lower multifunctionality. However, as will be shown in Chapter 4, the rival constructions under examination show different distributions according to a number of different formal and

semantic parameters, as well as to other factors, such as diachrony, textual genres and registers.

3.8. Summary

The present chapter has dealt with the polysemy of agent nouns. After a brief overview of the origin and etymology of the agentive suffixes employed in the formation of compounds in Russian, I described the polysemy patterns that characterize these suffixes not only in Russian, but also in other languages.

In Sections 3.3 to 3.5, I have reported on different approaches to polysemy in word-formation, i.e. cognitive, constructionist and diachronic approaches. The common basis of these approaches consists in the fact that they all regard polysemy in word-formation as a regular phenomenon which shows patterns that are common to many languages.

In Section 3.6, I have provided a synchronic constructionist account of the polysemy of Russian deverbal agent noun suffixes, and particularly, of those suffixes that are featured in the word-formation database of the RNC, from which the compounds analyzed in Chapter 4 have been retrieved. The analysis has shown that these suffixes show varying patterns of polysemy that not only include the meanings of Prototypical Agent and Carrier of State, along with that of Instrument, which build the typical polysemy pattern of agentive suffixes, but they can also extend to comprise other meanings, i.e. Object/Patient, Place, and Event, though less frequently.

Section 3.7 has treated separately the suffixless construction in Russian derivatives and compounds, showing that this construction displays a large polysemy pattern, which covers all the range of meanings covered by the other suffixed agentive constructions.

The next chapters will compare the features of such rival suffixes in compounding, and will address questions concerning their formal and semantic properties, their diachronic and stylistic distribution, and their productivity in contemporary Russian.

4. Compound agent nouns in Russian: A qualitative and quantitative analysis

The present chapter describes the qualitative and quantitative analyses carried out for this study and compares rival constructions giving rise to compound agent nouns in Russian. In Section 4.1, I briefly describe the RNC, which I have used throughout this study to check the data and retrieve examples. Section 4.2 outlines the processes of extraction and selection of the data from the RNC word-formation database. In Section 4.3, I propose a comparison of the rival constructions selected, based on a number of formal and semantic parameters. In Section 4.4, some statistical tests (i.e. binomial and multinomial logistic regression, conditional inference trees and random forests) are carried out to compare the behavior of rival word-formation constructions. Section 4.5 focuses on the distribution of rival constructions over time, and across genres and registers. In Section 4.6, I discuss rivalry at the word level and analyze 29 doublets (i.e. pairs of compounds sharing the same lexical bases, but belonging to different constructions) that I found in the database.

4.1. The Russian National Corpus

The RNC is a corpus of modern Russian that was created in 2003 by the Institute of Russian Language (Russian Academy of Sciences). It presently contains over 600 million words and is one of the most important resources worldwide for the study of Russian.

The RNC includes various subcorpora, the largest being the main corpus, a collection of texts that represents standard Russian and contains written texts from the 18th century to the present day. The RNC also includes the following subcorpora: a syntactic Treebank, the corpus of newspapers and news, the parallel corpus, the corpus of Russian dialects, the corpus of Russian poetry, the Russian-for-schools corpus, the corpus of spoken Russian, the accentuated corpus, the multimedia corpus, and a diachronic corpus including Old Church Slavonic, Old and Middle Russian.

The RNC is provided with annotation at different levels. The metatextual annotation gives information about the texts: the text's title, information about the author(s), the creation date, the genre to which the text belongs, the text's size, and so forth. The morphosyntactic annotation includes lemmatization, part-of-speech tagging, and the annotation of inflectional categories. For part of the texts included in the main corpus, disambiguation is also available, i.e. homonyms have been manually disambiguated. The RNC is also provided with lexico-semantic annotation, which includes different lexical and semantic tags, according to the part of speech being tagged: taxonomy (for nouns, verbs, adjectives and adverbs), mereology (for concrete and abstract nouns), topology (for concrete nouns), causation (for verbs), auxiliary status (for verbs), evaluation (for nouns, adjectives and adverbs), derivational features (morpho-semantic features, class of the motivating word, taxonomic type of the motivating word, morphological type of derivation). The newest level of annotation, which is still being implemented, is the word-formation annotation.[26] This regards the annotation of prefixes, suffixes and roots, as well as the annotation of compounds. As regards compounds, however, the only information available at the moment is the "compound status" of the word, while no information regarding the compound type is currently provided, and not all compounds in the corpus have been tagged.

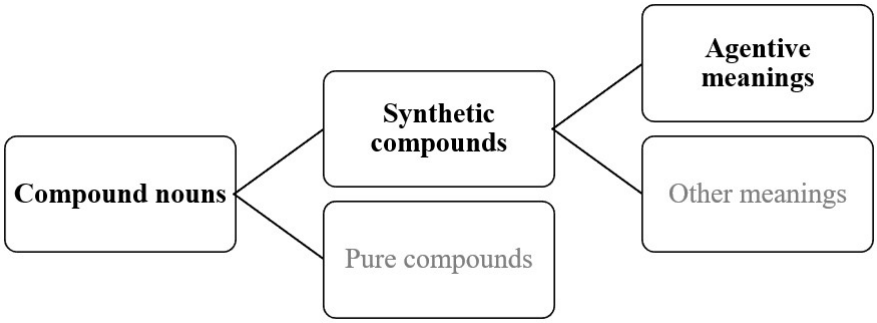
4.2. Extraction and selection of the data

The analysis of Russian compound agent nouns carried out throughout this study is based on the word-formation database of the RNC, which includes a list of compounds retrieved from dictionaries of modern Russian. This list has served as a basis for the first steps of the annotation of compounds in the corpus. Since the annotation process has not yet been completed, I could not use the RNC interface to collect the relevant data for the analysis proposed in the present chapter. Thus, I used the RNC word-formation database as a source for my investigations, while the RNC interface was used to retrieve examples and to check the number of occurrences for each compound in the database.

[26] See Grišina et al. (2009) and Tagabileva & Berezuckaja (2010) for more information about the word-formation annotation of the RNC.

From the RNC word-formation database (which contains over 16,000 compounds), I manually selected synthetic agentive compounds, following the steps shown in Figure 8. First, I selected compound nouns, leaving aside compound adjectives, adverbs and verbs. Second, I selected synthetic compounds, i.e. compounds deriving from the combination of compounding and suffixation/conversion (cf. Chapter 1, Sections 1.3, 1.6; Chapter 2, Section 2.3). Third, I selected compounds belonging to those constructions that bring about agentive meanings. In some cases, the same agentive constructions can also form noun-based compounds, as in *bel-o-gvarde-ec* 'member of the White Guard', from *Belaja Gvardija* 'White Guard'. These cases were excluded from the analysis.

Figure 8. Selection process from the RNC word-formation database



The figures corresponding to this selection process are summarized in Table 6, where the data are grouped into two main constructions.

Table 6. Absolute (AF) and relative (RF) frequencies of the synthetic suffixed and suffixless constructions in the RNC word-formation database

Construction	AF	RF (%)
[X-LV-V-SUFF]$_N$	438	52.7
[X-LV-V-Ø]$_N$	393	47.3
Total	**831**	**100**

The first construction, represented as [X-LV-V-SUFF]$_N$, includes compounds in which the first member is a non-verbal base and the second member is a verbal base (the two

bases being separated by a linking vowel), followed by one of the following agentive suffixes: *-ec*, *-lec*, *-tel'*, *-nik*, *-ščik/čik*, *-l'ščik*, *-ka*, *-lka*. Examples of each suffix are found in (1) through (8).

(1) *oruž-e-nos-ec* 'squire'
weapon-*e*-bear-*ec*

(2) *zeml-e-vlade-lec* 'land owner'
land-*e*-own-*lec*

(3) *zakon-o-da-tel'* 'legislator'
law-*o*-give-*tel'*

(4) *prav-o-zaščit-nik* 'human rights activist'
right-*o*-defend-*nik*

(5) *kvartir-o-s''em-ščik* 'tenant'
apartment-*o*-rent-*ščik*

(6) *stal-e-plavi-l'ščik* 'steel maker'
steel-*e*-melt-*l'ščik*

(7) *mjas-o-rub-ka* 'meat grinder'
meat-*o*-grind-*ka*

(8) *sen-o-kosi-lka* 'lawnmower'
hay-*o*-mow-*lka*

The figures corresponding to each suffix are shown in Table 7.

Table 7. Absolute (AF) and relative (RF) frequencies corresponding to each subschema of the suffixed construction in the RNC word-formation database

Construction	AF	RF (%)
[X-LV-V-*ec*]$_N$	143	32.6
[X-LV-V-*lec*]$_N$	7	1.6
[X-LV-V-*tel'*]$_N$	175	40
[X-LV-V-*nik*]$_N$	28	6.4
[X-LV-V-*ščik/čik*]$_N$	23	5.2

[X-LV-V-*l'ščik*]$_N$	4	0.9
[X-LV-V-*ka*]$_N$	47	10.7
[X-LV-V-*lka*]$_N$	11	2.5
Total	**438**	**100**

The second construction, represented as [X-LV-V-Ø]$_N$, includes compounds in which the first member is a non-verbal base and the second member is a verbal base (the two bases being separated by a linking vowel). These cases are regarded as instances of conversion, as already discussed (cf. Chapter 3, Section 3.7). One example of this construction is given in (9).

(9) *jazyk-o-ved-ø* 'linguist'
language-*o*-know-ø

I left out from my analysis compounds that, although formally belonging to one of the above-mentioned constructions, denote Objects/Patients (e.g. *novobranec* 'recruit'), Places (e.g. *lesopilka* 'sawmill'), and Events (e.g. *snegopad* 'snowfall'), which are rare compared to the other meanings (i.e. Prototypical Agents, Carriers of State, and Instruments) and are not found for all constructions. This choice was due to two main reasons. In the first place, considering that the main aim of the analysis consists in the comparison of rival word-formation constructions, the inclusion of nouns denoting Objects/Patients, Places and Events would have required a further comparison with other rival constructions that are used to form such nouns in Russian, which in turn are not used to form agent nouns. In the second place, the inclusion of these compounds would have weakened the semantic cohesion of the selected data. Indeed, the agentive suffixes shown above all share the property of producing so-called "subject names", i.e. nouns that act as subjects with respect to the base verb (e.g. *oruženosec* 'squire, one who bears weapons'), which makes them different from object/patient nouns, place nouns and event nouns.

4.3. A comparison of rival word-formation constructions

A cursory look at the compounds collected shows that the nine rival word-formation constructions giving rise to compound agent nouns in Russian do not show a clear-cut distribution. In other words, their functions seem to overlap and, in some cases, they appear to behave as completely synonymous constructions. So, why do we have *basn-o-pis-**ec*** 'fable writer', but *byt-o-pisa-**tel'*** 'everyday-life writer'? Why do we find *sladk-o-ež-**ka*** 'sweet tooth, lit. one who eats sweets', but *mjas-o-ed-**ø*** 'meat eater/lover'?

To determine what factors contribute most to the existence of these rival constructions in Russian, and prevent one construction from ousting all others, I collected all compounds formed with agent noun suffixes and analyzed them according to a number of parameters discussed in Sections 4.3.1 through 4.3.6.

4.3.1. Part of speech of the non-verbal element

The first parameter considered for the comparison of rival constructions is the word class of the compound's non-verbal element, which occurs in first position. The question addressed in this section is whether different constructions show different tendencies as regards the selection of the non-verbal element of the compound and whether the incorporation of a certain part of speech triggers the selection of a specific construction. The possible options are exemplified in (10) through (13).

(10) ***rabot***$_{N}$*-o-da-tel'* 'employer'
job$_{N}$*-o*-give-*tel'*

(11) ***vol'n***$_{A}$*-o-dum-ec* 'free thinker'
free$_{A}$*-o*-think-*ec*

(12) ***vs***$_{PRO}$*-e-derži-tel'* 'the almighty'
all$_{PRO}$*-e*-hold-*tel'*

(13) ***pjat***$_{NUM}$*-i-bor-ec* 'pentathlete'
five$_{NUM}$*-i*-fight-*ec*

The non-verbal element of the compound can be a nominal base (10), an adjectival/adverbial base (11), a pronominal base (12), or a numeral base (13).

As regards adjectival/adverbial bases, I intentionally chose not to specify the part-of-speech status of these elements, in agreement with Bogdanov (2011: 167). The author suggests that the status of first-position adjectival/adverbial bases in compounds is far from clear. Indeed, although these bases usually form adverbs, in most cases the compound cannot be paraphrased by using an adverb. The author exemplifies this situation by referring to expressions such as *svežezamorožennye ovošči* 'freshly frozen vegetables', that cannot be paraphrased as **ovošči, zamorožennye sveže/svežo* 'vegetables frozen freshly', but rather as *ovošči, zamorožennye v svežem sostojanii* 'vegetables frozen in a fresh state' (Bogdanov 2011: 167). According to the author, in these cases it would be wrong to say that we are dealing with adverbial bases. Regarding compound agent nouns, the options are the following:

a) the adjectival/adverbial base has a modal meaning, and thus can be paraphrased by using an adverb (14);
b) the adjectival/adverbial base can be paraphrased with what are known as "depictive constructions" (*depiktivnye konstrukcii*; cf. Rachilina 2010: 159 ff.) or "secondary predicates" (*vtoričnye predikaty*; cf. Bogdanov 2011: 190–192), as shown in (15), that is constructions of the type *on est mjaso*$_{\text{ACC}}$ *syrym*$_{\text{INS}}$ (Bogdanov 2011: 190) 'he eats the meat raw';
c) the adjectival/adverbial base is substantivized and thus functions as a noun (16).

(14) *gromk*$_{\text{A}}$*-o-govori-tel'* 'loudspeaker'
loud$_{\text{A}}$-*o*-speak-*tel'*

(15) *perv*$_{\text{A}}$*-o-prochod-ec* 'earliest explorer'
first$_{\text{A}}$-*o*-pass-*ec*

(16) *sladk*$_{\text{A}}$*-o-ež-ka* 'sweet tooth'
sweet$_{\text{A}}$-*o*-eat-*ka*

The compound in (14) could be paraphrased as 'an instrument that speaks loud' (*pribor, kotoryj govorit gromko*$_{\text{ADV}}$). The compound in (15) could be paraphrased as 'a person that

arrived first (to a place)' (*čelovek, kotoryj prošel pervym*$_{INS}$). Finally, the compound in (16) could be paraphrased as 'a person that eats sweets' (*čelovek, kotoryj est sladkoe*$_{N}$). In no case do we find adjectival bases used with attributive functions, as in compounds where the second base is nominal, e.g. *sin*$_{A}$*-e-glaz*$_{N}$*-yj* 'blue-eyed'. Thus, it seems reasonable to keep the adjectival/adverbial class together, keeping in mind that these bases can act as adverbial modifiers, secondary predicates, or nouns.

Let us now compare the behavior of the different constructions. Table 8 shows the absolute and relative frequencies of the four possible parts of speech in the different constructions.

Table 8. Absolute (AF) and relative (RF) frequencies corresponding to the non-verbal element's part of speech for rival constructions

	N		**A**		**PRO**		**NUM**	
	AF	**RF (%)**	**AF**	**RF (%)**	**AF**	**RF (%)**	**AF**	**RF (%)**
-ec[27]	117	81.8	15	10.5	7	4.9	4	2.8
-lec	7	100	0	-	0	-	0	-
-tel'	164	93.7	9	5.1	2	1.1	0	-
-nik	27	96.4	1	3.6	0	-	0	-
-čik/ščik	23	100	0	-	0	-	0	-
-l'ščik	4	100	0	-	0	-	0	-
-ka	37	78.7	4	8.5	5	10.6	1	2.1
-lka	11	100	0	-	0	-	0	-
-ø	333	84.7	42	10.7	17	4.3	1	0.3
Total	**723**	**87**	**71**	**8.5**	**31**	**3.7**	**6**	**0.7**

As can be observed, there is a strong tendency for all constructions to incorporate a nominal base as non-verbal element. Four constructions (*-lec*, *-ščik/čik*, ***-**l'ščik*, and *-lka*) only allow for the incorporation of nominal bases. The constructions that show a slightly higher variation (N lower than 90%) are those formed with *-ec*, *-ka* and the suffixless construction. Generally, however, it does not seem that the categorial status of the non-verbal element is a factor contributing to the diversity of rival constructions, and the

[27] For reasons of space, from now on I will only indicate the suffix, and not the whole construction.

incorporation of a particular part of speech does not seem to trigger the choice of a specific construction.

4.3.2. Semantic role of the non-verbal element

The second parameter considered for the comparison of rival constructions is the semantic role played by the non-verbal element with respect to the verbal base. Argument structure realization in verbal (or synthetic) compounds has been discussed in a number of studies within different theoretical frameworks (cf. Roeper & Siegel 1978; Selkirk 1982; Booij 1988; Kiefer 1993; Scalise 1994; Di Sciullo 2005), where it is generally pointed out that in these compounds the non-verbal element can be interpreted as the Theme/Patient argument of the underlying verb. According to Benigni & Masini (2009), in Russian compounds of the type N+V(+SUFF), "[…] N acts as the internal argument of the verb and has the semantic role of patient or theme, but never agent. Sometimes, though, N might have the role of instrument (*parochod*)" (Benigni & Masini 2009: 184). Tagabileva (2013: 198) adds to these two options the possibility of incorporating an element playing the semantic role of Manner. In fact, the analysis shows that many more options are possible.

Before showing the results of the analysis, I will briefly describe the inventory of the semantic roles found in the data. Considering that the literature on semantic roles and their classification is quite extended and varied (cf. Daniel 2014 for a summary of predicate-centered and argument-centered approaches), and especially that there is no consensus on the number and kind of necessary semantic roles (cf. Newmeyer 2010; Kittilä, Västi & Ylikoski 2011; Luraghi & Narrog 2014), in Table 9 I give my own description of the labels employed in the present analysis, which mostly follow the current terminology. The list of semantic roles summarized in Table 9 is not comprehensive of all the existing semantic roles, but only includes those that I have found in the data analyzed for this study.

Table 9. Inventory of semantic roles

Semantic role	Description	Example
Agent	entity that brings about the event voluntarily	**John** broke the window.
Theme/Patient	entity that undergoes a change of state or that is passively affected by the event	John broke **the window**.
Recipient	entity to whom the event is addressed	John gave **me** a present.
Location	place where the event takes place	John works **in London**.
Goal	place to which the event is directed	John is travelling **to Paris**.
Path	place through which the event takes place	John is walking **through the forest**.
Time	time at which the event takes place	John is going to Paris **tomorrow**.
Purpose	aim of the event	John fights **for civil rights**.
Instrument	entity used to bring about the event	John broke the window **with a stone**.
Manner	way in which the event is brought about	John is reading **carefully**.
Comitative	entity with whom the event is brought about	John is going out **with his friends**.
Topic	subject matter of the event	John is talking **about his family**.

In the analyzed compounds, I have found that the non-verbal element in first position can encode one of the semantic roles shown in Table 9. One instantiation for each semantic role is given in examples (17) through (27).

(17) Theme/Patient

posud$_{\text{TH/PT}}$-*o-moj-ka* 'dishwasher, lit. sth. that washes **the dishes**$_{\text{TH/PT}}$'

dish$_{\text{TH/PT}}$-*o*-wash-*ka*

(18) Recipient

bog$_{\text{REC}}$-*o-mol-ec* 'pilgrim, devotee, lit. one who prays **God**$_{\text{REC}}$'

god$_{\text{REC}}$-*o*-pray-*ec*

(19) Location

dom$_{\text{LOC}}$-*o-sed* 'homebody, lit. one who sits **at home**$_{\text{LOC}}$'

home$_{\text{LOC}}$-*o*-sit

(20) Goal

ogn$_{\text{GOAL}}$-*e-poklon-nik* 'fire worshipper, lit. one that bows **toward the fire**$_{\text{GOAL}}$'

fire$_{\text{GOAL}}$-*e*-bow-*nik*

(21) Path

mor$_{\text{PATH}}$-*e-chod-ec* 'seaman, lit. one who goes **across the sea**$_{\text{PATH}}$'

sea$_{\text{PATH}}$-*e*-go-*ec*

(22) Time

dolg$_{\text{TIME}}$-*o-ži-tel'* 'long liver, lit. one who lives **for a long time**$_{\text{TIME}}$'

long$_{\text{TIME}}$-*o*-live-*tel'*

(23) Purpose

pravd$_{\text{PUR}}$-*o-bor-ec* 'truth fighter, lit. one who fights **for the truth**$_{\text{PUR}}$'

truth$_{\text{PUR}}$-*o*-fight-*ec*

(24) Instrument

molot$_{\text{INSTR}}$-*o-bo-ec* 'hammerer, lit. one who beats **with a hammer**$_{\text{INSTR}}$'

hammer$_{\text{INSTR}}$-*o*-beat-*ec*

(25) Manner

gromk$_{\text{MNR}}$-*o-govori-tel'* 'loudspeaker, lit. sth. that speaks **loud**$_{\text{MNR}}$'

loud$_{\text{MNR}}$-*o*-speak-*tel'*

(26) Comitative

skot$_{\text{COM}}$-*o-lož-ec* 'zoophile, lit. one who lies **with the cattle**$_{\text{COM}}$'

cattle$_{\text{COM}}$-*o*-lie-*ec*

(27) Topic

byt$_{\text{TOP}}$-*o-pisa-tel'* 'life writer, lit. one who writes **about life**$_{\text{TOP}}$'

life$_{\text{TOP}}$-*o*-write-*tel'*

The incorporation of elements other than Theme/Patient (which, as we will see, is by far the most frequent case) can be due to several reasons. One important factor concerns the verbal valency. If we take example (21), for instance, we can see that the verbal base corresponds to the verb *chodit'* 'go, walk', which is monovalent and takes no direct object. In such cases, then, the incorporation of an element playing the semantic role of

Theme/Patient is impossible. However, we do find cases in which the incorporation of elements other than Theme/Patient is due to factors other than verbal valency. Let us consider the compound *očevidec* 'eye witness, lit. one who sees with (his own) eyes' (from *oko* 'eye' and *videt'* 'see'). In this case, the incorporation of the direct object would have been possible in principle, because *videt'* 'see' is a bivalent verb. However, what is relevant here is the Instrument (the eyes): a person can be qualified as an eye witness if he/she sees a crime with his/her own eyes. Therefore, the type of argument also plays a crucial role in the incorporation of certain elements rather than others (cf. Kiefer 1993).

As stressed by Benigni & Masini (2009: 184) and Tagabileva (2013: 198), the non-verbal element cannot denote the subject of the verbal base, as these formations as a whole already denote subjects with respect to the underlying verbs. My analysis confirms this assumption and shows that no first arguments are incorporated in such compounds.

Let us now compare the different constructions based on the semantic role played in the compound by the non-verbal element. The results are summarized in Table 10.

Table 10. Relative frequencies (RF)[28] corresponding the non-verbal element's semantic role for rival constructions

	Th/Pt	**Rec**	**Loc**	**Goal**	**Path**	**Time**	**Pur**	**Instr**	**Mnr**	**Com**	**Top**
-ec	75.5	0.7	2.1	-	2.1	0.7	1.4	3.5	12.6	1.4	-
-lec	100	-	-	-	-	-	-	-	-	-	-
-tel'	90.3	-	1.1	0.6	1.1	0.6	-	0.6	5.1	-	0.6
-nik	82.1	-	-	17.9	-	-	-	-	-	-	-
-ščik/čik	78.3	-	-	-	8.7	-	-	13	-	-	-
-l'ščik	100	-	-	-	-	-	-	-	-	-	-
-ka	83	-	-	-	-	-	-	2.1	14.9	-	-
-lka	100	-	-	-	-	-	-	-	-	-	-
-ø	81.4	0.3	0.5	1.3	2.5	0.8	0.3	2.5	10.4	-	-
Total	**82.8**	**0.2**	**0.8**	**1.3**	**2**	**0.6**	**0.4**	**2.4**	**9**	**0.2**	**0.1**

As expected, there is a strong preference for all constructions to incorporate a non-verbal element playing the semantic role of Theme/Patient. However, certain constructions (cf.

[28] Absolute frequencies are not shown for reasons of space.

particularly *-ec*, *-tel'* and the suffixless construction) allow for more variation and can incorporate elements playing many different semantic roles, many more than those identified in previous studies about compounding in Russian.

4.3.3. Transitivity of the verbal element

The third parameter regards the transitivity of the verbal element. Compound agent nouns can show both transitive (28) and intransitive (29) verbal bases.

(28) *mašin-o-**stroi***$_{TR}$*-tel'* 'machinist, lit. one who **builds**$_{TR}$ cars'
car-*o*-**build**$_{TR}$-*tel'*

(29) *kanat-o-**chod***$_{INTR}$*-ec* 'rope walker, lit. one who **walks**$_{INTR}$ on a rope'
rope-*o*-**walk**$_{INTR}$-*ec*

The figures corresponding to each construction are shown in Table 11.

Table 11. Absolute (AF) and relative (RF) frequencies corresponding to transitive and intransitive verbal bases for rival constructions

	Transitive		**Intransitive**	
	AF	**RF (%)**	**AF**	**RF (%)**
-ec	123	86	20	14
-lec	0	-	7	100
-tel'	160	91.4	15	8.6
-nik	21	75	7	25
-ščik/čik	21	91.3	2	8.7
-l'ščik	4	100	0	-
-ka	46	97.9	1	2.1
-lka	11	100	0	-
-ø	340	86.5	53	13.5
Total	**726**	**87.4**	**105**	**12.6**

As shown in Table 11, although some constructions (i.e. *-ec*, *-nik* and the suffixless construction) allow for greater variation, all constructions show a strong preference for

transitive verbal bases. By contrast, the construction with the suffix *-lec* only includes intransitive verbal bases. Remarkably, however, this construction includes compounds formed on the basis of a single verb, i.e. *vladet'* 'own, possess', which is, indeed, intransitive and governs the instrumental case.

4.3.4. Formal aspect of the verbal element

In the present section, I compare the verbal bases embedded in compounds based on their (formal) belonging to a certain grammatical aspect. The question addressed in this section is whether rival constructions show preferences or restrictions regarding the incorporation of verbal bases whose aspect is formally perfective or imperfective. I refer to the "formal" aspect of the verbal bases because, from a semantic point of view, all the events underlying the compounds considered must be understood as imperfective. Indeed, a fable writer (*basnopisec*) is one who writes fables habitually, as a professional; a fire extinguisher (*ognetušitel'*) is an instrument that is habitually employed to extinguish fire, and so on. However, from the formal point of view, several situations are possible, and we might wonder whether this could be a significant factor when it comes to distinguishing rival constructions. In most cases, the verbal base is either clearly imperfective (30) or its status is unclear (31).

(30) *skor-o-**pis***$_{\text{IPFV}}$*-ec* 'shorthand typist, lit. one who **writes**$_{\text{IPFV}}$ fast'
fast-*o*-**write**$_{\text{IPFV}}$-*ec*

(31) *ver-o-**otstup**-nik* 'apostate, lit. one who **abandons** the faith'
faith-*o*-**abandon**-*nik*

In (30) the verbal base corresponds to the verb *pisat'* 'write', which is imperfective (the corresponding perfective is *napisat'*, containing the perfectivizing prefix *na-*). In (31) we cannot determine whether the verbal base is perfective or imperfective, as the only difference between the two members of the aspectual pair is the suffix *-a*/*i-*, which does not appear in the compound: *otstupat'sja* ('abandon', imperfective) and *otstupit'sja* ('abandon', perfective) (cf. also Itkin 1996: 19–20). In a few cases, the verbal base can

be classified as biaspectual, as in (32), where the base verb is *nasledovat'* 'inherit' (biaspectual).

(32) *prestol-o-**nasled***$_{\text{BIASP}}$*-nik* 'successor, lit. one who **inherits**$_{\text{BIASP}}$ the throne'
throne-*o*-**inherit**$_{\text{BIASP}}$-*nik*

In some other cases, the verbal base is clearly perfective (33).

(33) *zeml-e-**ustroi***$_{\text{PFV}}$*-tel'* 'land organizer, lit. one who **organizes**$_{\text{PFV}}$ the land'
land-*e*-**organize**$_{\text{PFV}}$-*tel'*

However, in all cases in which a perfective base appears, in the aspectual pair the imperfective is secondary. So, the imperfective corresponding to the perfective *ustroit'* 'organize' is *ustraivat'*, which is formed through the addition of the imperfective suffix -*iva*-. Thus, it is reasonable to believe that in these cases the perfective base is selected because it is morphologically more basic than its imperfective counterpart. Interestingly, this mostly regards compounds in *tel'*, which show a perfective base in 23.4% of the cases (see Table 12).

Table 12. Absolute (AF) and relative (RF) frequencies corresponding to the formal aspect of the verbal bases for rival constructions

	Imperfective		**Biaspectual**		**Uncertain status**		**Perfective**	
	AF	**RF (%)**	**AF**	**RF (%)**	**AF**	**RF (%)**	**AF**	**RF (%)**
-ec	143	100	0	-	0	-	0	-
-lec	7	100	0	-	0	-	0	-
-tel'	129	73.7	5	2.9	0	-	41	23.4
-nik	7	25	1	3.6	18	64.3	2	7.1
-ščik/čik	14	60.7	0	-	8	34.8	1	4.3
-l'ščik	4	100	0	-	0	-	0	-
-ka	45	95.7	0	-	1	2.1	1	2.1
-lka	11	100	0	-	0	-	0	-
-ø	377	95.9	1	0.3	14	3.6	1	0.3
Total	**737**	**88.7**	**7**	**0.8**	**41**	**4.9**	**46**	**5.5**

The high number of perfective bases in compounds belonging to the construction in *-tel'* could be seen as partially correlated to its semantics. As will be discussed further on, the construction in *-tel'* is frequently associated with instrumental meanings, much more frequently than the other high-frequency constructions. These instrumental meanings are often brought about by compounds based on prefixed verbs whose semantics is more specific than the semantics of the base unprefixed imperfective. For instance, the compound *paroočistitel'* 'fallow fields purifier' is based on the prefixed verb *očistit'*, whose meaning 'purify' is more specific than the meaning of the unprefixed *čistit'* 'clean'. Considering that *očistit'* develops the secondary imperfective *očiščat'* that conserves the specific meaning 'purify', the perfective remains more basic in the aspectual pair and is thus employed in compounding.

In general, most constructions show a clear preference for imperfective bases (*-ec*, *-lec*, *-l'ščik* and *-lka* do not even allow for other options) or for bases whose status is uncertain (cf. *-nik*). Much less frequently do we observe perfective bases in compounds, and this almost exclusively regards the suffix *-tel'*, as I have just discussed.

4.3.5. Animacy of the compound's referent

In the present section, I compare rival constructions based on the type of referents they denote. In particular, I consider the parameter of animacy and the three possible outcomes of the animacy hierarchy: human > animal > inanimate (cf. Croft 1989: 185). The compounds included in my database show that all these outcomes are possible, as they can denote human beings (34), animals (35), and inanimate entities (36).

(34) *oč-e-vid-ec* 'eye witness, lit. one who sees with his own eyes'
eye-*e*-see-*ec*

(35) *bron-e-nos-ec* 'armadillo, lit. one who has an armour'[29]
armour-*e*-carry-*ec*

[29] The compound *bronenosec* also has the meaning of 'battleship'.

(36) *tepl-o-nosi-tel'* 'heat carrier, lit. sth. that carries heat'
heat-*o*-carry-*tel'*

Do rival constructions behave differently as regards the animacy of the referents they denote? Is there a tendency of certain constructions to form exclusively a specific type of referents, or more referents belonging to a specific type with respect to other constructions?

Table 13 shows that animacy appears as a relevant factor to highlight the differences among rival constructions.

Table 13. Absolute (AF) and relative (RF) frequencies corresponding to the animacy of the compound's referent for rival constructions

	Human		**Animal**		**Inanimate**	
	AF	**RF (%)**	**AF**	**RF (%)**	**AF**	**RF (%)**
-ec	131	91.6	4	2.8	8	5.6
-lec	7	100	0	-	0	-
-tel'	122	69.7	0	-	53	30.3
-nik	16	57.1	0	-	12	42.9
-ščik/čik	15	65.2	0	-	8	34.8
-l'ščik	4	100	0	-	0	-
-ka	13	27.7	6	12.8	28	59.6
-lka	0	-	0	-	11	100
-ø	247	62.8	35	8.9	84	21.4
Total	**555**	**66.8**	**45**	**5.4**	**204**	**24.5**

The constructions in *-ec*, *-lec* and *-l'ščik* mainly create nouns denoting human beings, while *-ka* and *-lka* mostly denote inanimate entities. The constructions in *-tel'*, *-nik*, *-ščik/čik*, and the suffixless construction seem to allow for greater variation: they mostly denote human beings, but they are also frequently employed to denote inanimate entities. The parameter of animacy is strictly connected to the next parameter, i.e. the compound's semantics.

4.3.6. Semantics of the compound

As discussed in Chapter 3, the morphological formants used to create agent nouns often display polysemy patterns that are common cross-linguistically. In Chapter 3 (cf. Sections 3.6 and 3.7), I have shown that Russian deverbal agentive suffixes show a polysemy pattern that includes, besides Prototypical Agents, Carriers of State, Instruments, Objects/Patients, Places, and Events. Having excluded from the analysis nouns denoting Objects/Patients, Places and Events for the reasons discussed in Section 4.2, we are left with nouns denoting Prototypical Agents, Carriers of State and Instruments. Table 14 summarizes the properties that I attribute to each of these labels.

Table 14. Semantics of the compound

Semantics	**Description**
Prototypical Agent	entity that performs a certain action
Carrier of State	entity that is characterized by the fact of being in a certain state
Instrument	entity that is used to perform a certain action

In examples (37) through (39), I show one example for each semantic label.

(37) *granat-o-met-čik* 'grenadier, lit. one who throws grenades'
grenade-*o*-throw-*čik*

(38) *knig-o-ljub* 'bibliophile, lit. one who loves books'
book-*o*-love

(39) *gazon-o-kosi-lka* 'lawnmower, lit. sth. that is used to mow the lawn'
land-*o*-mow-*lka*

Example (37) displays a prototypical agent noun: a grenadier is one who performs (as a profession) the action of throwing grenades. The compound in (38), instead, is classified as a Carrier of State: a bibliophile is a person who has the characteristic of loving books. In example (39), an instrument noun is shown: a lawnmower is a machine employed to perform the action of mowing the land. Table 15 shows the behavior of rival constructions based on the semantics of the nouns they form.

Table 15. Absolute (AF) and relative (RF) frequencies corresponding to the type of nouns denoted by rival constructions

	Prototypical Agents		**Carriers of State**		**Instruments**	
	AF	**RF (%)**	**AF**	**RF (%)**	**AF**	**RF (%)**
-ec	85	59.4	51	35.7	7	4.9
-lec	0	-	7	100	0	-
-tel'	103	58.9	21	12	51	29.1
-nik	14	50	2	7.1	12	42.9
-ščik/čik	16	69.6	0	-	7	30.4
-l'ščik	4	100	0	-	0	-
-ka	19	40.4	3	6.4	25	53.2
-lka	0	-	0	-	11	100
-ø	246	62.6	79	20.1	68	17.3
Total	**487**	**58.6**	**163**	**19.6**	**181**	**21.8**

As shown in Table 15 (and as already discussed in Chapter 3, Section 3.6), not all constructions are able to form all types of nouns. Compounds in *-lec* can only denote Carriers of State, compounds in *-l'ščik*, by contrast, can only denote Prototypical Agents, while compounds in *-lka* can only denote Instruments. Among the most frequent constructions, we can observe greater variation: *-ec* compounds mainly denote Prototypical Agents (59.4%), but they also frequently denote Carriers of State (35.7%), while only rarely do they denote Instruments (4.9%); the construction in *-tel'* seems to be more strictly related to actions, as it mostly creates agent nouns (58.9%) and instrument nouns (29.1%), while it forms nouns denoting Carriers of State less frequently (12%); the same is true for compounds in *-nik*, *-ščik/čik* and *-ka*: these suffixes are rarely (or never, in the case of *-ščik/čik*) employed in compounding to create nouns denoting Carriers of State, while they are more frequently used to form agent or instrument nouns; the suffixless construction, which is by far the most frequent, mainly denotes Prototypical Agents (62.6%), but it also shows significant percentages of nouns denoting Carriers of State (20.1%) and Instruments (17.3%).

4.3.7. To sum up

The results of the analysis described throughout Section 4.3 are summarized in Table 16. In this table, each column shows the results (in the form of relative frequencies) for each of the parameters considered in Sections 4.3.1 through 4.3.6. For reasons of space, the column including the parameter "Part of speech" only shows the relative frequencies of the outcome "N" (noun) compared to all the other possible outcomes, i.e. adjective/adverb, pronoun and numeral. Similarly, the column corresponding to the parameter "Semantic role" compares the relative frequencies of the most common outcome, i.e. "Th/Pt" (theme/patient), with all other possible semantic roles. As regards the parameter "Aspect", the cases of perfective verbal bases are compared to the cases of imperfective or other (biaspectual or unclear) verbal bases. Finally, the column corresponding to the parameter "Animacy" compares the relative frequencies of animate referents (including both human and animal) with inanimate referents.

Remarkably, there are striking differences between the low-frequency constructions formed with the suffixes *-lec*, *-l'ščik*, *-lka* and all the other constructions. Indeed, the constructions in *-lec*, *-l'ščik* and *-lka* show no variation at all for each of the parameters considered. In other words, the compounds formed with these suffixes are instantiations of very specific constructions, in which there is no room for variation, neither regarding the features of the two lexical bases forming the compound, nor in terms of the type of referents that the compounds denote. So, for instance, the construction in *-lec* only gives rise to compounds in which the first element is a noun playing the semantic role of Theme/Patient, and the second element is constituted by an intransitive, imperfective verbal base (specifically, that of the verb *vladet'* 'own'). All compounds belonging to this construction denote human entities with the meaning of Carrier of State.

By contrast, the other constructions show different behaviors, allowing for a certain degree of variation, higher or lower depending on the parameter considered. The parameters "Part of speech", "Semantic role", and "Transitivity" show similar tendencies for all the constructions. The construction in *-ec* and the suffixless construction show on average more variation. The parameter "Aspect" highlights the different behavior of the suffix *-tel'* with respect to the other constructions, as it is the only suffix that allows a high percentage of perfective verbal bases (23.4% of the cases). The parameters

"Animacy" and "Semantics" show that the constructions in -*ec* and -*ø* are more strictly associated with meanings of Human Agents, while the other constructions are more often employed to create nouns denoting Instruments.

Given the large number of data and parameters considered, in Section 4.4, statistical analyses are carried out to shed more light on the results obtained so far.

Table 16. Relative frequencies of rival constructions according to the six parameters considered

	Part of speech		Semantic role		Transitivity		Aspect		Animacy		Semantics		
	N	Other	Th/Pt	Other	Tr	Intr	Ipfv or other	Pfv	Hum/An	Inan	Agent	CoS	Instr
-ec	81.8	18.2	75.5	24.5	86	14	100	-	94.4	5.6	59.4	35.7	4.9
-lec	100	-	100	-	-	100	100	-	100	-	-	100	-
-tel'	93.7	6.3	90.3	9.7	91.4	8.6	76.6	23.4	69.7	30.3	58.9	12	29.1
-nik	96.4	3.6	82.1	17.9	75	25	92.9	7.1	57.1	42.9	50	7.1	42.9
-ščik/čik	100	-	78.3	21.7	91.3	8.7	95.7	4.3	65.2	34.8	69.6	-	30.4
-l'ščik	100	-	100	-	100	-	100	-	100	-	100	-	-
-ka	78.7	21.3	83	17	97.9	2.1	97.9	2.1	40.4	59.6	40.4	6.4	53.2
-lka	100	-	100	-	100	-	100	-	-	100	-	-	100
-ø	84.7	15.3	81.4	18.6	86.5	13.5	99.7	0.3	78.6	21.4	62.6	20.1	17.3
Total	87	13	82.8	17.2	87.4	12.6	94.5	5.5	75.5	24.5	58.6	19.6	21.8

4.4. Statistical analyses of rival word-formation constructions

In the present section, I resort to statistical methods that can help discriminate the behavior of rival word-formation constructions giving rise to compound agent nouns in Russian, based on all the parameters discussed in Section 4.3. In Section 4.4.1, I describe the data pre-processing that was carried out in order to facilitate statistical analyses. Section 4.4.2 focuses on logistic regression models, whereas Section 4.4.3 deals with conditional inference trees and random forests.

4.4.1. Data pre-processing

Before performing statistical analyses, the data and the parameters described in Section 4.3 were reconsidered and reorganized to facilitate the computation of results.

Concerning the parameters used to classify compounds, a few changes were made before starting the analysis:

a) the possible outcomes of the parameter "Part of speech" were reduced to the two options "N" and "other";
b) the possible outcomes of the parameter "Semantic role" were reduced to the two options "TH/PT" and "other";
c) the possible outcomes of the parameter "Aspect" were reduced to the two options "PFV" and "other";
d) the possible outcomes of the parameter "Animacy" were reduced to the two options "AN" (including humans and animals[30]) and "INAN".

Regarding the constructions, a few changes were made according to the type of statistical analysis performed:

a) in the case of binomial logistic regression (cf. Section 4.4.2), the comparison regards the constructions showing an expressed suffix on the one side, and the

[30] Animals, though, represent a small percentage (5.4%) of the total with respect to nouns denoting humans and inanimates (cf. Section 4.3.5).

suffixless construction on the other, that is all the constructions with an expressed suffix were merged and compared to the suffixless construction;

b) in the case of multinomial logistic regression, conditional inference trees and random forests, the low-frequency constructions (i.e. *-lec*, *-l'ščik*, and *-lka*) were merged with the simple suffixes from which they originate as a result of amalgamation or metanalysis, as shown in Sections 4.4.2 and 4.4.3.

4.4.2. Logistic regression models

Logistic regression is a statistical method used in linguistics to investigate the distribution of rival words or constructions based on different parameters or contextual features (cf. Baayen 2008, Chapter 6; Speelman 2014; Levshina 2015, Chapters 12 and 13). This technique "models the relationships between a categorical response variable with two or more possible values and one or more explanatory variables, or predictors. (…) If there are two possible outcomes (i.e. near synonyms), the logistic model is called binomial, or dichotomous. In case of three and more outcomes, we deal with a multinomial, or polytomous regression" (Levshina 2015: 253). Logistic regression techniques have been widely employed in linguistics to compare grammatical constructions (cf., among others, Bresnan et al. 2007; Baayen et al. 2013; Fabiszak et al. 2014; Klavan 2014; Levshina, Geeraerts & Speelman 2014; Shank, Plevoets & Cuyckens 2014; Kang 2016), and investigate questions of lexical semantics (cf., among others, Arppe 2008; Deshors & Gries 2014; Glynn 2014; Robinson 2014). In the present section, I resort to logistic regression techniques to compare the rival constructions that constitute the object of this study.

4.4.2.1. Binomial logistic regression

As a first step, to compare the suffixed constructions to the suffixless construction, I apply binomial logistic regression (cf. Gries 2013, Chapter 5; Levshina 2015, Chapter 12) using R[31] version 3.3.2, and specifically, the package "rms" (Regression Modeling Strategies).

[31] R is a free software environment for statistical computing and graphics, see https://www.r-project.org/.

The input document is a tab-delimited file in which each row corresponds to a different lemma, while each column corresponds to a different parameter (part of speech of the non-verbal base, semantic role of the non-verbal base, transitivity of the verbal base, aspect of the verbal base, animacy of the referent, and semantics of the compound). The last column indicates the number of occurrences of each lemma in the RNC.[32] To perform an analysis based on token numbers, I multiply each row by the number of tokens indicated in the last column and I get a new file in which each row corresponds to an occurrence in the RNC.

By applying the function lrm(), which produces a logistic regression model, I get the results shown in Figure 9.

Figure 9. Binomial logistic regression model: suffixed vs. suffixless

```
Logistic Regression Model

 lrm(formula = suffix ~ semantics + animacy + aspect + transitivity +
     pos + sr, data = y_n)

                       Model Likelihood      Discrimination    Rank Discrim.
                          Ratio Test            Indexes           Indexes
 Obs          186186    LR chi2    73945.91    R2        0.440    C       0.811
  n           106741    d.f.               7    g         2.013    Dxy     0.621
  y            79445    Pr(> chi2)  <0.0001    gr        7.482    gamma   0.687
 max |deriv|  2e-06                           gp        0.326    tau-a   0.304
                                              Brier     0.167

                  Coef    S.E.   Wald Z Pr(>|Z|)
 Intercept         0.7034 0.0084  83.88 <0.0001
 semantics=cos     0.7290 0.0171  42.68 <0.0001
 semantics=instr   0.1095 0.0415   2.64 0.0084
 animacy=inan     -1.2277 0.0393 -31.21 <0.0001
 aspect=pfv        5.0797 0.1967  25.83 <0.0001
 transitivity=intr -1.1564 0.0172 -67.11 <0.0001
 pos=other        -0.3606 0.0171 -21.09 <0.0001
 sr=other         -0.8258 0.0182 -45.27 <0.0001
```

After the formula, the output shows some statistics. On the left, it shows the total number of observations and the frequency of each outcome ("n" stands for "no", that is "suffixless", whereas "y" stands for "yes", that is "suffixed"). The model seems

[32] The occurrences of each compound have been checked in the RNC main corpus, which contains texts from the 18th century to the present day belonging to different genres (fiction, drama, memoirs and biographies, journalism and literary criticism, scientific and popular scientific texts, instructional texts, religious and philosophical texts, technical texts, business and jurisprudence texts, letters and diaries), for a total of over 200 million words.

significant, since the *p*-value[33] is lower than 0.05, and the concordance index C (also known as the area under the ROC-curve, i.e. the receiver operating characteristic curve) indicates that the model discriminates well, as its value is higher than 0.80 (cf. Levshina 2015: 259). Concerning the coefficients, we see that the first row corresponds to the intercept, that is, the estimated log odds (i.e. the logarithm of the odds) of the outcome when all predictors are at their reference levels. The reference levels are automatically selected by the program according to the alphabetical order of the possible outcomes, if not else specified by the user. In this case, the reference levels were manually set to the following outcomes: part of speech = noun, semantic role = theme/patient, transitivity = transitive, aspect = other (i.e. non-perfective), animacy = animate, semantics = agent, which constitute the most common combination of factors.

To interpret the results, we have to keep in mind the following: a negative coefficient boosts the chances of the reference level (in this case, the suffixless construction), while a positive coefficient boosts the chances of the second level (that is, the suffixed constructions).

Thus, when all predictors are at their reference levels, the chances of the constructions with an expressed suffix are greater than the chances of the suffixless construction. By contrast, if we look at the results below the intercept, we see that the chances of the suffixless construction to occur are greater than the chances of the constructions with an expressed suffix when the referent is inanimate, when the verb base is intransitive, when the first element of the compound is not a noun, and when the semantic role played by the first element is not that of Theme/Patient. Conversely, the chances are greater for the constructions with an expressed suffix when the compounds denote Carriers of State or Instruments and when the verbal base is perfective.

In the last column on the right, we find the *p*-values, which indicate whether the null hypothesis of no difference can be rejected or not. In our case, the *p*-values are lower than 0.05, so the null hypothesis of no difference can be rejected.

By applying the function validate(), we verify whether the model is subject to overfitting. Overfitting usually occurs when a model has too many parameters compared to the number of observations, and this results in poor predictive performances (cf.

[33] The *p*-value or probability value is the probability to get the data we have (i.e. the actual observed results) under the assumption that the null hypothesis is true.

Levshina 2015: 166–167, 274–275). By applying this function, we perform a resampling validation of the regression model, i.e. the model is validated many times (200 in our case) by using random subsets of the data.

Figure 10. Validation of the logistic regression model

```
                Backwards Step-down - Original Model

No Factors Deleted

Factors in Final Model

[1] semantics    animacy      aspect       transitivity pos
[6] sr
          index.orig training    test optimism index.corrected   n
Dxy           0.6214   0.6214  0.6213    0e+00          0.6214 200
R2            0.4402   0.4403  0.4402    1e-04          0.4401 200
Intercept     0.0000   0.0000 -0.0004    4e-04         -0.0004 200
Slope         1.0000   1.0000  0.9999    1e-04          0.9999 200
Emax          0.0000   0.0000  0.0001    1e-04          0.0001 200
D             0.3972   0.3972  0.3971    1e-04          0.3971 200
U             0.0000   0.0000  0.0000    0e+00          0.0000 200
Q             0.3972   0.3972  0.3971    1e-04          0.3971 200
B             0.1666   0.1666  0.1666    0e+00          0.1666 200
g             2.0125   2.0154  2.0152    2e-04          2.0124 200
gp            0.3255   0.3255  0.3255    0e+00          0.3255 200
```

As can be seen in Figure 10, this function produces in output some goodness-of-fit measures. The most important information about potential overfitting is shown in the column “optimism”, which displays the differences between the training and the test statistics: high optimism values indicate overfitting. Considering that in our case the optimism values are lower than 0.05, we can conclude that the model does not overfit the data.

To see whether all predictors are worth keeping in the model, i.e. whether they significantly discriminate between the two constructions or not, I resort to the glm() function, which creates generalized linear models, often called mixed models because they contain both fixed effects and random effects (cf. Levshina 2015: 192). First, I apply this function keeping all the predictors included in the binomial logistic regression model and I obtain the results shown in Figure 11.

Figure 11. Generalized linear model: suffixed vs. suffixless (all predictors included)

```
Call:
glm(formula = suffix ~ semantics + animacy + aspect + transitivity +
    pos + sr, family = "binomial", data = y_n)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.0575  -0.6061  -0.3507   0.8967   2.4176

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)       0.703366   0.008385  83.883  < 2e-16 ***
semanticscos      0.728993   0.017081  42.679  < 2e-16 ***
semanticsinstr    0.109471   0.041526   2.636  0.00838 **
animacyinan      -1.227660   0.039333 -31.212  < 2e-16 ***
aspectpfv         5.079692   0.196663  25.829  < 2e-16 ***
transitivityintr -1.156364   0.017232 -67.105  < 2e-16 ***
posother         -0.360637   0.017098 -21.092  < 2e-16 ***
srother          -0.825819   0.018243 -45.269  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 254092  on 186185  degrees of freedom
Residual deviance: 180146  on 186178  degrees of freedom
AIC: 180162

Number of Fisher Scoring iterations: 8
```

Notably, the model produces the same results as those reached through the lrm() function. However, the advantage of using this function is that it allows comparing different models based on the number of predictors included. In other words, we can build different glm models and eliminate one predictor at a time to see whether it significantly discriminates between the two constructions or not. Thus, we can eliminate the last predictor included in the formula in Figure 11, i.e. “sr” (first member’s semantic role). This yields the results shown in Figure 12. A first way to compare the two models is to look at the AIC values. The AIC (Akaike information criterion) is a measure that estimates the quality of the model and has to be compared with the AIC of similar models including more or fewer predictors. The lower the AIC value, the better the model. In this case, we see that the first model including all predictors seems to be better (AIC: 180,162) than the second model, which does not include the predictor “sr” (AIC: 182,226). The two models can also be compared by applying an ANOVA, i.e. analysis of variance (cf. Levshina 2015: 171 ff.), to see whether the model that includes the predictor “sr” tells us more than the model that does not include this predictor.

Figure 12. Generalized linear model: suffixed vs. suffixless ("sr" eliminated)

```
Call:
glm(formula = suffix ~ semantics + animacy + aspect + transitivity +
    pos, family = "binomial", data = y_n)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.9887  -0.6641  -0.3195   0.9142   2.4138

Coefficients:
                 Estimate Std. Error  z value Pr(>|z|)
(Intercept)       0.656383   0.008261   79.460   <2e-16 ***
semanticscos      0.742577   0.017203   43.165   <2e-16 ***
semanticsinstr   -0.092300   0.040983   -2.252   0.0243 *
animacyinan      -1.229415   0.039104  -31.440   <2e-16 ***
aspectpfv         5.120007   0.196685   26.032   <2e-16 ***
transitivityintr -1.550187   0.014633 -105.941   <2e-16 ***
posother         -0.734217   0.014668  -50.056   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 254092  on 186185  degrees of freedom
Residual deviance: 182212  on 186179  degrees of freedom
AIC: 182226

Number of Fisher Scoring iterations: 8
```

Figure 13. ANOVA: model with all predictors vs. model without the predictor "sr"

```
Analysis of Deviance Table

Model 1: suffix ~ semantics + animacy + aspect + transitivity + pos
Model 2: suffix ~ semantics + animacy + aspect + transitivity + pos +
    sr
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1    186179     182212
2    186178     180146  1   2065.1 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 13 shows the results of this comparison. If the greater model significantly reduces the deviance with respect to the model without a predictor, this means that the predictor is worth keeping in the model. In this case, the model including all predictors significantly reduces the deviance, so we can conclude that it is worth keeping the predictor "sr" in the model.

The same operation can be repeated and, by eliminating one predictor at a time, we can see whether their role in the model is significant or not. By applying the function drop(), this operation is repeated automatically for each predictor.

Figure 14. Single term deletions

```
Single term deletions

Model:
suffix ~ semantics + animacy + aspect + transitivity + pos +
    sr
             Df Deviance    AIC    LRT  Pr(>Chi)
<none>            180146 180162
semantics     2   182118 182130 1971.7 < 2.2e-16 ***
animacy       1   181117 181131  970.6 < 2.2e-16 ***
aspect        1   186004 186018 5857.8 < 2.2e-16 ***
transitivity  1   184623 184637 4476.5 < 2.2e-16 ***
pos           1   180588 180602  441.3 < 2.2e-16 ***
sr            1   182212 182226 2065.1 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 14 shows that the lower AIC corresponds to the model including all predictors (180,162) and that the deviance changes significantly every time one of the predictors is eliminated (the last column shows that the *p*-values are very small), which indicates that each predictor is useful for the model. However, these results should be interpreted cautiously in light of the possible effect size problems due to the large quantity of data examined.

4.4.2.2. Multinomial logistic regression

When we want to compare more than two rival words or constructions, we have to resort to multinomial logistic regression (cf. Gries 2013, Chapter 5; Levshina 2015, Chapter 13), or, as will be shown in Section 4.4.3, to alternative methods such as conditional inference trees and random forests.

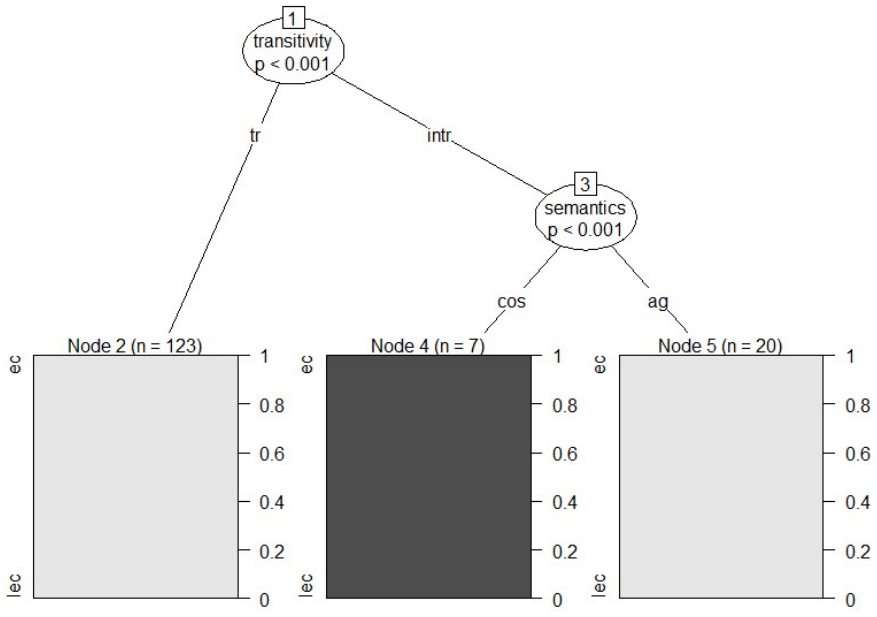
Before performing multinomial logistic regression, a few changes were made concerning the number of possible outcomes of the analysis. These changes are related to the low-frequency constructions in *-lec*, *-l'ščik*, and *-lka*. These constructions are the result of amalgamation or metanalysis (see Chapter 3), and, therefore, are strictly related to the simple suffixes from which they originate: *-lec* is strictly related to *-ec*, *-lka* to *-ka*, and *-l'ščik* to *-ščik/čik*. The suffix *-ščik/čik*, in turn, is strictly related to *-nik*, as they both originate from OCS *-ikŭ* (cf. Section 3.1.2). Considering that these constructions only include few compounds compared to the other constructions, and thus are not equally

represented in the corpus, they could be merged with the bigger constructions to which they are historically related in order to facilitate statistical analyses (i.e. *-lec* with *-ec*, *-lka* with *-ka*, *-l'ščik* and *-ščik/čik* with *-nik*).

However, before doing so, it is important to check whether statistical tests show significant differences between the suffixes that we want to merge. In order to do so, I resort to the technique of conditional inference trees (that will be discussed more in detail in Section 4.4.3), i.e. tree-structure models of regression and classification based on binary recursive partitioning (Levshina 2015: 291).

The comparison of the suffixes *-ec* and *-lec* gives in output the tree shown in Figure 15, where we see that the differences between the two suffixes seem to mainly depend on the parameters "transitivity" and "semantics". More specifically, the construction in *-lec* seems to attract intransitive verbal bases and compounds denoting Carriers of State, differently from *-ec*.

Figure 15. Conditional inference tree: *-ec* vs. *-lec*

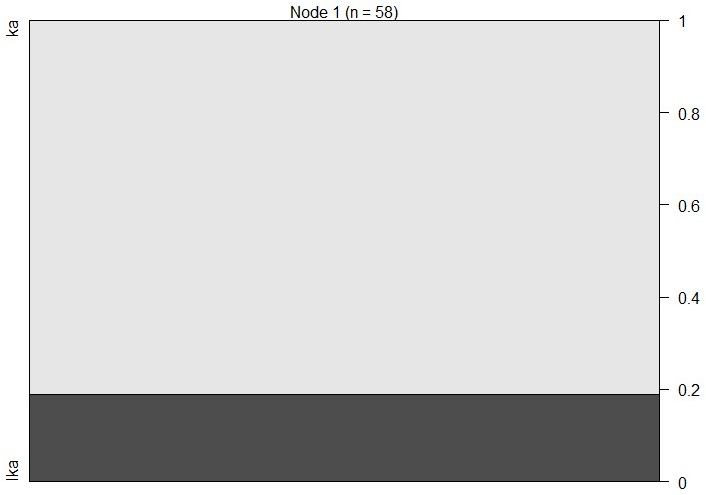


However, these results are the result of the fact that the suffix *-lec* in our database is only used within one specific construction based on the verb *vladet'* 'own, possess'. Thus, all

compounds formed with this suffix will show the same features precisely because they belong to the same construction. The features "intransitive verbal base" and "carrier of state" are nonetheless significantly represented also in compounds formed with the suffix *-ec* (14% of them show intransitive bases, and 35.7% of them denote Carriers of State), but this does not emerge from the tree shown in Figure 15 because it is significant that *all* compounds in *-lec* show these features, while *not all* compounds in *-ec* do so. However, we can consider that *-lec* shows some features that are also typical of *-ec*, and that the two suffixes can be merged before applying the subsequent analyses.

Regarding the suffixes *-ka* and *-lka*, the comparison produces the tree represented in Figure 16.

Figure 16. Conditional inference tree: *-ka* vs. *-lka*

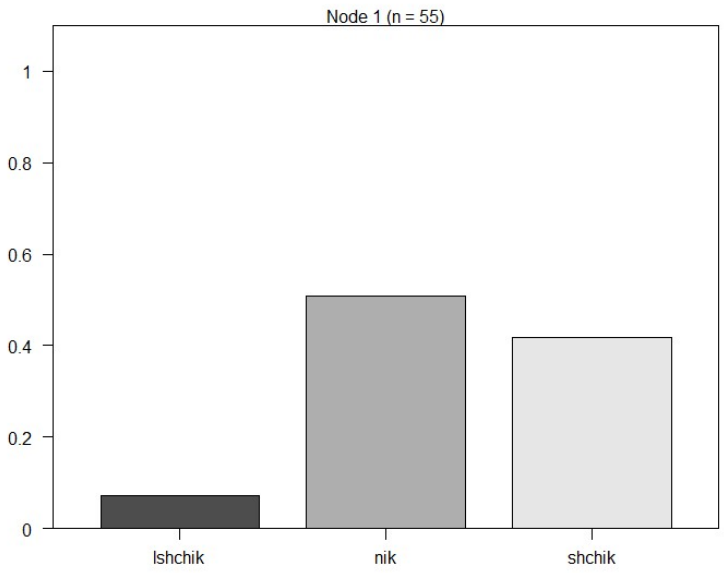


Notably, the model is not able to identify any parameter that would distinguish significantly between the two suffixes, which confirms that it is reasonable to merge the two suffixes in the analyses that follow.

Finally, I compare the suffixes *-l'ščik*, *-ščik/čik* and *-nik* to see whether their behavior is similar enough and, consequently, whether it makes sense to merge the three suffixes. The tree in Figure 17 seems to confirm this assumption, as the model does not identify any parameter that discriminates between the three suffixes. Thus, here again, it

seems reasonable to merge the suffixes before moving to the general comparison that is described in Section 4.4.3.

Figure 17. Conditional inference tree: *-l'ščik* vs. *-ščik/čik* vs. *-nik*



Now that the low-frequency constructions have been merged with the constructions to which they are historically related, I can apply multinomial logistic regression, a statistical model that allows investigating the distribution of rival constructions when the possible outcomes are more than two. The comparison now includes five possible outcomes: *-ec* (which also includes *-lec*), *-tel'*, *-ik* (which includes *-l'ščik*, *-ščik/čik*, and *-nik*), *-ka* (which also includes *-lka*), and the suffixless construction.

To fit the multinomial logistic regression model, I use the function polytomous() available in the R package "polytomous" (cf. Levshina 2015: 283). The results in Figure 18 show the table of regression coefficients, represented as log odds. Positive coefficients boost the chances of the corresponding construction, while negative coefficients decrease the chances of the corresponding construction. As shown in the regression table, when the referent is inanimate, the favored suffixes are *-ka* and the suffixless construction, which show positive coefficients. A perfective verb base favors the occurrence of the suffix *-ik* and, even more significantly, of the suffix *-tel'*. The results corresponding to

the predictor "part of speech" show that the constructions subject to more variation are *-ec*, *-ka* and the suffixless construction, since they show positive coefficients for values different from "noun". The coefficients related to the semantics of the compounds confirm what already discussed in Section 4.3.6, i.e. that the construction in *-ec* is more often associated with Carriers of State ("cos" in Figure 18) compared to the other constructions, and that the constructions in *-ik*, *-ka* and *-tel'* are more frequently employed to denote Instruments as compared with *-ec* and the suffixless construction.

Figure 18. Multinomial logistic regression model

```
Formula:
suffix ~ semantics + animacy + aspect + transitivity + pos +
    sr

Heuristic:
one.vs.rest

Log-odds:
                      ec     ik     ka     tel    zero
(Intercept)       -0.9658  -2.91 -3.772 -0.8233 -0.7034
animacyinan       -0.2563 -1.759  2.241  -4.831   1.228
aspectpfv        (-16.67) 0.8704 -4.246   3.061   -5.08
posother           0.2104 -3.504  1.102  -0.819  0.3606
semanticscos        1.048 -2.148 -1.727 -0.1536  -0.729
semanticsinstr     -1.849  1.582 0.4772   3.777 -0.1095
srother          -0.06068 -1.029 -2.564 -0.6618  0.8258
transitivityintr  -0.6416  0.216 -6.456 -0.8853   1.156

Null deviance:              429700  on  930930  degrees of freedom
Residual (model) deviance:  321000  on  930890  degrees of freedom

R2.likelihood:  0.2529
AIC:            321100
BIC:            321500
```

If we now consider the statistics related to the regression model (cf. Figure 19), we see some goodness-of-fit measures. The R^2 coefficient (coefficient of determination) shows how good the model fits the data. Since values from 0.2 to 0.4 are considered to indicate a very good fit (cf. Levshina 2015: 280), this model seems to fit the data well ($R^2 = 0.25$).

Figure 19. Multinomial logistic regression model: statistics

```
$df.null
[1] 930930

$df.model
[1] 930890

$AIC.model
[1] 321127.3

$BIC.model
[1] 321532.6

$loglikelihood.null
[1] -214858.8

$loglikelihood.model
[1] -160523.6

$deviance.null
[1] 429717.7

$deviance.model
[1] 321047.3

$R2.likelihood
[1] 0.252888

$R2.nagelkerke
[1] 0.4909836

$crosstable

          ec     ik     ka    tel   zero
  ec     8753      0      0      0  25482
  ik      130      0      0    799   4159
  ka      117      0      0      4   6275
  tel    3886      0      0   6937  22903
  zero   5622      0      0     26 101093

$accuracy
[1] 0.6272384

$recall.predicted
       ec        ik        ka       tel      zero
0.2556740 0.0000000 0.0000000 0.2056870 0.9470869

$precision.predicted
       ec        ik        ka       tel      zero
0.4729306       NaN       NaN 0.8932526 0.6321789

$lambda.prediction
[1] 0.1264019

$tau.classification
[1] 0.3815901

$d.lambda.prediction
[1] 47.05381
```

```
$d.tau.classification
[1] 202.8287

$p.lambda.prediction
[1] 0

$p.tau.classification
[1] 0
```

The cross table on which the accuracy measure is based and the measures "recall" and "precision" show null values for the constructions in *-ik* and *-ka*, which is probably due to the low frequencies of these constructions with respect to the others. Thus, I repeat the analysis after removing the constructions in *-ik* and *-ka* from the data, and I obtain the statistics shown in Figure 20.

Figure 20. Multinomial logistic regression model: statistics for *-ec*, *-tel'* and *-ø*

```
$df.null
[1] 524106

$df.model
[1] 524082

$AIC.model
[1] 240759.3

$BIC.model
[1] 241001

$loglikelihood.null
[1] -163859.1

$loglikelihood.model
[1] -120355.6

$deviance.null
[1] 327718.1

$deviance.model
[1] 240711.3

$R2.likelihood
[1] 0.2654929

$R2.nagelkerke
[1] 0.4632538
```

```
$crosstable

          ec    tel   zero
  ec    7719      0  26516
  tel   3877   6937  22912
  zero  5132     26 101583

$accuracy
[1] 0.6653559

$recall.predicted
       ec       tel      zero
0.2254710 0.2056870 0.9516774

$precision.predicted
       ec       tel      zero
0.4614419 0.9962660 0.6726861

$lambda.prediction
[1] 0.1397566

$tau.classification
[1] 0.3926862

$d.lambda.prediction
[1] 46.6107

$d.tau.classification
[1] 181.8309

$p.lambda.prediction
[1] 0

$p.tau.classification
[1] 0
```

The R^2 coefficient still shows that the model fits the data well (0.26). The accuracy measure shows how correctly the model predicts the distribution of the constructions and it is calculated by dividing the sum of the numbers in the diagonal of the cross table (showing the number of correct predictions for each construction) by the number of observations. In this case, the accuracy of the model is 66% (higher than in the previous model, where the accuracy was 62%). The measure "recall" indicates the proportion of compounds for each construction that were predicted for the model. Unfortunately, only the suffixless construction shows a good recall value, while for *-ec* and *-tel'* the values are quite low. The measure "precision" shows how many times the predictions made by the model for each construction were correct. Here *-tel'* shows the highest precision value, followed by the suffixless construction and *-ec*.

Figure 21. Multinomial logistic regression model: *-ec* vs. *-tel'* vs. *-ø*

```
Formula:
suffix ~ semantics + animacy + aspect + transitivity + pos +
    sr

Heuristic:
one.vs.rest

Log-odds:
                        ec      tel     zero
(Intercept)         -0.8711 -0.7127  -0.492
animacyinan         -0.2063  -4.771   1.382
aspectpfv          (-16.76)   6.427  -5.285
posother             0.2087 -0.8423  0.2943
semanticscos         0.9868 -0.2439 -0.8191
semanticsinstr       -1.863    3.94  0.3668
srother           (-0.03203) -0.7983  0.4927
transitivityintr    -0.7321 -0.9495   1.072

Null deviance:              327700  on  524106  degrees of freedom
Residual (model) deviance:  240700  on  524082  degrees of freedom

R2.likelihood:  0.2655
AIC:            240800
BIC:            241000
```

Figure 21 shows the results produced by the multinomial logistic regression model applied to the suffixes *-ec*, *-tel'* and to the suffixless construction. The results mostly confirm the previous statistics and show the following: inanimate referents disfavor the occurrence of *-ec* and *-tel'*, while they favor the occurrence of the suffixless construction; the perfective aspect of the verb base boosts the chances of the suffix *-tel'*; first member's parts of speech other than noun favor the construction in *-ec* and the suffixless construction; compounds denoting Carriers of State favor the occurrence of the suffix *-ec*, while compounds denoting Instruments favor the occurrence of the suffix *-tel'* and, to a lesser extent, the suffixless construction. The occurrence of the suffixless construction is also favored by first member's semantic roles other than Theme/Patient and by intransitive verbal bases.

4.4.3. Conditional inference trees and random forests

Other statistical models that can be used to compare rival constructions are conditional inference trees and random forests (cf. Breiman 2001; Strobl, Malley & Tutz 2009), which were introduced to linguistic analysis by Tagliamonte & Baayen (2012), and have

been employed in a number of studies to discriminate the behavior of linguistic constructions (cf. Baayen et al. 2013; Lohmann 2013; Endresen 2014; Nesset & Makarova Forth.). As models of logistic regression, random forests also aim at determining which outcome is most probable when we have a series of predictors. However, while logistic models predict the outcome based on a mathematical equation that indicates how each predictor affects the outcome, random forests "work through the data and, by trial and error, establish whether a variable is a useful predictor" (Tagliamonte & Baayen 2012). Tagliamonte & Baayen (2012) explain how random forests work: "The algorithm works through all predictors, splitting (partitioning) the data into subsets where justified, and then recursively considers each of the subsets, until further splitting is not justified. In this way, the algorithm partitions the input space into subsets that are increasingly homogeneous with respect to the levels of the response variable. The result of this recursive binary splitting of the data is a conditional inference tree". When more than one useful predictor is found, the model selects the predictor with the strongest association with the response, while the algorithm stops when no useful predictors are found. Random forests produce a large number of conditional inference trees by randomly sampling different subsets of the data. As pointed out by Tagliamonte & Baayen (2012) and Levshina (2015: 292), conditional inference trees and random forests are particularly well applicable to situations in which the token number is relatively small, while the number of predictors is large.

To apply this model to the data, I used the functions ctree() and cforest() available in the R package "party" (A Laboratory for Recursive Partitioning). The input document is a tab-delimited file in which each row corresponds to a different lemma, while each column corresponds to a different parameter (semantics of the compound, animacy of the referent, aspect of the verbal base, transitivity of the verbal base, part of speech of the non-verbal base, semantic role of the non-verbal base). In this case, I keep the column indicating the number of tokens as a parameter, and do not multiply each row by the number of tokens as done for binomial and multinomial logistic regression.

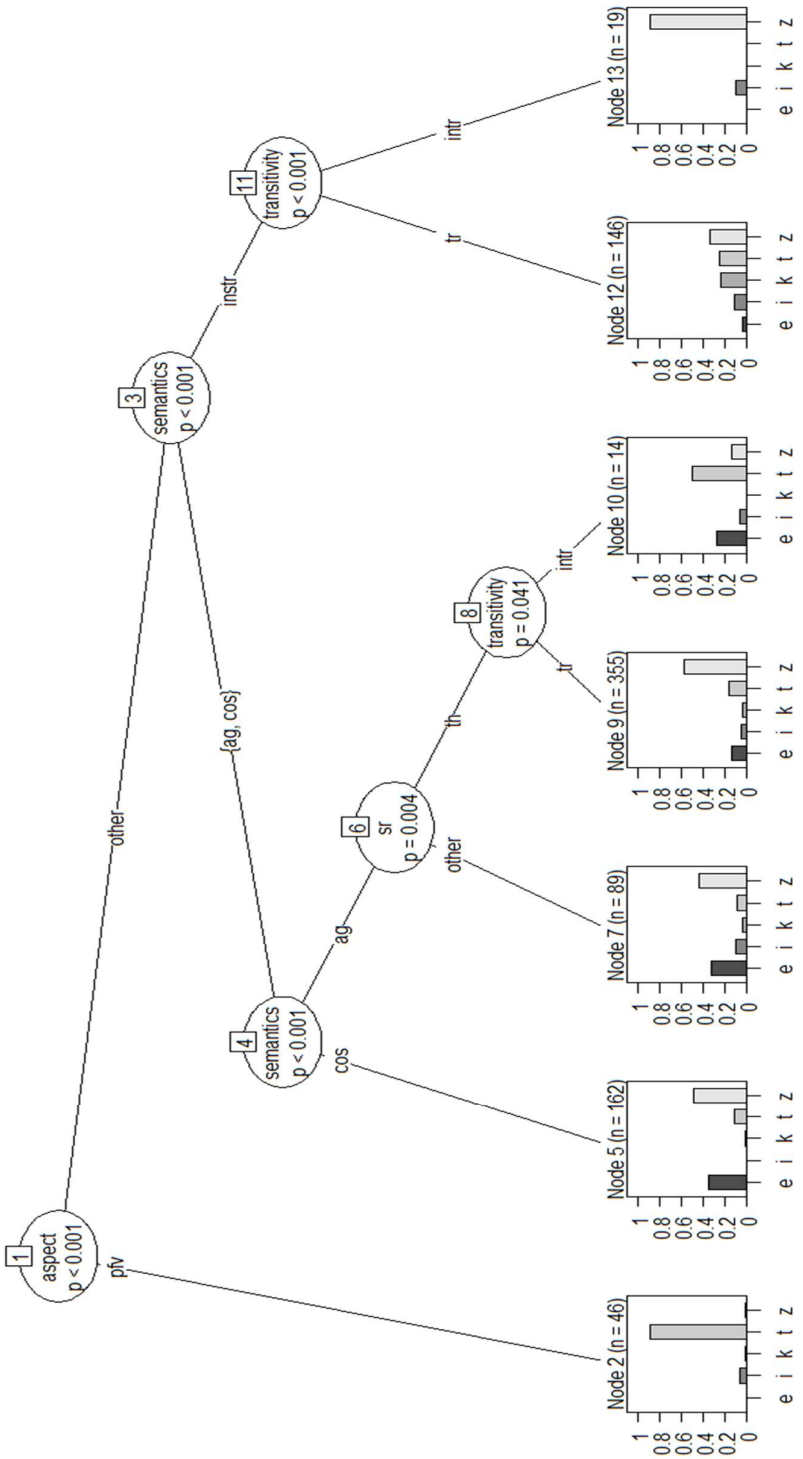
When the function ctree() is applied, the outcome is the tree reproduced in Figure 22. As can be observed, the ovals contain the names of the variables and the corresponding *p*-values, while the branches specify the levels of the variables. At the

bottom of the figure, the bar plots indicate the proportion of each suffix[34] in each end node, and the number of observations for each end node is shown above the bar plot.

The first split at the top divides perfective and non-perfective verbal bases. Perfective bases are contained in the end Node 2, where most observations correspond to the suffix *-tel'*. Node 3 splits compounds with non-perfective bases according to the parameter "semantics" into Prototypical Agents and Carriers of State on the one side, and Instruments on the other. Node 4 splits Carriers of State and Prototypical Agents. Carriers of State are contained in the end Node 5, where most occurrences correspond to the construction in *-ec* and to the suffixless construction. Prototypical Agents are split at Node 6 according to the semantic role of the first element of the compounds: those compounds in which the first element plays a semantic role other than Theme/Patient end up in Node 7, while those in which the first element plays the semantic role of Theme/Patient are further split at Node 8 according to the transitivity of the verbal element. Compounds showing a transitive verbal base end up in Node 9, while those showing an intransitive verbal base end up in Node 10. Instruments are split at Node 11 according to the transitivity of the verbal base, and most of them end up in Node 12, which includes compounds based on transitive verbs. Most of the compounds (42.7%) end up in Node 9, which includes agent nouns based on non-perfective transitive bases showing a first element playing the semantic role of Theme/Patient. Node 5 includes 162 compounds (19.5%) denoting Carriers of State based on non-perfective verbs. Node 12 includes 146 compounds (17.6%) denoting Instruments based on non-perfective transitive verbs. These three nodes correspond to the most frequent combinations of factors, while the remaining end nodes include compounds showing less common patterns.

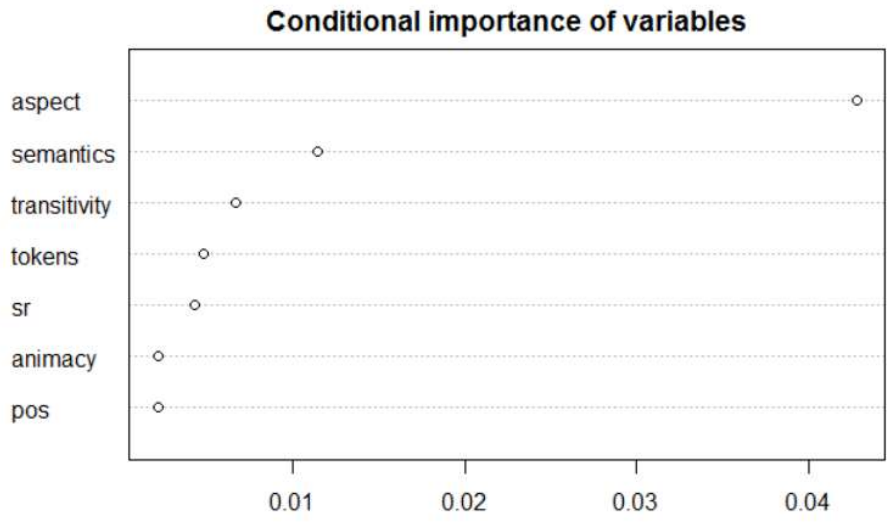
[34] To improve visualization, the following abbreviations are employed in the plot: e = *-ec*, i = *-ik*, k = *-ka*, t = *-tel'*, z = suffixless construction.

Figure 22. Conditional inference tree: *-ec*, *-ik*, *-ka*, -tel' and *-ø*



If we now apply the function cforest() to obtain a random forest, and then calculate the measures of variable importance for the random forest through the function varimp(), we get the results showed in the dot chart in Figure 23.

Figure 23. Dot chart of conditional variable importance



The dot chart shows that "aspect" is the predictor that best discriminates the data (0.042), followed by "semantics" (0.011), "transitivity" (0.007), "tokens" (0.004), "semantic role" (0.004), "part of speech" (0.003), and "animacy" (0.002).[35] The variables whose values are around zero are irrelevant in discriminating the data. To see how well the model fits the data, the accuracy measure can be computed. In this case, the accuracy is about 52%, that is twice as large as the one that we could expect by chance.

4.4.4. To sum up

The statistical analyses carried out in Section 4.4 have allowed shedding more light on the results of the comparison of rival word-formation constructions giving rise to compound agent nouns.

[35] The exact values are obtained by applying the function round().

First, the results of logistic regression models have shown that the differences among rival constructions are better captured when all the constructions are compared with each other, rather than when opposing the suffixed constructions to the suffixless construction. Indeed, although the binomial logistic model opposing "suffixed" vs. "suffixless" shows significant results (see Section 4.4.2.1), it cannot determine which of the parameters considered is more relevant in distinguishing the opposed constructions. The application of generalized linear models and their comparison through ANOVA have led to the conclusion that every parameter is significant for the model, which prevents from drawing significant conclusions.

The results of multinomial logistic regression have given us a better understanding of the similarities/differences of the rival constructions analyzed, as significant differences are found among different constructions showing an expressed suffix. In particular, the results have shown which constructions behave more similarly depending on the parameter considered. For example, the constructions in *-ec*, *-ka* and the suffixless constructions are similar in that they allow for greater variation with respect to the other constructions as regards the part of speech of the first element of the compound. The constructions in *-tel'* and *-ik* (including *-nik*, *-čik*/*ščik* and *-l'ščik*), instead, are similar in that they are based on perfective bases more often that the other constructions, and in that they often give rise to nouns denoting Instruments, which is less typical of the other constructions. The construction in *-ec* stands out with respect to the other constructions when the parameter "semantics" is considered. Indeed, this construction is much more often associated with nouns denoting Carriers of State compared to the other constructions, while it is rarely employed to create instrument nouns. In these respects, the suffixless construction is the construction that behaves more similarly to the construction with the suffix *-ec*.

Conditional inference trees and random forests have proved to be especially useful to determine which parameters are more relevant in splitting the data into different groups. As has been shown (see Section 4.4.3), the parameter "aspect" reveals a substantial difference between the construction with the suffix *-tel'* and the other constructions: it is almost only compounds in *-tel'* that show perfective verb bases. The parameter "semantics" clearly shows the different behavior of the construction with the suffix *-ec* compared to the other constructions: *-ec* is much less often associated with

instrument nouns, while its association with nouns denoting Carriers of State seems much more significant, which also confirms the results of multinomial logistic regression. The parameters "part of speech", "semantic role" and "transitivity" have appeared to contribute less to distinguishing between rival constructions. The parameter "animacy" has also appeared as less significant, although the semantics of the compound is strictly related to the animacy of the referent.

To further investigate the factors playing a role in the existence (and survival) of rival constructions creating compound agent nouns in Russian, I examine the diachronic and stylistic components in Section 4.5.

4.5. Diachronic and stylistic differentiation of rival word-formation constructions

In the present section, I discuss the role of diachronic and stylistic factors in the distribution and differentiation of rival word-formation constructions. In order to do so, I exploit the information provided by the RNC and, in particular:

a) the date of creation of the texts in which compounds occur;
b) the textual genres in which compounds occur;
c) the occurrence of compounds in written and spoken registers.

The aim of this analysis is to determine whether these factors significantly affect the distribution of rival constructions.

4.5.1. Distribution over time

To investigate the distribution of rival word-formation constructions over time, I checked the number of occurrences of the compounds belonging to each construction in different time spans. To carry out this search, I divided the whole period covered by the RNC main corpus in time spans of fifty years each (excluding the last time span, which covers seventeen years, i.e. from 2000 to 2016) and determined how many occurrences for each construction are found in each time span. Table 17 shows the results of this search.

Table 17. Observed frequencies (RNC) of rival constructions in different time spans

	1700-1749	**1750-1799**	**1800-1849**	**1850-1899**	**1900-1949**	**1950-1999**	**2000-2016**
-ec	156	928	3,201	5,378	10,257	7,006	5,462
-lec	0	0	13	701	1,465	403	458
-tel'	230	2,808	2,567	4,674	6,963	6,915	10,282
-nik	14	38	125	310	484	727	1,366
-ščik/čik	46	53	33	78	54	230	240
-l'ščik	0	0	1	1	10	140	1,221
-ka	2	15	109	514	1,304	2,433	1,541
-lka	0	0	0	12	146	199	251
-ø	26	574	1,991	11,027	30,573	38,665	27,432

In order to determine whether the differences in the distribution of rival constructions in different time spans are statistically significant, I compare the observed and expected frequencies and calculate the Pearson residuals, i.e. the differences between the observed and expected frequencies divided by the square root of the expected value (cf. Levshina 2015: 218). To calculate the expected frequencies, I first multiply the total of each row by the total of each column, and then divide the result by the total number of observations. The results are shown in Table 18.

Table 18. Expected frequencies (rounded to integers) of rival constructions in different time spans

	1700-1749	**1750-1799**	**1800-1849**	**1850-1899**	**1900-1949**	**1950-1999**	**2000-2016**
-ec	80	745	1,356	3,828	8,645	9,566	8,132
-lec	7	70	127	360	812	899	764
-tel'	85	793	1,444	4,075	9,203	10,183	8,657
-nik	8	71	128	363	819	906	770
-ščik/čik	67	17	31	87	196	217	185
-l'ščik	3	32	58	162	367	406	345
-ka	15	136	248	700	1,581	1,750	1,488
-lka	2	14	25	72	162	180	153
-ø	273	2,539	4,623	13,049	29,471	32,611	27,723

Now I can calculate the Pearson residuals using the formula shown in (40).

(40) (observed frequency – expected frequency) / √expected frequency

The results are shown in Table 19. If the residual is positive, the observed frequency is greater than expected; if the residual is negative, the observed frequency is smaller than expected. The values corresponding to the Pearson residuals should be interpreted as follows: "[t]he greater the absolute value of a residual, the greater the discrepancy between the observed and expected frequencies, and the more it contributes to the test statistic. In fact, the χ^2-statistic is the sum of squared Pearson residuals" (Levshina 2015: 218). Thus, great absolute values indicate statistically significant results. For instance, the Pearson residuals of the construction in *-ec* for the first half of the 19th century equal 50.103. The high positive value tells us that this construction is strongly overrepresented in the given time span. Results closer to zero indicate lower discrepancy between the observed and expected frequencies and, thus, less significant results.

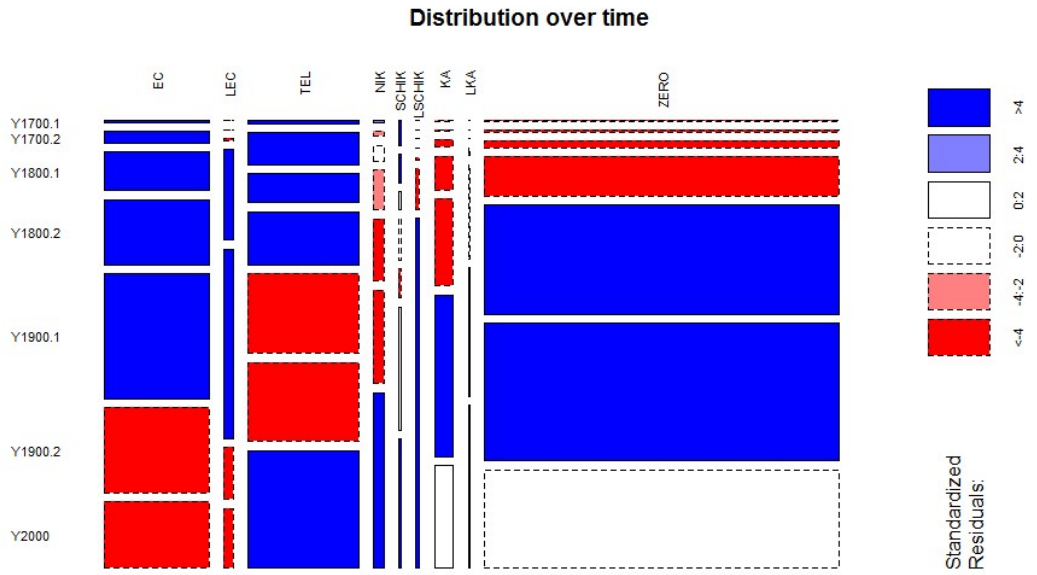
Table 19. Distribution over time: Pearson residuals

	1700-1749	**1750-1799**	**1800-1849**	**1850-1899**	**1900-1949**	**1950-1999**	**2000-2016**
-ec	8.497	6.705	50.103	25.052	17.337	-26.174	-29.608
-lec	-2.646	-8.367	-10.116	17.972	22.916	-16.542	-11.071
-tel'	15.727	71.555	29.553	9.383	-23.350	-32.385	17,465
-nik	2.121	-3.916	-0.265	-2.782	-11.706	-5.947	21.478
-ščik/čik	-2.566	8.731	0.359	-0.965	-10.143	0.882	4.044
-l'ščik	-1.732	-5.657	-7.484	-12.649	-18.635	-36.241	83.203
-ka	-3.357	-10.376	-8.826	-7.030	-6.966	16.327	1.374
-lka	-1.414	-3.742	-5	-7.071	-1.257	1.416	7.923
-ø	-14.949	-38.997	-38.710	-17.701	6.419	33.524	-1.748

The Pearson residuals can be represented in a mosaic plot, which can be very useful if we want to visualize the results in a more straightforward way. This can be done by resorting to the function "mosaicplot()" implemented in the R package "vcd" (Visualizing categorical data), and the output is the plot shown in Figure 24. The different time spans are represented on the vertical axis (Y1700.1 stands for the first half of the 18th century,

Y1700.2 stands for the second half of the 18th century, and so on), while rival constructions are represented on the horizontal axis. Blue-shaded rectangles show which constructions are overrepresented in a given time span (and indeed correspond to positive residuals in Table 19), while red-shaded rectangles show which constructions are underrepresented in a given time span (corresponding to negative residuals in Table 19). White rectangles indicate that there is not a significant difference between the observed and the expected frequencies. The size of the rectangles reflects the proportions of the cells in the contingency table: the suffixless construction occupies a larger area with respect to the other constructions because it shows a higher number of total occurrences; the construction in *-lka* is almost invisible in the plot because it shows a very low number of total occurrences with respect to the other constructions.

Figure 24. Distribution over time: mosaic plot



If we focus on the three most frequent constructions (i.e. *-ec*, *-tel'* and the suffixless construction), we can make the following observations. The construction in *-ec* is overrepresented in earlier texts, while it is underrepresented in texts belonging to the 20th and 21st centuries. The suffixless construction shows the opposite situation: it is underrepresented in earlier texts, while it is overrepresented in later texts (cf. Chapter 5,

Section 5.4). Finally, the construction in *-tel'* is overrepresented especially starting from the second half of the 20th century.

Turning to low-frequency constructions, we can observe a general trend for them to be underrepresented in earlier texts, while they are normally represented or overrepresented in later texts, with the exception of the suffix *-lec*, which seems to behave more similarly to the suffix *-ec*, which is understandable if we consider that the two suffixes are strictly related.

In general, it seems that diachrony is a significant factor in discriminating the behavior of rival word-formation constructions. In particular, the most striking difference is observed between the construction in *-ec* and the suffixless construction, which could lead us to assume that one (i.e. the suffixless construction) is slowly replacing the other (i.e. *-ec*) as time goes by. As we will see, this assumption can be somehow confirmed if we compare doublets (see Section 4.6), that is pairs of compounds sharing the same lexical bases, but belonging to different constructions. These cases mainly concern the opposition of the construction in *-ec* and the suffixless construction, and, as I will show, diachrony plays a major role in distinguishing between the two members of each pair.

4.5.2. Distribution across genres

Another important aspect that could give us some interesting insights about the differentiation of rival word-formation constructions regards their distribution across different textual genres. The RNC interface allows, for each word searched, to visualize some statistics, including the distribution of the occurrences of that word across different textual genres (the label *sfera funkcionirovanija* is used in the corpus interface to indicate the textual genre). Thus, for each compound, I checked the number of occurrences in different textual genres in the RNC. Consistently with the approach adopted in other studies (cf. Nesset & Makarova Forth.), I grouped the labels used in the RNC to specify the textual genres in which compounds occur into five broader genres:

a) Journalism (including journalistic, advertising, non-fiction, and everyday life texts);
b) Fiction (including fiction, drama, memoirs, and biographies);

c) Scientific/Educational (including scientific, popular scientific, instructional, religious, and philosophical texts);
d) Technical;
e) Official/Business.

This choice was made to simplify statistical analyses, considering that the labels adopted in the corpus are extremely fine-grained and, therefore, the results would have included too many groups with no clear-cut differences between one another.

By following the same procedure described in Section 4.5.1 for the distribution of rival constructions over time, I now compare the observed and expected frequencies of rival constructions across different textual genres and calculate the Pearson residuals to see which constructions are over- or underrepresented in a certain genre. Table 20 shows the observed frequencies of each construction in each textual genre, while Table 21 shows the expected frequencies.

Table 20. Observed frequencies (RNC) of rival constructions in different textual genres

	Journalism	**Fiction**	**Scientific/ Educational**	**Technical**	**Official/ Business**
-ec	13,806	11,190	5,926	42	317
-lec	1,700	521	552	0	187
-tel'	16,221	8,292	6,694	921	1,610
-nik	1,471	1,018	375	44	94
-ščik/čik	307	249	107	19	39
-l'ščik	1,034	46	220	1	65
-ka	2,035	3,460	295	14	10
-lka	262	260	39	35	1
-ø	48,988	48,316	7,045	983	1,373

Table 21. Expected frequencies (rounded to integers) of rival constructions in different textual genres

	Journalism	**Fiction**	**Scientific/ Educational**	**Technical**	**Official/ Business**
-ec	14,419	12,324	3,571	352	621
-lec	1,364	1,166	338	33	59

-tel'	15,552	13,292	3,851	373	670
-nik	1,384	1,183	343	33	60
-ščik/čik	332	284	82	8	14
-l'ščik	630	538	156	15	27
-ka	2,680	2,291	664	64	115
-lka	275	235	68	7	12
-ø	49,187	42,039	12,180	1,180	2,118

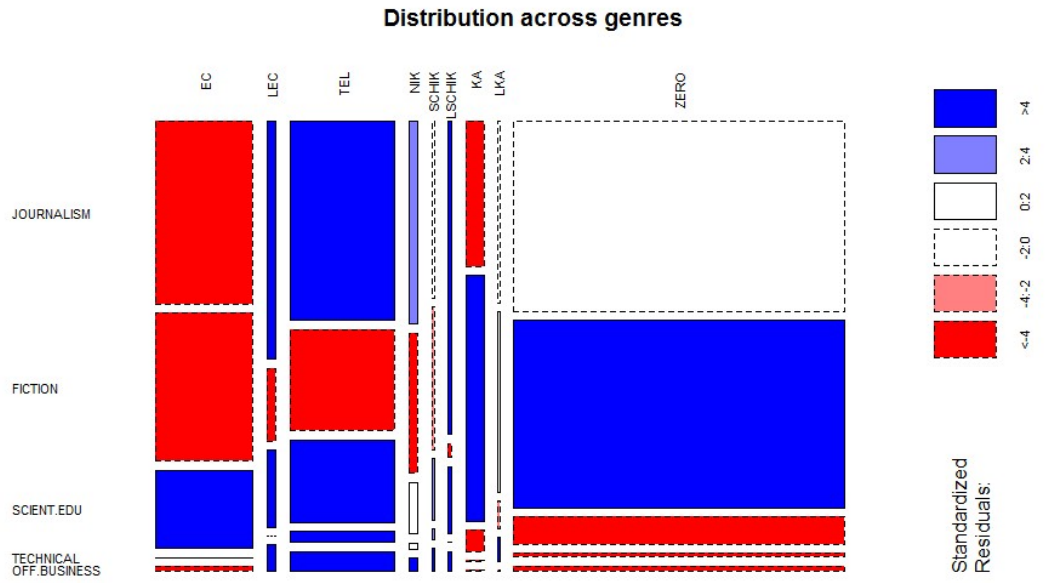
The residuals, calculated using the formula in (40), are shown in Table 22. Again, if the residual is positive, the observed frequency is greater than expected; if the residual is negative, the observed frequency is smaller than expected. Greater absolute values indicate greater statistical significance.

Table 22. Distribution across genres: Pearson residuals

	Journalism	**Fiction**	**Scientific/ Educational**	**Technical**	**Official/ Business**
-ec	-5.105	-10.215	39.409	-16.523	-12.199
-lec	9.098	-18.889	11.640	-5.745	16.664
-tel'	5.364	-43.368	45.813	28.374	36.315
-nik	2.339	-4.797	1.728	1.915	4.389
-ščik/čik	-1.372	-2.077	2.761	3.889	6.681
-l'ščik	16.096	-21.212	5.124	-3.615	7.313
-ka	-12.459	24.423	-14.320	-6.25	-9.791
-lka	-0.784	1.631	-3.517	10.583	-3.175
-ø	-0.897	30.614	-46.528	-5.735	-16.188

The plot representing the results corresponding to the distribution of the constructions across genres is shown in Figure 25. Textual genres are represented on the vertical axis, while rival constructions are represented on the horizontal axis. Again, blue-shaded rectangles show which constructions are overrepresented in a given textual genre (and indeed correspond to positive residuals in Table 22), while red-shaded rectangles show which constructions are underrepresented in a given textual genre (corresponding to negative residuals in Table 22). White rectangles indicate that there is not a big difference between the observed and the expected frequencies.

Figure 25. Distribution across genres: mosaic plot



If we look at the plot, we can draw some interesting conclusions. First, some constructions appear to be overrepresented in technical and official discourses, i.e. *-lec*, *-tel'*, *-nik*, *-ščik/čik*, *-l'ščik* and *-lka*. This appears to be consistent with the features of these constructions.

As regards the low-frequency constructions in *-lec*, *-l'ščik* and *-lka*, it must be noted that they show little variation in terms of the type of nouns that they create, so their semantics is usually very specific. The construction in *-lec* only forms compounds based on the verb *vladet'* 'own, possess' (e.g. *zemlevladelec* 'land owner'), which are likely to occur in the official discourse. The construction in *-l'ščik* only includes four compounds and 1358 out of its 1366 total occurrences belong to the compound *nalogoplatel'ščik* 'tax payer', which is obviously very likely to be used in the official discourse. The construction in *-lka*, instead, is only used to create nouns denoting Instruments and is often used in the technical discourse.

Turning to the more frequent constructions in *-tel'*, *-nik* and *-ščik/čik*, these are among the constructions that most frequently give rise to instrument nouns, which could justify their high frequency in technical discourse compared to other constructions.

By contrast, the constructions in *-ec* and *-ka*, and the suffixless construction appear to be underrepresented in technical and official discourses. The suffix *-ka* is indeed characterized by a colloquial nature, which also explains its overrepresentation in fictional texts. The suffix *-ec* and the suffixless construction, in turn, could be seen as more "neutral", less specialized constructions, which is mostly due to the fact that they show large polysemy patterns (cf. Chapter 3). However, despite being similar in terms of their underrepresentation in technical and official discourses, they nonetheless show different behaviors concerning their distribution in the remaining genres. Indeed, while the suffixless construction is normally represented or overrepresented in journalistic and fictional texts, the construction in *-ec* is underrepresented in these textual genres, but it is overrepresented in scientific/educational texts. This might also be related to diachronic factors: the scientific/educational genre includes religious texts, which make up a large portion of the earlier texts included in the RNC. This, in turn, explains why the construction in *-ec* is so frequent in this genre: as already mentioned in section 4.5.1, the semantics of the compounds belonging to this construction is often associated with religious concepts.

Thus, the analysis has shown that stylistic features also play a significant role when it comes to determine the differences between rival word-formation constructions. As we have seen, some of them (especially the low-frequency constructions) seem to have acquired more specialized meanings, which makes them overrepresented in technical and official/business discourses, whereas the other constructions (showing higher frequencies) seem to be characterized by a less specialized semantics, which makes them underrepresented in specialized discourses.

4.5.3. Written vs. spoken registers

To further investigate the stylistic differentiation of the rival word-formation constructions under examination, I consider their distribution in spoken texts and compare it to their distribution in written texts. In order to do so, I used the spoken subcorpus of the RNC to check the frequency of the compounds analyzed so far. The spoken subcorpus

of the RNC includes about 12 million words and is made up of transcriptions of public and private discourses, as well as transcriptions of movies.

Similarly to what I have done for time and genres, I now compare the observed and expected frequencies of rival constructions across different registers (i.e. written vs. spoken registers) and calculate the Pearson residuals to determine which constructions are over- or underrepresented in a certain register. Table 23 and Table 24 show the observed and expected frequencies of each construction in each register.

Table 23. Observed frequencies (RNC) of rival constructions in written and spoken registers

	Written	**Spoken**
-ec	32,352	473
-lec	3,040	24
-tel'	34,439	808
-nik	3,064	107
-ščik/čik	734	22
-l'ščik	1,373	47
-ka	5,918	184
-lka	608	8
-ø	110,288	3,485

Table 24. Expected frequencies (rounded to integers) of rival constructions in written and spoken registers

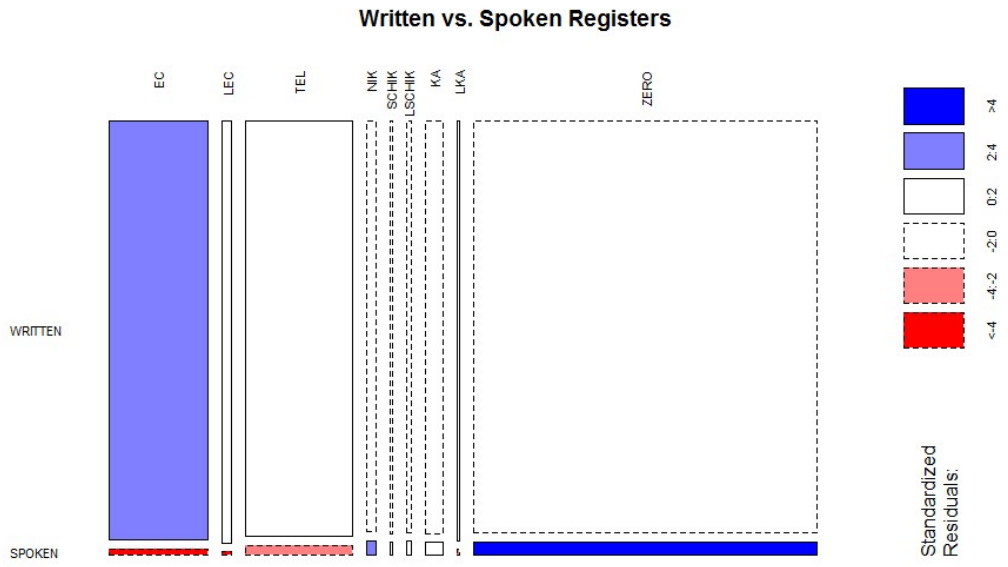
	Written	**Spoken**
-ec	31,965	860
-lec	2,984	80
-tel'	34,324	923
-nik	3,088	83
-ščik/čik	736	20
-l'ščik	1,383	37
-ka	5,942	160
-lka	600	16
-ø	110,794	2,979

The Pearson residuals, calculated using the formula in (40) are shown in Table 25, while the plot in Figure 26 represents the results corresponding to the distribution of rival constructions in written and spoken registers.

Table 25. Written vs. spoken registers: Pearson residuals

	Written	**Spoken**
-ec	2.1646	-13.197
-lec	1.025	-6.261
-tel'	0.621	-3.785
-nik	-0.432	2.634
-ščik/čik	-0.074	0.447
-l'ščik	-0.269	1.644
-ka	-0.311	1.897
-lka	0.327	-2
-ø	-1.520	9.271

Figure 26. Written vs. spoken registers: mosaic plot



The comparison shows a clear imbalance between written and spoken registers, which is partly due to the different sizes of the corpora: the main (written) subcorpus includes over 200 million words, whereas the spoken subcorpus only contains about 12 million words. However, this also depends on the fact that most of the compounds included in our database are only common in written registers, while they are very rarely (or never) used

in the spoken discourse. Despite the huge size difference between the two registers, the plot in Figure 26 shows some interesting trends for some of the constructions under examination.

Once again, the most interesting observations arise from the comparison of the construction in *-ec* with the suffixless construction. Notably, while the suffix *-ec* is overrepresented in the written corpus and underrepresented in the spoken corpus, the suffixless construction shows the opposite distribution: it is slightly underrepresented in the written corpus, whereas it is overrepresented in the spoken corpus. Most compounds in *-ec* show their first occurrence in the RNC in texts belonging to the 18th and 19th centuries (cf. Section 6.4), they are often perceived as archaisms and rarely appear in spoken texts. More than half of the compounds in *-ec* contained in the database do not appear in the spoken subcorpus of the RNC (77 out of 143). Even when the compounds are featured in the spoken subcorpus, they often show low frequencies and are mostly used in specific contexts. Example (41)[36] is taken from an interview with the protopope Oleg Stenjaev that was broadcast live on Radonež radio (an orthodox radio station), and the topic is religion. Thus, although the discourse is spoken, the register is not colloquial. The same is true for example (42), which is taken from an interview with the philosopher and literary critic Michail Bachtin.

(41) «[protoierej Oleg Stenjaev, muž, 1956] *Vot voz'mite naprimer evangel'skich christian baptistov kažetsja / oni sostojalis' v Rossii kak / èè nekaja takaja konfessija / sostojalis'. No oni že* ***ikonoborcy****. A svjatye otcy učat čto netu / zlee eresi čem neželi ikonoborčeskaja*.» [Protoierej Oleg Stenjaev. Besedy na Evangelie ot Matfeja na radio «Radonezh» v prjamom èfire // http://tv.radonezh.ru/audioarhiv01/stenyev/, 2002]

'[protopope Oleg Stenjaev, male, 1956] Let's take, for example, the Evangelical Christian Baptists. It seems that they were formed in Russia as a particular confession. Well, they are **iconoclasts**. And the Church Fathers teach that there is no worse heresy than iconoclasm.'

[36] All the examples are retrieved from the RNC.

(42) «[Bachtin, muž] *Potomu čto u nas / naprimer / sčitajut / čto Blok byl čut' li ne ateistom. Vot. No drugie utverždajut / čto / net / Blok ateistom nikogda ne byl. Čto on* ***bogoborcem*** *byl / čto net takogo poèta v mire bol'šogo / kotoryj by ne byl by* ***bogoborcem*** *i kotoryj byl by čistym / da ešče estestvenno-naučnym ateistom*.» [M. M. Bachtin. Besedy s M. Bachtinym // V. D. Duvakin. Besedy s Michailom Michajlovičem Bachtinym. Moskva, 2002, 1973]

'[Bachtin, male] Because in Russia, for instance, they think that Blok was almost an atheist. But others say that, no, Blok never was an atheist. That he was a **theomachist**, that there is no such a big poet in the world that would not be a **theomachist** and that would be pure, and even a scientific atheist.'

Turning to the suffixless construction, here too we see that many of the compounds featured in the main subcorpus do not appear in the spoken subcorpus (194 out of 393). However, it seems that suffixless compounds are more often employed in colloquial contexts. Thus, for instance, we see that the compound *pravdoljubec* 'honest, lit. one who loves truth' occurs only once in the spoken subcorpus in a movie script (43), while the compound *pravdoljub* 'honest, lit. one who loves truth' is also used in colloquial contexts, such as informal domestic conversations (48).

(43) «[Feofan Grek, Nikolaj Sergeev, muž, 77, 1894] *A tebja-to kak zvat'?* [Kirill, Ivan Lapikov, muž, 49, 1922] *Kirillom*. [Feofan Grek, Nikolaj Sergeev, muž, 77, 1894] *Kirillom*… [slyšit krik na ploščadi / idet k dveri] *Čto že / pravoslavnye /* ***pravdoljubcy*** *da christiane? Dolgo budete mučit' zlodeja? Skoro vy èto končite? On už sem' raz mukami svoimi iskupil grechi vse!*» [Andrej Tarkovskij, Andron Michalkov-Končalovskij, Andrej Rublev, k/f (1971)]

'[Feofan Grek, Nikolaj Sergeev, male, 77, 1894] And what's your name? [Kirill, Ivan Lapikov, male, 49, 1922] Kirill. [Feofan Grek, Nikolaj Sergeev, male, 77, 1894] Kirill… [hears a scream on the square, goes to the door] So what, **honest** orthodox Christians? How long will you torment the villain? Will you finish soon? With his agony, he has already paid seven times for all his sins!'

(44) «[Aleksej, muž] *Zato on čestnyj. On vse vremja govorit / èto nečestno / èto nečestno / ja vot čestno delaju*. [Lida, žen] *Da / da / on* ***pravdoljub***.» [Razgovor za prosmotrom semejnych fotografij // Iz kollekcii NKRJA, 2008]

'[Aleksej, male] But he is honest. He always says, "this is not fair, this is not fair, I want to be honest". [Lida, female] Yes, yes, he's a **truth lover**.'

Concerning the remaining constructions, we can note that the construction in *-tel'* behaves similarly to *-ec*, as it is underrepresented in the spoken corpus, while the construction in *-nik* is slightly overrepresented in the spoken corpus, similarly to the suffixless construction. The other constructions do not show statistically significant results, which is mainly due to their low frequencies.

Considering the analyses carried out throughout this section and comparing the three plots (cf. Figure 24, Figure 25, and Figure 26), we can draw the following conclusions. The construction in *-ec* and the suffixless construction show the strongest opposition. While being similar in that they both give rise to compounds with less specialized meanings with respect to the other constructions, they behave differently especially in terms of diachronic and stylistic differentiation (as will be also demonstrated in Section 4.6.2): while the constructions in *-ec* is overrepresented in earlier texts by comparison to the suffixless construction and its current usage is mostly restricted to written texts, the suffixless construction is overrepresented in later texts and its usage is not restricted to written texts, but shows overrepresentation in the spoken subcorpus as well.

4.6. Competition at the word level: rival forms

In his paper about English word-formation, Bauer points out that competition takes place between word-formation patterns, but also between individual words (Bauer 2006: 181), and that micro-level competition (i.e. competition between lexeme pairs) seems to have an effect at the macro-level of derivational patterns (Bauer 2006: 196).

In the present section, I consider some cases of micro-level competition that I found in my database of Russian compound agent nouns to see whether this analysis can

provide some further explanations of the rivalry of word-formation constructions in Russian. Specifically, I examine 29 doublets, i.e. pairs of compounds sharing the same lexical bases, but belonging to different constructions. In general, the possible situations are the following:

a) the two compounds denote different referents (semantic differentiation, see Section 4.6.1);
b) the two compounds denote the same referent (diachronic or stylistic differentiation, see Section 4.6.2).

4.6.1. Semantic differentiation

In what follows, I examine nine pairs of compounds sharing the same lexical bases, but belonging to different constructions. Although the motivation for the survival of both forms in these cases must be found in semantic differentiation, different situations are possible.

In some cases, one construction denotes a human referent, while the other construction denotes an inanimate referent, as shown in examples (45) through (54). Example (45) shows the occurrence of the compounds *granatometčik* 'grenadier' and *granatomet* 'grenade launcher', the former denoting a Human Agent, the latter an Instrument.

(45) «***Granatometčik***$_{\text{HUM}}$ *uže sobiralsja nažat' na spusk, no v ètot moment emu za šivorot upal gorjaščij ugolek ot mangala, on dernulsja i instinktivno nažal na spuskovoj krjučok* ***granatometa***$_{\text{INAN}}$.» [Viktor Docenko. Tridcatogo uništožit'! (2000)]

'The **grenadier** was already going to pull the trigger, but at that moment a burning piece of coal from the brazier fell onto his collar, he twitched and instinctively pulled the trigger of the **grenade launcher**.'

In (46) a similar situation is shown: within the same sentence, we find both the Human Agent *zvezdoletčik* 'astronaut' and the instrument noun *zvezdolet* 'spaceship'.

(46) «*Kto-nibud' iz* ***zvezdoletčikov***$_{\text{HUM}}$*, čašče vsego Olla Dez, korotko pojasnjal temu stereofil'ma, i* ***zvezdolet***$_{\text{INAN}}$ *izčezal*.» [I. A. Efremov. Čas byka (1968-1969)]

'One of the **astronauts**, most often Olla Dez, briefly explained the content of the 3D film, and the **spaceship** disappeared.'

In both cases, however, the compound in *-čik* denoting a Human Agent seems to be derived from the suffixless compound denoting an Instrument: a grenadier is one who uses a grenade launcher, and an astronaut is one who uses a spaceship (cf. Kuznecov's dictionary, which supports this interpretation). If this interpretation is correct, then these cases are not to be considered as cases of rivalry, as one member of the pair is derived from the other. The figures in
Table **26** do not seem to confirm this reading for the pair *granatometčik–granatomet*, as the first attestation of *granatometčik* in the RNC precedes the first attestation of *granatomet*. However, this might well be mere coincidence due to the inevitable incompleteness of the corpus.

Examples (47) and (48) display the opposition of the compound *samoučka* 'self-taught person' and its rival *samoučitel'* 'manual for self-tuition'.

(47) «*Dlja togo čtoby dat' polnyj portret Krekšina, nado otmetit', čto on byl i astrologom-****samoučkoj***$_{\text{HUM}}$.» [Pavel Krotov. «Prekrasnych vymyslov pletja iskusno nit'…» // «Rodina», 2009]

'To give a full portrait of Krekšin, it has to be considered that he also was a **self-taught** astrologist.'

(48) «*Na šestnadcatiletie emu podarili gitaru i* ***samoučitel'***$_{\text{INAN}}$.» [Roman Senčin. Eltyševy (2008) // «Družba Narodov», 2009]

'For his sixteenth birthday, they gave him a guitar and a **manual for self-tuition**.'

A similar situation is displayed by the opposition *maslobojščik–maslobojka* 'oil mill', where the former denotes a human being ('oil miller') and the latter an Instrument ('butter churn'), as shown in examples (49) and (50).

(49) «*God vmeste s molodoj ženoj batračili oni u svistunovskogo* ***maslobojščika***$_{\text{HUM}}$*, a potom kupili korovu s lošad'ju i zažili svoim chozjajstvom*.» [Boris Možaev. Živoj (1964–1965)]

'He and his young wife worked as farm laborers for a year at the **oil miller**'s of Svistunovskij, and then they bought a cow and a horse and started to live on their own business.'

(50) «*Na kuchne tesnee, tam – pečka, stol obedennyj, gorka s posudoj, u poroga – odežda, da obuv', da vsjakaja snast' chozjajskaja, kotoroja nužna pod rukoj: podojnik, moločnye fljagi, separator,* ***maslobojka***$_{\text{INAN}}$.» [Boris Ekimov. Prosnetsja den'… (1997)]

'In the kitchen it is tight. There is an oven, a dining table, some shelves for the cutlery. On the doorstep there are clothes, shoes, and any home utensils that are needed at hand: the milk container, milk pails, the separator, and the **butter churn**.»

An interesting case is represented by the opposition *kitoboec–kitoboj*. Although *kitoboec* is normally used to denote the Instrument 'whaler' (51) and *kitoboj* to denote the Human Agent 'whale hunter' (52), we do find examples (though rare) in which *kitoboec* is used to denote human referents (53) and *kitoboj* to denote inanimate referents (54).

(51) «*Na nosu každogo* ***kitobojca***$_{\text{INAN}}$ *nachoditsja osobaja puška, streljajuščaja garpunom…*» [V. I. Kurskij. Kitobojnyj promysel v SSSR // «Nauka i žizn'», 1936]

'On the prow of every **whaler** there is a special cannon that shoots harpoons…'

(52) «*V gody moej junosti, kogda devuška vychodila zamuž za rybaka, a už tem bolee* ***kitoboja***$_{\text{HUM}}$*, podružki vzdychali: povezlo!*» [Tat'jana Poljak. Vyjti zamuž za rybaka (2003) // «Rybak primor'ja», 2003.01.30]

'When I was young, when a girl married a fisherman, let alone a **whale hunter**, her friends would sigh: lucky her!'

(53) «*Ach, čto tvorilos', kogda* ***kitobojcy***$_{\text{HUM}}$ *prichodili iz vos'mimesjačnogo rejsa! Ich golodnye ženy s krasnymi gubami i sinimi cvetami davili svoich morjakov v ob''jat'jach…*» [Roman Karcev. Maloj, suchoj i pisatel' (2000-2001)]

'Ahh, what happened when the **whale hunters** came back from an eight-month sailing! Their hungry wives with red lips and blue flowers clung to their seamen…'

(54) «*Ljubopytno, čto p'janstvo ej, točno kapitanu kakogo-nibud'* ***kitoboja***$_{\text{INAN}}$ *ili piratskogo korablja, niskol'ko ne prepjatstvovalo deržat' rul' krepkoju rukoju.*» [A. V. Amfiteatrov. Mar'ja Lus'eva (1903)]

'Interestingly, drinking for her was as easy as for the captain of some **whaler** or pirate ship to hold the rudder.'

Another possible situation is that one construction denotes a human referent, while the other construction denotes an animal. In example (55) the compound *bogomolec* denotes the human referent 'pilgrim', while *bogomol* in (56) denotes the animal 'praying mantis'.

(55) «*Pritok* ***bogomol'cev***$_{\text{HUM}}$ *postepenno uveličivalsja, i prežnij chram stanovilsja tesen*.» [Christorožděstvenskij ženskij mosnastyr' v Tveri (2004) // «Žurnal moskovskoj patriarchii», 2004.02.23]

'The flow of **pilgrims** gradually grew and the old cathedral became too narrow.'

(56) «*Nenavižu nasekomych. Tol'ko* ***bogomoly***$_{\text{AN}}$ *nravjatsja. Ne znaju počemu*.» [Tat'jana Solomatina. Otojti v storonu i posmotret' (2011)]

'I hate insects. I only like the **praying mantis**. I don't know why.'

However, in earlier texts *bogomol* was also used to denote human referents, as exemplified in (57).

(57) «*Čelovek on – chorošij,* ***bogomol***$_{\text{HUM}}$*, myslej strogich i gramoten, nu, a vorovat' – ljubit!*» [Maksim Gor'kij. V ljudjach (1915-1916)]

‘He’s a good person, he’s a **devotee** with a strict moral code and he’s a literate person, but he likes stealing!’

In examples (58) through (60) I show the opposition of the two compounds *posudomojka* and *posudomoj*. While the former can denote both human and inanimate referents, ‘kitchen porter’ (58) and ‘dishwasher’ (59), the latter can only denote human referents (60).

(58) «*Masha edina vo mnogich licach: ona v odinočku vypolnjaet nelegkie objazannosti povara, uborščicy,* ***posudomojki***$_{\text{HUM}}$*, oficiantki, kasteljanši, prački*.» [Bachyt Kenžeev. Iz Knigi sčast’ja (2007) // «Novyj Mir», 2008]

‘Masha is one and many persons: she alone serves as cook, cleaning lady, **kitchen maid**, waitress, linen-keeper, laundress.’

(59) «***Posudomojki***$_{\text{INAN}}$*, stiral’nye mašiny i pylesosy pridumali mužčiny, kotorym suprugi postojanno zakatyvali skandaly na temu: «Ja vsju krasotu poterjala, pytajas’ navesti čistotu i porjadok».*» [Dar’ja Doncova. Mikstura ot kosoglazija (2003)]

‘**Dishwashers**, washing machines and vacuum cleaners were invented by men whose wives constantly made scenes like «I lost all my beauty trying to bring cleanness and order».’

(60) «*Mimo golovy i pleča Pavla vidno Šuchovu: dve ruki povara postavili dve miski v okošečke i, deržas’ za nich, ostanovilis’, kak by v razdum’e. Dolžno, on povernulsja i* ***posudomoev***$_{\text{HUM}}$ *rugaet*.» [Aleksandr Solženicyn. Odin den’ Ivana Denisoviča (1961)]

‘By Pavel’s head and shoulders Šuchov could see that the hands of a cook put two bowls on the window and, holding them, they stopped, as if they had changed their mind. Then he turned around and scolded the **kitchen porters**.’[37]

[37] N.B. that the compound *posudomoj* is a *hapax legomenon*.

The data also include a triplet, where two compounds denote a human referent, while a third compound denotes an inanimate referent. The compounds *dušegubec* (61) and *dušegub* (62) have the meaning of 'torturer, murderer' (the difference between the two must be found in diachronic and stylistic differentiation, see Section 4.6.2), while the compound *dušegubka* (63) has the meaning of 'mobile gas chamber' (as well as the metaphorical meaning of a place with no air).

(61) «*Koroče govorja, na drugoj den' utrom s Rajmondoj slučilsja obširnyj infarkt legkogo – a večerom javilsja* ***dušegubec***$_{\text{HUM}}$ *Fedja*.» [Marina Palej. Kabirija s Obvodnogo kanala (1990)]

'In brief, the next morning Rajmonda had a pulmonary embolism and in the evening the **murderer** Fedja appeared.'

(62) «*Ne* ***dušegub***$_{\text{HUM}}$ *on, Aleša, ja v ètom ponimaju*.» [Aleksej Motorov. Prestuplenie doktora Parovozova (2013)]

'He's not a **murderer**, Aleša, I know that.'

(63) «*Ostavšiesja na svobode proklinali potom ètu svobodu vsju žizn'. V èti* ***dušegubki***$_{\text{INAN}}$ *vpuskali vychlopnye gazy. I poka mašina doezžala do okrainy goroda – ljudi v nej zadychalis'*.» [Ljudmila Gurčenko. Aplodismenty (1994-2003)]

'Those who remained free then cursed that freedom for the rest of their lives. In those **mobile gas chambers**, they let in exhaust fumes. And while the car was reaching the city border, people suffocated in there.'

However, in earlier texts *dušegubka* was also used with the meaning of 'murderer' (64).

(64) «*Ona choš' i p'janica, a tol'ko ne* ***dušegubka***$_{\text{HUM}}$.» [I. D. Putilin. 40 let sredi grabitelej i ubijc (1889)]

'She might be a drunkard, but not a **murderer**.'

A different situation is exemplified by the opposition *raketonosec–raketonositel'*. Both compounds denote an inanimate referent, but while *raketonosec* refers to a ship or an

aircraft that carries rockets (65), *raketonositel'* (sometimes also written as *raketa-nositel'*) denotes a rocket used to carry a payload from Earth's surface into outer space (66).

(65) «*No esli zakupku strategičeskich* ***raketonoscev***$_{\text{INAN}}$ *i atomnych podlodok sokratit' nel'zja – ich v ljubom slučae nado dostraivat'...*» [Sergej Sarkisjanc. Vojna budet èkonomnoj // «Èkspert», 2015]

'But if it is impossible to reduce the purchase of strategic **rocket carriers** and atomic submarines, they have to be completed in any case…'

(66) «*V 4 časa 44 minuty po moskovskomu vremeni rossijskij kosmičeskij modul' «Zvezda», dostavlennyj na okolozemnuju orbitu* ***raketonositelem***$_{\text{INAN}}$ *«Proton» v avtomatičeskom režime, sostykovalsja s ranee sobrannym na orbite blokom iz dvuch modulej – rossijskogo «Zarja» i amerikanskogo Unity*.» [Boris Javelov. Chronograf (2004) // «Vokrug Sveta», 2004.07.15]

'At 4:44 AM (Moscow time) the Russian Space Station Module "Zvezda", reached in near-Earth orbit by the **carrier rocket** "Proton" in automatic mode, joined the block formed by the two Station Modules - the Russian "Zarja" and the American "Unity" - that had already reached the orbit before.'

The compound *raketonositel'* seems to have a different structure with respect to the other compounds included in the analysis. Indeed, this compound should not be interpreted as 'sth. that carries rockets', but rather as 'a rocket that carries sth.'. Thus, the first element (*raket-*) must be understood as the head of an appositional compound, where the verbal element in second position specifies its property of carrying something. If we take for granted this interpretation, then we should not consider the pair *raketonosec–raketonositel'* as an example of rivalry, as the two words are the result of different word-formation processes.

Table **26** summarizes the results discussed so far and regarding the nine cases of semantic differentiation. The controversial cases discussed are highlighted in grey.

Table 26. Rival forms showing semantic differentiation and corresponding first occurrences in the RNC

	Meaning	**I Occ.**		**Meaning**	**I Occ.**
granatometčik	grenadier [+hum]	1930	*granatomet*	grenade launcher [-an]	1933
zvezdoletčik	astronaut [+hum]	1961	*zvezdolet*	spaceship [-an]	1935
samoučka	self-taught person [+hum]	1782	*samoučitel'*	manual for self-tuition [-an]	1833
maslobojščik	oil miller [+hum]	1863-1864	*maslobojka*	oil mill [-an]	1863
kitoboec	whaler [-an] (whale hunter [+hum])	1936	*kitoboj*	whale hunter [+hum] (whaler [-an])	1855
bogomolec	pilgrim [+hum]	1682-1709	*bogomol*	mantis [+an] (devotee [+hum])	1741-1752
posudomojka	kitchen porter [+hum] dishwasher [-an]	1961	*posudomoj*	kitchen porter [+hum]	1961
dušegubec/ dušegub	murderer [+hum]	1766/ 1833-1834	*dušegubka*	mobile gas chamber [-an] (murderer [+hum])	1833
raketonosec	rocket carrier [-an]	1970	*raketonositel'*	carrier rocket [-an]	1958

4.6.2. Diachronic or stylistic differentiation

In this section, I consider 20 pairs of rival compounds not showing semantic differentiation, but rather diachronic or stylistic differentiation.

Of these 20 pairs, 17 show the opposition *-ec*/*-ø*, as displayed in Table 27. These compounds, all denoting human referents, are nicely described in Tagabileva (2013), where the author suggests that the suffixless construction seems to be ousting the older construction in *-ec*, but several factors prevent the complete extrusion of the construction in *-ec*. A first important factor is constituted by morphonological constraints on the suffixless construction, which only forms compounds on the basis of one-syllable

unprefixed imperfective verb stems,[38] and shows some restrictions also concerning the first stem, which is rarely constituted by more than two syllables. Second, stylistic differentiation also plays an important role in the survival of the model in *-ec*, as the archaic forms in *-ec* are still used in written texts and they acquire a high stylistic value that the neutral suffixless construction does not have.

Table 27. Rival forms in *-ec* and *-ø* showing diachronic and stylistic differentiation and corresponding token frequencies and first occurrences in the RNC

	Tokens	**I Occ.**		**Tokens**	**I Occ.**	**Meaning**
čelovekoljubec	87	1717	*čelovekoljub*	9	1846-1847	philanthropist
čestoljubec	329	1766-1794	*čestoljub*	2	1894	ambitious
duchoborec	90	1824	*duchobor*	283	1890-1899	lit. spirit warrior
krjučkotvorec	29	1772	*krjučkotvor*	35	1788-1822	pettifogger
morechodec	167	1766-1894	*morechod*	370	1830-1837	seaman
odnodumec[39]	3	2001	*odnodum*	15	1887-1894	opinion ally
pravdoljubec	112	1808-1822	*pravdoljub*	48	1847	truth lover
samoljubec	25	1765	*samoljub*	8	1742	self-lover
sebjaljubec	75	1829	*sebjaljub*	3	1985-1995	self-lover
serdcevedec	80	1757	*serdceved*	49	1832	God
slastoljubec	147	1764	*slastoljub*	1	1786-1787	voluptuous
trudoljubec	30	1760-1775	*trudoljub*	68	1789	hard-working
verchovodec	1	1843-1847	*verchovod*	34	1862	leader
vodonosec	6	1826	*vodonos*	107	1792	water carrier
vol'nodumec	398	1779-1790	*vol'nodum*	13	1784-1792	free thinker
ženoljubec	22	1750	*ženoljub*	37	1882	ladies' man
žizneljubec	20	1877	*žizneljub*	81	1906-1913	life lover

[38] This, however, is not always true, as demonstrated by examples featured in my database, such as *vodovypusk* 'culvert' and *gromootvod* 'lightning rod', whose base verbs are prefixed (*vy-puskat'* 'let out' and *ot-vodit'* 'reject').
[39] See Tagabileva (2013: 204) for discussion.

The remaining three pairs show the opposition of different constructions. The compounds *vozduchoduvka* (67) and *vozduchoduv* (68) 'compressed air releaser' show rather different frequencies (21 tokens vs. 6 tokens respectively), and *vozduchoduvka* is older than *vozduchoduv* (first occurrences in the RNC: 1896 vs. 1933).

(67) «*V otličie ot rasprostranennych v promyšlennosti centrobežnych* ***vozduchoduvok***$_{\text{INAN}}$*, èto ustrojstvo osnovyvalos' na principe osevogo kompressora*.» [D. A. Sobolev. Russkoe zarubež'e // «Aviacija i kosmonavtika», «Aviacionnyj sbornik», 1994]

'Differently from the centrifugal **compressed air releasers** that are commonly used in industries, this device was based on the principle of an axial compressor.'

(68) «***Vozduchoduv***$_{\text{INAN}}$ *oni vključili, – skazal Samodelkin, razgljadyvaja pul't upravlenija raketoj*.» [Valentin Postnikov. Priključenija Karandaša i Samodelkina na «Dryndolete» (1997)]

'They turned on the **compressed air releaser**, – Samodelkin said, examining the control panel of the rocket.'

Although one might think that diachronic differentiation is taking place here, a quick Google search shows that these compounds are both currently used (apparently interchangeably) to designate leaf blowers. A similar situation is exemplified by the opposition *samosejka* (69) vs. *samosev* (70) 'self-sown plant'.

(69) «*Na meste stojavšej zdes' chaty tjanulas' iz sornjakov k svetu kolječaja gruša-dička – možet, nepotrebnyj otprysk nekogda rosšich zdes' gruš-spasovok, a možet, slučajnaja* ***samosejka***$_{\text{INAN}}$*, zanesennaja iz lesa pticami*.» [Vasil' Bykov. Znak bedy (1982)]

'Instead of the hut that used to be here, a thorny wild pear stretched toward the light from the weeds, maybe a worthless offspring of the pear "Spasovka" that never grew here, or maybe a casual **self-sown plant**, brought from the forest by the birds."

(70) «*Oni sobirajutsja na sornjakach iz semejstva paslenovych, na* ***samosevach***$_{\text{INAN}}$ *kartofelja, kartofel'nych burtach, no čašče vsego na rannich posadkach kartofelja*. [kollektivnyj. Naši konsul'tacii // «Chimija i žizn'», 1968]

'They (beetles) gather on the weeds of the Solanaceae (or nightshades), on potato **sprouts** and crusts, but most often on young potato plants.'

The compound *samosev* is much more frequent (83 tokens in the RNC) and older (first occurrence in the RNC: 1867) than *samosejka* (5 tokens, first occurrence in the RNC: 1982), but its high frequency might be due to the fact that this item is polysemous and can also mean 'natural sowing', thus indicating a process and not a self-sown plant.

Finally, let us consider the opposition between the compounds *bumagomaratel'* (71) and *bumagomaraka* (72) 'scribbler', both of which are negatively connotated.

(71) «*Ved' èto strašno podumat', čem dolžen byt' pisatel'… esli on ne chočet byt', konečno, tol'ko* ***bumagomaratelem***$_{\text{HUM}}$.» [N. G. Garin-Michajlovskij. Studenty (1895)]

'Because it's scary to think of what a writer must be… of course, if he doesn't want to be only a **scribbler**.»

(72) «*V samom dele, počemu Merkurij Avdeevič dolžen byl by uklonit'sja ot peredači o podrobnostjach dosadivšego emu razgovora s ljud'mi raspuščennych nravov, kakimi-to licedejami i* ***bumagomarakami***$_{\text{HUM}}$?» [K. A. Fedin. Pervye radosti (1943-1945)]

'Indeed, why sould Merkurij Avdeevič avoid telling the details of the conversation with those people of loose moral, those hypocrites and **scribblers**?'

The compound *bumagomaratel'* is older (first occurrence in the RNC: 1769) than *bumagomaraka* (first occurrence in the RNC: 1836). Both compounds show low frequencies (17 tokens for *bumagomaratel'* and 24 for *bumagomaraka* in the RNC).

The main difference between the rival forms in the last three pairs of compounds must be found in stylistic differentiation: the construction in *-ka* is often characterized by

greater expressiveness and colloquiality. The figures corresponding to the last three pairs are summarized in Table 28.

Table 28. Rival forms showing diachronic and stylistic differentiation (or polysemy of one of the members) and corresponding token frequencies and first occurrences in the RNC

	Tokens	**I Occ.**		**Tokens**	**I Occ.**	**Meaning**
vozduchoduvka	21	1896	*vozduchoduv*	6	1933	compressed air releaser
samosejka	5	1982	*samosev*	83	1867	self-sown plant
bumagomaratel'	17	1769	*bumagomaraka*	24	1836	scribbler

Thus, the analysis of the 29 cases of micro-level competition has led to the following conclusions:

i. in 31% of the cases (9 out of 29 pairs), semantic differentiation is taking place. The two constructions denote different referents, although in some cases the two forms seem to admit a semantic overlap (45) – (66);

ii. in 58.6% of the cases (17 out of 29 pairs), rivalry takes place between the construction in *-ec* and the suffixless construction in denoting human referents. The distribution of the two constructions here is due to several factors, i.e. diachronic differentiation, morphonological constraints and stylistic differentiation (cf. Tagabileva 2013);

iii. in 10.3% of the cases (3 out of 29 pairs), it is much more difficult to motivate the existence of two rival forms. Indeed, the two constructions seem to denote the same referents and the causes of their rivalry must be found either in the polysemy of one of the two compounds or in stylistic differentiation (67) – (72);

iv. in 82.8% of the cases (24 out of 29 pairs), one of the members of the pair belongs to the suffixless construction, which might lead us to wonder whether this construction is slowly ousting its rivals in the creation of agent nouns;

v. the constructions involved in cases of rivalry are the following: *-ec*, *-tel'* (only three pairs), *-ščik/čik* (only two pairs), *-ka* and the suffixless construction. The fact that the most common situation in cases of rivalry regards the suffix *-ec* and the suffixless construction (and, to a certain extent, the suffix *-ka*) might be seen as a further indicator of the fact that

these constructions are more neutral and less specialized with respect to the other constructions, as discussed in Section 4.5.2.

4.7. Summary

In the present chapter, I have compared rival constructions forming compound agent nouns in Russian. The analysis shown in Section 4.3, based on a number of formal and semantic parameters, has allowed putting forward some observations that have been confirmed by the statistical analyses described in Section 4.4. The main conclusions arising from these observations can be summarized as follows:

i. the low-frequency constructions in *-lec*, *-l'ščik* and *-lka* show no variation and greater specificity with respect to the other constructions;

ii. the remaining high-frequency constructions show greater variation and lower specificity;

iii. the differences and similarities among rival constructions are captured to a greater or lesser extent depending on the parameter considered:

a) the rival word-formation constructions examined show lower variation in terms of first member's part of speech and semantic role, and in terms of transitivity of the verbal element, i.e. all constructions show similar tendencies, although the construction in *-ec* and the suffixless construction show, on average, greater variation;
b) the remaining parameters better highlight the different behaviors of rival constructions: the construction in *-tel'* stands out with respect to all the others for its capability of forming compounds based on perfective verbs; the parameters "Animacy" and "Semantics" show that the constructions in *-ec* and *-ø* are more strictly associated with meanings of Human Agents, while the other constructions are also often used to denote Instruments; the construction in *-ec* stands out with respect to all the others for its greater capability of giving rise to nouns denoting Carriers of State.

The analysis described in Section 4.5 has shown that rival constructions differ significantly also in terms of their distribution over time and across textual genres and registers. In general, this analysis has brought to the conclusion that the two constructions showing a stronger opposition are the construction in *-ec* and the suffixless construction, which are more similar to one another in terms of the semantics of the compounds they form, but differ significantly in terms of diachronic distribution (*-ec* being more frequent than *-ø* in earlier texts, but less frequent than *-ø* in recent texts) and of register (*-ec* being more typical of written texts and thus belonging to a higher register compared to the suffixless construction). Remarkably, their underrepresentation in technical and official texts confirms their less specialized (and more polysemous) semantics with respect to the other constructions.

The analysis of 29 pairs of rival compounds in Section 4.6 has provided further evidence of the observations already put forward, showing that, in most cases of rivalry in which semantic differentiation is not at play, the opposition regards the construction in *-ec* and the suffixless construction.

5. The productivity of rival word-formation constructions

The present chapter addresses the controversial question of productivity in word-formation. Productivity is one of the crucial issues that must be dealt with when comparing rival word-formation constructions: "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). Thus, the comparison of rival constructions giving rise to compound agent nouns in Russian raises the following questions: Do rival constructions behave differently as regards their degree of productivity? Which construction is the most productive? Which construction is the least productive?

Although productivity in word-formation is a well-studied topic, it is still not clear what is the best way to address this issue and many diverging approaches exist in the literature. Section 5.1 is devoted to the definitions and measurements of productivity that have been proposed so far. In Section 5.2, I consider the main problems that arise when measuring productivity. In Section 5.3, I show how calculations of productivity significantly depend on the nature of the data available and on the corpora employed by looking at three low-frequency constructions in the ruTenTen Corpus. In Sections 5.4 and 5.5, I address the questions of productivity over time, as well as across genres and registers. In Section 5.6, I consider possible restrictions affecting the productivity of rival constructions. In Section 5.7, I carry out two case studies comparing partially specified schemas to test the differences in productivity of the constructions in *-ec* and *-ø*. Finally, in Section 5.8, I briefly discuss the relation between productivity and polysemy.

5.1. Defining and measuring productivity

Most linguistic research dealing with word-formation ends up being concerned with the question of productivity. Thus, when consulting the literature about a certain word-

formation process in a certain language, we often get to a point in which the productivity of that process is discussed.

If we take, for example, the section of the AG-80 devoted to word-formation (§§ 191–1110), we see that the word-formation processes and formants are often described in terms of their productivity in modern Russian. With respect to compound agent nouns, the authors claim that:

a) the suffix *-tel'* is productive in technical/scientific terminology (§ 558);
b) the suffix *-ec* exhibits a limited productivity, i.e. only when the deverbal noun exists outside compounding (§ 559);
c) the suffixes *-nik*, *-ščik/čik*, and *-l'ščik* are only productive in technical terminology (§§ 560, 561, 562);
d) the suffixes *-ka* and *-lka* are mainly productive in technical discourse and everyday language (§§ 563, 564).

Nothing is said about the productivity of the suffixless construction (cf. § 579). Moreover, it is not quite clear on what basis it is claimed that certain processes or formants are productive in modern Russian, whereas others are not. Without any explanation being given, one might suppose that the criteria on which the determination of the productivity of certain word-formation constructions is based solely rely on native speaker's intuitions.

Recent studies on word-formation are, instead, increasingly concerned with establishing more precise criteria on which definitions of productivity should be based. In general, productivity is usually intended as the ability of an affix or of a morphological process to be actively used to produce new words (cf., for instance, Bauer 1983: 18; Plag 2003: 44; Booij 2005: 68; Haspelmath & Sims 2010: 114). However, possible approaches to this issue are quite varied and sometimes they significantly diverge one from the other (see Bauer 2005 for an overview of different productivity theories).

Aronoff (1976) intends productivity as the relationship between possible words and actual words, and sees the productivity of a certain affix as related to the type of bases to which it can be added. The main drawback of this approach is that we cannot really determine when a possible word becomes an actual word. Indeed, dictionaries cannot be used as evidence of the fact that a possible word has become actual, because many actual words are not listed in dictionaries. In fact, dictionaries can rather prove the past productivity of some word-formation process, and not their actual productivity.

Corbin (1987) draws an important distinction between two possible ways of intending productivity, which lies at the basis of the distinction between qualitative and quantitative approaches to productivity. The author distinguishes the notions of *disponibilité* 'availability' and *rentabilité* 'profitability'. In a qualitative approach, productivity is intended as a yes/no question: a word-formation process is either available or not for the creation of new words. In a quantitative approach, morphological productivity, i.e. profitability, is understood instead as the extent, the degree to which a certain morphological process can be employed to create new words. The latter notion of productivity is more focused on the actual use of a certain morphological process and aims at determining the exact value corresponding to the productivity of the morphological process considered during a certain time span. However, the two notions of productivity can also be understood as strictly interrelated concepts: "quantitative and qualitative notions of productivity (…) are closely related. Thus the idea of potentiality, which is central to qualitative definitions of productivity, can be expressed in the statistical terms of probability" (Plag 1999: 22).

Quantitative approaches to productivity strongly rely on the notions of type and token frequencies. Type frequency "is concerned with the number of items in the language that contain the item or process under consideration" (Bauer 2001a: 47). Thus, if we consider the process giving rise to compound agent nouns with the suffix *-tel'* in Russian, the type frequency of this word-formation process in the RNC will equal the number of different compounds formed with this suffix in the corpus. By contrast, the number of occurrences of each compound in the corpus corresponds to its token frequency.

The most widespread corpus-based quantitative approach to morphological productivity is represented by the works of Baayen and his associates (Baayen 1992, 1993, 2001; Baayen & Lieber 1991), who have elaborated on a number of corpus-based statistical measures of productivity.

One of the proposed measures is type frequency (V), i.e. the number of types formed with a given affix[40] in a corpus of N tokens. However, as pointed out by Bauer (2001a: 144), type frequency can only indicate that a given affix has been productive in

[40] Although the measures of productivity proposed by Baayen usually refer to the productivity of affixes, it would be better to talk more generically of word-formation processes or constructions, which would also include cases in which word-formation is not affix-based.

the past, but it does not tell us anything about the availability of that process in the language.

A second measure is "potential productivity" or "productivity in the narrow sense", which is calculated by dividing the number of *hapax legomena* (V1) with a given affix by the number of tokens (N) with that affix, as shown in (1).

(1) P = V1/N

According to Baayen & Lieber (1991: 809–810), this measure "estimates the probability of coming across new, unobserved types, given that the size of the sample of relevant observed types equals N". As I will discuss later (cf. Section 5.2), this measure has been subject to several criticisms.

A third measure proposed by Baayen & Lieber (1991: 819) and Baayen (1992: 124, 1993: 190) is "global productivity" (P*), i.e. a function of P and V, represented through a two-dimensional graph with the degree of productivity P on the horizontal axis and the value V on the vertical axis. Thus, productive affixes will show large values for P and V, while unproductive affixes will show low values for P and V. However, as pointed out both by Baayen himself (1992: 124) and Bauer (2001a: 154), it is difficult to rank different affixes in terms of global productivity due to their disparate positions on the graph.

Finally, another measure proposed by Baayen (1993: 192) is the "hapax-conditioned degree of productivity" (P*), which is calculated by dividing the number of *hapax legomena* with a given affix by the total number of *hapax legomena* in the corpus. However, as Bauer (2001a: 155) points out, "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."

Considering that productivity is about the creation of new words, the correlation between *hapax legomena* and productivity has often been criticized in view of the fact that *hapax legomena* are not necessarily new formations. However, as Gaeta & Ricca (2015: 847) suggest, "for corpora of many tenths of millions tokens, most hapaxes indeed turn out to be un-established words", and if the corpus data are manually checked, *hapax*

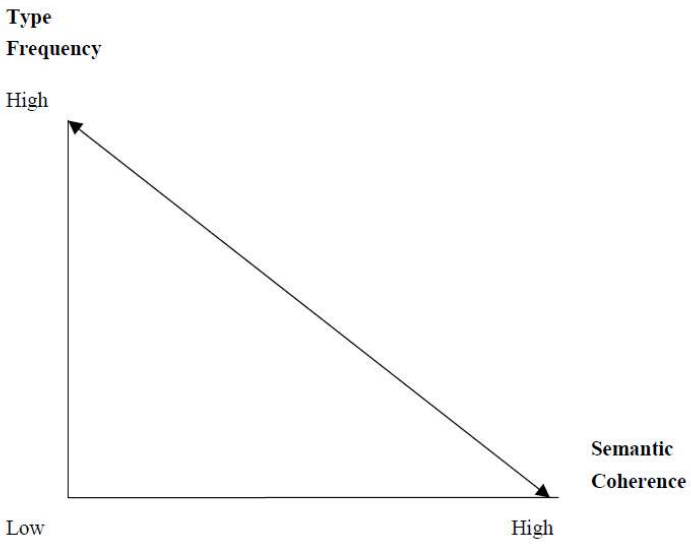
legomena can be considered as a reliable indicator of productivity (see also Haspelmath & Sims 2010: 130; Plag, Dalton-Puffer & Baayen 1999).

In recent studies about productivity in word-formation, the quantitative corpus-based approach adopted by Baayen and the measure of potential productivity have been extensively employed to determine the value of productivity of certain word-formation processes in different languages (among others, see Gaeta & Ricca 2006; Štichauer 2009, 2015; Efthymiou, Fragaki & Markos 2012; Naccarato 2016). Nonetheless, this approach to productivity also entails some problematic aspects that need to be dealt with, as I will show in Section 5.2.

In their investigations about productivity in word-formation, Baayen (1993), Hay (2001) and Hay & Baayen (2002) attribute a central role to parsing, i.e. transparency and decompositionality. As the authors point out, while some words are highly decomposable, others are more opaque and are characterized by whole-word access, rather than by parsing (Hay & Baayen 2002). When we have high rates of parsing, we normally find increased rates of productivity (Hay & Baayen 2002). Baayen (1993) proposes the productivity measure A, which corresponds to "the number of words containing the affix which occur below a certain frequency threshold, each weighted by their frequency of occurrence" (Hay & Baayen 2002: 204). The main drawback of the measure A consists in the fact that it relies on the assumption that low-frequency forms require parsing, while high-frequency forms do not. This assumption is also in line with other previous studies arguing that high frequency is correlated with low decomposability and low semantic transparency (cf., among others, Modor 1992; Baayen 1992, 1994b; Baayen et al. 1997; Bybee 1988, 1995). However, as Hay & Baayen (2002: 204) suggest, "the absolute frequency of a derived form is not straightforwardly related to parsing. More relevant is the nature of the frequency relations between the derived form and the base". Hay (2001) describes the results of an experiment in which subjects were presented with pairs of simple/prefixed (e.g. *patient* vs. *impatient*) and simple/suffixed (e.g. *general* vs. *generally*) words, and notes that derived forms that are more frequent than the base they contain are more easily subject to whole-word access, while derived forms that are less frequent than the base they contain are more easily subject to parsing (Hay 2001: 1049–1050).

Within a constructionist approach, Barðdal (2008) conceives productivity as the function of type frequency and semantic coherence and the inverse correlation between the two (Barðdal 2008: 34). This correlation is represented as shown in Figure 27.

Figure 27. The inverse correlation between type frequency and semantic coherence (Barðdal 2008: 35)



This conception of productivity implies that "the higher the type frequency of a construction, the lower the degree of semantic coherence[41] is needed for a construction to be productive. Conversely, the lower the type frequency of a construction, the higher degree of semantic coherence is needed for a construction to be extendable" (Barðdal 2008: 34).

5.2. Problems in measuring productivity

As mentioned previously, one of the most widespread approaches to productivity in word-formation is the one based on Baayen's corpus-based measures. In particular, most studies aiming at determining the productivity of certain word-formation processes employ

[41] In morphology, the notion of semantic coherence is usually employed to refer to cases in which the meaning of a derivative is transparent and compositional, i.e. it is the result of the meaning of the base and that of the affix (cf. Aronoff 1976; Bauer 1983).

Baayen's measure of potential productivity, as shown in (1) (cf. Gaeta & Ricca 2006; Štichauer 2009, 2015; Efthymiou, Fragaki & Markos 2012; Naccarato 2016). However, if we apply this measure to determine the productivity of the rival constructions forming compound agent nouns in Russian based on data from the RNC word-formation database used throughout this study, we face some problems that are partly related to the measure itself, and partly to the nature of the data.

A first problem is related to the measure of potential productivity proposed by Baayen. Indeed, since the value of P is highly dependent on the sample size (P is a function of N), this measure seems inadequate to compare differently-sized corpora (cf. Lüdeling, Evert & Heid 2000; Evert & Lüdeling 2001; Gaeta & Ricca 2006; Štichauer 2009; Efthymiou, Fragaki & Markos 2012; Naccarato 2016). Let us assume, for instance, that we want to determine how the productivity of rival word-formation constructions changes over time. To do so, we would need to compare texts belonging to different time periods. In a corpus such as the RNC, this would result problematic. Indeed, when creating different subcorpora for each century represented in the corpus, we get incomparable results: the amount of texts belonging to the 18th century, for example, does not equal the amount of texts belonging to the 19th century, or of those belonging to the 20th century. Still, this problem could be solved in different ways. One solution would be to divide the larger subcorpora into smaller pieces comparable to the shortest one. Another solution would consist in resorting to parametric statistical models of frequency distribution known as LNRE (Large Number of Rare Events) models (Baayen 2001), as already done in previous studies (Štichauer 2009, 2015; Efthymiou, Fragaki & Markos 2012; Naccarato 2016).

A second (and more crucial) problem is related to the nature of the data on which the study is based. Since the measure of potential productivity is based on the number of *hapax legomena* of a word-formation process in a corpus, it does not seem suitable to all types of data. In this case, the data at my disposal do not meet the requirements that would allow applying this measure. Remarkably, although the compounds included in the database used throughout this study are represented in the RNC and the analysis relies on corpus data, the compounds were not extracted from the corpus following a bottom-up approach, because the RNC is not yet provided with a search system that allows extracting all the compound words in the corpus. As previously discussed, the data used throughout

this study were retrieved from the RNC word-formation database, which was used as a basis for starting the tagging process of the compounds in the corpus. Since the database consists of dictionary data, we will not find high numbers of *hapax legomena*, if we find them at all. Indeed, as previously mentioned, dictionaries cannot give us any evidence of the actual productivity of word-formation processes. Rather, they can prove that a certain word-formation process was productive in the past.

For each construction included in my database, I checked in the RNC the number of types, tokens and *hapax legomena*. The corresponding figures are shown in Table 29.

Table 29. Type frequencies, token frequencies and *hapax legomena* of rival constructions (word-formation database of the RNC)

Construction	Type frequency	Token frequency	*Hapax legomena*
-ec	143	31,278	5
-lec	7	2,957	0
-tel'	175	34,439	2
-nik	28	3,064	0
-ščik/čik	23	734	1
-l'ščik	4	1,373	1
-ka	47	5,917	2
-lka	11	608	1
-ø	393	110,288	21

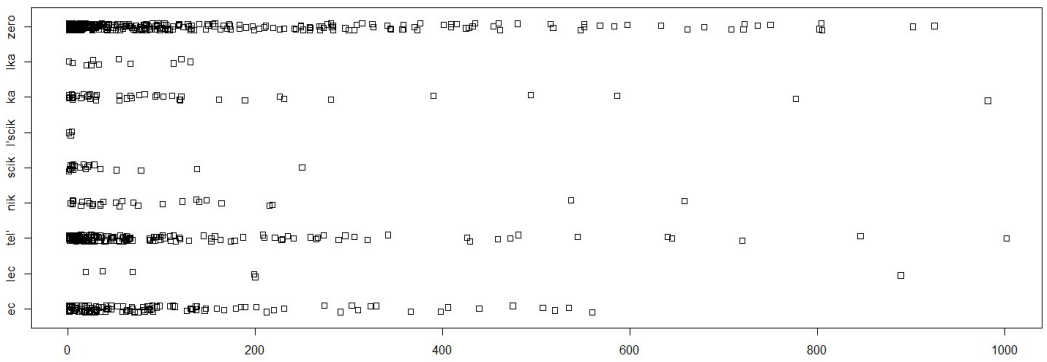
As expected, very few *hapax legomena* were retrieved and, in some cases (i.e. for the constructions in *-lec* and *-nik*), no hapaxes at all were found. If we apply the potential productivity measure shown in (1) to these data, we get the results shown in the last column of Table 30, which indicate very low degrees of productivity, or even no productivity at all. These results cannot be taken as reliable for two main reasons. First, the constructions differ significantly in terms of tokens and the measure of potential productivity, which is a function of the token number, cannot be applied when the token difference is so big. Second, and more important, the data at our disposal include very few *hapax legomena* (as the result of the data retrieval process already discussed), which would give us a wrong impression of the productivity of the constructions under examination in Russian.

Table 30. Token frequencies, hapax legomena and potential productivity P of rival constructions (word-formation database of the RNC)

Construction	Token frequency	*Hapax legomena*	P
-ec	31,278	5	0.00016
-lec	2,957	0	0
-tel'	34,439	2	0.00006
-nik	3,064	0	0
-ščik/čik	734	1	0.00136
-l'ščik	1,373	1	0.00073
-ka	5,917	2	0.00034
-lka	608	1	0.00164
-ø	110,288	21	0.00019

If we consider the frequency distributions of the compounds in the RNC, as shown in Figure 28, we see the typical patterns of productive constructions, i.e. low numbers of high-frequency words and high numbers of low-frequency words (cf. Plag 2002: 292, 2006: 548). The plot, created with the function stripchart() of the R package "ggplot2", shows the frequencies of the compounds selected for the analysis (each box in the plot represents a compound), and includes compounds showing a maximum of 1,000 occurrences in the corpus. The frequency patterns are better visible for the most frequent constructions (i.e. *-ec*, *-tel'* and the suffixless construction), which show a high concentration of low-frequency compounds and a low concentration of high-frequency compounds.

Figure 28. Token frequencies of the compounds in the RNC: strip chart



Thus, the low numbers of *hapax legomena* for the constructions under examination are not evidence of their low degrees of productivity, but instead confirm that the data at our disposal are unsuitable for the application of the potential productivity measure.

It must be clarified that a bottom-up approach to the retrieval of data from the RNC was excluded for two main reasons. First, the corpus is not yet provided with a thorough annotation of compounds, so it is not possible to carry out a search that would give us in output a complete list of the compounds contained in the corpus. Thus, the only way to obtain a full list of compounds formed with a certain suffix would be to carry out a search of the type **ka*, which would give us in output all the occurrences of words ending in *-ka*, from which compounds might be manually selected. However, in the main corpus this search corresponds to more than five million occurrences distributed over more than 100 thousand documents. The corpus interface allows visualizing a maximum of one hundred documents in each page and, at the bottom of each page, the lemma list for all the occurrences included in that page is given. This means that the only possible solution would be to choose the option producing one hundred documents per page and to look for compounds in the lemma list at the bottom of each page. However, for the search **ka*, which I have taken as an example, this would mean that we would have to check manually more than 1,000 pages (and this operation would have to be repeated for each suffix). A second problem, which is even more crucial, regards the suffixless construction, that, for obvious reasons, cannot be looked for in the corpus, as it does not show a fixed ending.

5.3. A bottom-up approach to productivity: the low-frequency constructions in the RNC and the ruTenTen Corpus

To test whether a bottom-up approach would produce significantly different results compared to the data retrieved from dictionaries, I conducted a case study involving the low-frequency constructions in *-lec*, *-l'ščik*, and *-lka*, which give rise to fewer compounds with respect to the other constructions and allow for feasible manual investigations in big corpora.

First, I searched for these suffixes in the main subcorpus of the RNC using the simple regular expressions **lec*, **l'ščik*, **lka*, which give in output all the occurrences of words ending in *-lec*, *-l'ščik*, and *-lka*, thus including not only compounds, but also simple (e.g. *palec* 'finger') and derived words (e.g. *vladelec* 'owner', from *vladet'* 'own, possess'). To select the relevant data, I had to proceed in the following way. For each construction, I selected the compounds from the lemma lists that are found at the end of each page (as already mentioned, the corpus interface allows visualizing a maximum of one hundred documents per page). For the suffix *-lec*, the search produced 128,539 occurrences contained in 21,134 documents, which means that I had to check 212 lemma lists (one every hundred documents). For the suffix *-l'ščik*, the search produced 17,966 occurrences and 6,220 documents, and consequently 63 lemma lists to be checked. Finally, for the suffix *-lka*, I found 241,024 occurrences and 31,801 documents, that is 319 lemma lists. The search yields the results shown in Table 31 along with the figures corresponding to the data contained in the RNC word-formation database used throughout this study.

Table 31. Type frequencies, token frequencies and *hapax legomena* for the constructions in *-lec*, *-l'ščik* and *-lka* in the RNC (word-formation database vs. bottom-up approach)

	Type frequency		**Token frequency**		***Hapax legomena***	
	Database	*Bottom-up*	*Database*	*Bottom-up*	*Database*	*Bottom-up*
-lec	7	51	2,957	3,758	0	17
-l'ščik	4	28	1,373	1,431	1	17
-lka	11	32	608	710	1	8

The results show that, by adopting a bottom-up approach, more types and more hapaxes are found, but only slightly higher token numbers. These results are clearly the consequence of the fact that compounds not included in the RNC word-formation database are probably not included in dictionaries because they represent innovations in the language or because they are not frequently used. Thus, the main difference with respect to the data contained in the RNC word-formation database is the presence of low-frequency items, often hapaxes (which is why we do not observe big differences in the

number of tokens retrieved), which is exactly what makes the difference when productivity is calculated.

Thus, if the measure of potential productivity shown in (1) is applied, this yields the results in the last column of Table 32, which are compared to the results corresponding to the data from the word-formation database of the RNC.

Table 32. Token frequencies, *hapax legomena* and productivity rates (P) for the constructions in *-lec*, *-l'ščik* and *-lka* in the RNC (word-formation database vs. bottom-up approach)

	Token frequency		***Hapax legomena***		**P**	
	Database	*Bottom-up*	*Database*	*Bottom-up*	*Database*	*Bottom-up*
-lec	2,957	3,758	0	17	0	0.00452
-l'ščik	1,373	1,431	1	17	0.00073	0.01188
-lka	608	710	1	8	0.00164	0.01127

Considering that the number of tokens retrieved changes only slightly, while the number of *hapax legomena* increases significantly within the bottom-up approach, the productivity rates obtained increase significantly. This demonstrates that the way in which data are retrieved affects substantially the results of productivity measurements and that dictionary data cannot tell us much about the synchronic productivity of a certain word-formation process.

Another factor playing a crucial role in the calculation of productivity regards the size of the corpus used. To demonstrate how the corpus size can impact calculations of productivity, I carried out another experiment based on data from a large corpus of modern Russian, i.e. the ruTenTen Corpus. This corpus, which is included in the Sketch Engine corpus tool (see https://www.sketchengine.co.uk/), is constituted by a collection of web texts for a total of almost fifteen billion words.[42]

In order to find all compounds formed with the relevant suffixes, I resorted to the function "Word list"[43] and employed the following regular expressions: .**l'ščik*, .**lec*, and .**lka*. Of course, these searches produced long lists of lemmas: simple words,

[42] By way of comparison, I recall that the RNC includes about 600 million words, and the main subcorpus used for this analysis includes about 200 million words (cf. Chapter 4).

[43] This function, which gives in output a list of lemmas with the required characteristics and the corresponding token frequencies in the corpus, is not available in the RNC.

derivatives and compounds. From the list obtained for each suffix, I selected only those compounds belonging to the model under examination, i.e. verb-based compounds, and obtained the results shown in Table 33. If we compare the results to the ones obtained from the RNC through the bottom-up approach, we observe higher values not only in terms of type frequency and *hapax legomena*, but also as regards the token frequencies of these constructions, which is due to the huge size of the corpus employed.

Table 33. Type frequencies, token frequencies and *hapax legomena* for the constructions in *-lec*, *-l'ščik* and *-lka* in the ruTenTen Corpus

	Type frequency	**Token frequency**	***Hapax legomena***
-lec	290	127,524	179
-l'ščik	129	385,275	74
-lka	514	44,741	289

Thus, we can observe that not only there are significant differences between the results obtained through a dictionary-based approach and those obtained through a bottom-up approach within the same corpus, but also that the corpus size affects significantly both type and token frequencies, as well as the number of *hapax legomena* retrieved, as shown in Table 34.

Table 34. Type frequencies, token frequencies and *hapax legomena* for the constructions in *-lec*, *-l'ščik* and *-lka* in the RNC (bottom-up approach) and in the ruTenTen Corpus

	Type frequency		**Token frequency**		***Hapax legomena***	
	RNC	*ruTenTen*	*RNC*	*ruTenTen*	*RNC*	*ruTenTen*
-lec	51	290	3,758	127,524	17	179
-l'ščik	28	129	1,431	385,275	17	74
-lka	32	514	710	44,741	8	289

The productivity rates based on the ruTenTen data were also calculated and compared to the results from the RNC (bottom-up approach), as shown in Table 35. However, as previously discussed, the productivity rates corresponding to the data from the RNC and those corresponding to the data from the ruTenTen Corpus cannot be directly compared because the difference between the two corpora in terms of token frequencies is too big.

Table 35. Token frequencies, *hapax legomena* and productivity rates (P) for the constructions in *-lec*, *-l'ščik* and *-lka* in the RNC (bottom-up approach) and in the ruTenTen Corpus

	Token frequency		***Hapax legomena***		**P**	
	RNC	*ruTenTen*	*RNC*	*ruTenTen*	*RNC*	*ruTenTen*
-lec	3,758	127,524	17	179	0.00452	0.001404
-l'ščik	1,431	385,275	17	74	0.01188	0.000192
-lka	710	44,741	8	289	0.01127	0.006459

To calculate productivity at equal token numbers, I resorted to parametric statistical models of frequency distribution known as LNRE (Large Number of Rare Events) models (Baayen 2001) and I estimated the number of expected *hapax legomena* for each construction at equal token numbers, as already done in previous studies (Štichauer 2009, 2015; Efthymiou, Fragaki & Markos 2012; Naccarato 2016). These statistical models are implemented in the R package "zipfR" (Baroni & Evert 2014), which was used to carry out this analysis. The Zipf-Mandelbrot model was used to estimate the expected number of *hapax legomena* for each construction at the equal token number of 10,000.

Table 36. Expected number of *hapax legomena* (rounded at integers) and productivity rates for the constructions in *-lec*, *-l'ščik* and *-lka* at the equal value of N = 10,000

	Token frequency		***Hapax legomena***		**P**	
	RNC	*ruTenTen*	*RNC*	*ruTenTen*	*RNC*	*ruTenTen*
-lec	10,000	10,000	22	38	0.0022	0.0038
-l'ščik	10,000	10,000	59	9	0.0059	0.0009
-lka	10,000	10,000	125	125	0.0125	0.0125

The results are shown in Table 36, in which we see that the difference in the productivity rates between the two corpora is significantly reduced once we carry out calculations based on equal token numbers. In particular, the construction in *-lka* appears as the most productive in both corpora, showing the same productivity rate of 0.0125, whereas the constructions in *-lec* and *-l'ščik* show opposite tendencies in the two corpora, the former appearing as less productive than the latter in the RNC, but more productive than the latter in the ruTenTen Corpus. However, the difference in their productivity rates is not big and could depend on the type of texts included in the two corpora.

The results obtained by using such a large corpus are certainly more adequate to carry out big quantitative investigations such as the calculation of synchronic productivity of the constructions under examination, as the corpus is very likely to include most of the current innovations in the language.

Overall, the bottom-up analyses carried out in the RNC and in the ruTenTen Corpus have also allowed finding out more about the nature of the constructions at issue, which are poorly represented in the database of compounds from the RNC.

In general, it is evident that the use of these suffixes in compounding is restricted to specific constructions. Let us consider the suffix *-lec*, for example. As previously observed (cf. Chapter 4), in the database of compounds from the RNC, this suffix is solely used to form compounds in which the base is constituted by the verb *vladet'* 'own, possess'. The bottom-up investigations have confirmed this tendency, since 45 out of 51 lemmas (88.2%) in the RNC and 258 out of 290 lemmas (89%) in the ruTenTen Corpus are instantiations of the construction [X-*vladelec*]. This tells us that the productivity of this suffix in compounding is almost exclusively restricted to this specific construction and that, while some of the compounds belonging to it are high-frequency words already fixed in dictionaries, most of them are occasionalisms in which the first element of the compound (i.e. the possessed item) can be an entity of every sort. In other words, once the speakers have become familiar with the construction [X-*vladelec*] through the words already fixed in dictionaries (which are also included in the database of compounds from the RNC, such as *domovladelec* 'landlord', *zemlevladelec* 'land owner', *rabovladelec* 'slave owner', and so on), they acquire the ability to produce compounds of the same type in which the first element is subject to great variation. Thus, for example, the search in the ruTenTen Corpus shows 12,922 tokens for the compound *avtovladelec* 'car owner'. This model is probably at the basis of the creation of many hapaxes found in the corpus, in which the first element *avto-* is substituted by the car type or brand: *gol'fovladelec* 'Golf owner', *mersedesovladelec* 'Mercedes owner', *subarovladelec* 'Subaru owner', etc. Of course, these items are not represented in dictionaries and, as a consequence, do not feature in the database of compounds on which the analysis in Chapter 4 is based.

Turning to the other suffixes, i.e. *-l'ščik* and *-lka*, we see that they also show a tendency to be mostly productive when they are part of specific constructions, such as [X-*platel'ščik*] (e.g. *nalogoplatel'ščik* 'tax payer', *rentoplatel'ščik* 'rent payer',

alimentoplatel'ščik 'alimony payer', and so on), or [X-*drobilka*] (e.g. *kamnedrobilka* 'stone crusher', *zernodrobilka* 'crop grinder', *orechodrobilka* 'nut grinder', and so on). Thus, the general trend indicates that these suffixes in compounding show a certain productivity only when they attach to specific verbal bases. In particular, their productivity seems to be restricted to cases in which the verbal base and the suffix form a derivative which exists as an autonomous word outside compounding, e.g. *vladelec* 'owner', *platel'ščik* 'payer', *drobilka* 'crusher, grinder' (cf. Section 5.5.1).

However, the huge corpus size also entails some drawbacks: such a large quantity of data would have prevented from carrying out a thorough analysis of the constructions under examination, such as the one presented in Chapter 4 (see also Appendix 1).

Another factor preventing from carrying out the whole analysis based on data from the ruTenTen Corpus is the absence of specific tags for compounds, which implies that data corresponding to the suffixless construction cannot be extracted, as the absence of an overt suffix makes it impossible to conduct a targeted search in the corpus.

Differently from the RNC, the ruTenTen Corpus does not even include metatextual tags containing information about the documents that make up the corpus, including the creation date, the textual genre and the register, which might prove useful to investigate the distribution of rival constructions (cf. Section 4.5).

In sum, the bottom-up analyses of the low-frequency constructions carried out using the RNC and the ruTenTen Corpus have demonstrated that the nature of the data has a significant impact on how productivity can be determined, and that dictionary data cannot tell us much about the synchronic productivity of word-formation processes. A bottom-up approach applied to a large corpus is definitely the best solution to retrieve adequate data on whose basis quantitative investigations can be performed. At the same time, a closer look at the data retrieved through bottom-up investigations reveals that the data extracted from the RNC word-formation database for the analysis shown in Chapter 4 can be considered as representative of the constructions under investigation. Indeed, although the data at our disposal are certainly not complete and problematic for calculations of productivity, they nonetheless include instantiations of the most productive schemas for each construction. In other words, the absence of formations such as the [car brand-*vladelec*] cases mentioned above does not entail a great loss with respect to the semantic analysis carried out in Chapter 4, as most of the compounds retrieved

through a bottom-up approach are attributable to the constructions that also feature in the RNC word-formation database.

5.4. Productivity over time

One important aspect that is inevitably disregarded when calculating productivity using the ruTenTen Corpus is the diachronic factor. The corpus includes texts from the Russian Web and is not adequate for diachronic investigations, also because it is not possible to carry out searches that allow selecting a specific time span. This function is available in the RNC, which would be an adequate source of data for the investigation of diachronic productivity, as it includes Russian texts from the 18th century to the present day. However, as previously discussed, the RNC raises other types of problems, such as the lack of a thorough annotation of compounds and the absence of an option to extract complete lemma lists.

However, the diachronic factor cannot be disregarded, as it might reveal interesting differences among the rival word-formation constructions under investigation. Moreover, this factor seems to be even more relevant when addressing the question of productivity in word-formation, as demonstrated by many studies devoted to this topic that are mostly concerned with the changes in productivity over time (among others, see Scherer 2003; Rosenberg 2007; Štichauer 2009, 2015; Bilynsky 2015; Palmer 2015; Naccarato 2016).

Still, as already discussed, the low frequencies of *hapax legomena* in the data make it impossible to determine the variation of productivity over time by applying measures based on the numbers of hapaxes.

One way to understand how the productivity of rival constructions changed over time consists in looking at the number of neologisms (or, to be more precise, the number of first occurrences) attested in the RNC over a certain time span. This way of intending productivity is sometimes referred to as "diachronic productivity" (Haspelmath & Sims 2010: 130) and can be determined only when a good historical dictionary is available, or when using large corpora that allow for diachronic investigations.

Thus, for each construction, I checked the date of creation of the text in which each compound belonging to that construction occurs for the first time. The RNC interface allows searching for a word and order the results by the creation date of the texts in which the word occurs. As the RNC contains texts from the 18th century to the present day (cf. Section 4.1), I divided the whole period covered by the main corpus in time spans of fifty years each (excluding the last time span, which covers seventeen years, i.e. from 2000 to 2016) and determined how many compounds belonging to a certain construction occur for the first time in a certain time span. The results are shown in Table 37.

Table 37. Absolute type frequencies corresponding to the first occurrences of rival constructions in the main corpus of the RNC

	1700–1749	**1750–1799**	**1800–1849**	**1850–1899**	**1900–1949**	**1950–1999**	**2000–2016**
-ec	21	37	23	27	23	10	2
-lec	0	0	2	4	1	0	0
-tel'	9	10	7	25	46	70	8
-nik	3	4	2	3	8	7	1
-ščik/čik	1	0	0	2	3	16	1
-l'ščik	0	0	1	0	1	2	0
-ka	2	3	14	9	8	10	1
-lka	0	0	0	2	5	4	0
-ø	16	53	51	77	125	61	10
Total	**52**	**107**	**100**	**149**	**220**	**180**	**23**

A first important observation to be made is that the vast majority of first occurrences of compounds (for most of the constructions) is concentrated in the 20th century, which might be due to the greater number of texts available in the RNC for that time span.[44] This being said, if we regard the number of first occurrences in each time span as an indicator of productivity, we can observe that rival constructions show different trends. In particular, the comparison of the three most frequent constructions, i.e. the suffixless construction (393 compounds), the construction in *-tel'* (175 compounds), and the

[44] Indeed, the number of texts (and, consequently, the number of tokens) included in the main corpus of the RNC is different for different time spans: 4,726,499 tokens for the 18th century, 53,090,226 tokens for the 19th century, 141,267,193 tokens for the 20th century, and 70,893,673 tokens for the first seventeen years of the 21st century. Most texts, then, are found in the 20th century.

construction in *-ec* (143 compounds), reveals some interesting facts. The results of the analysis corresponding to these three constructions are plotted in Figure 29, in which the last time span (2000–2016) was excluded because it would have led to misleading conclusions. Indeed, the low frequencies in the last time span are mostly the consequence of the fact that the data hardly include words coined in recent times, as these words are already included in dictionaries. Remarkably, the analysis carried out in the ruTenTen Corpus has demonstrated that the current picture is rather different and that the constructions under examination are productively employed in present-day texts.

Figure 29. First occurrences in the main corpus of the RNC for the three most frequent constructions

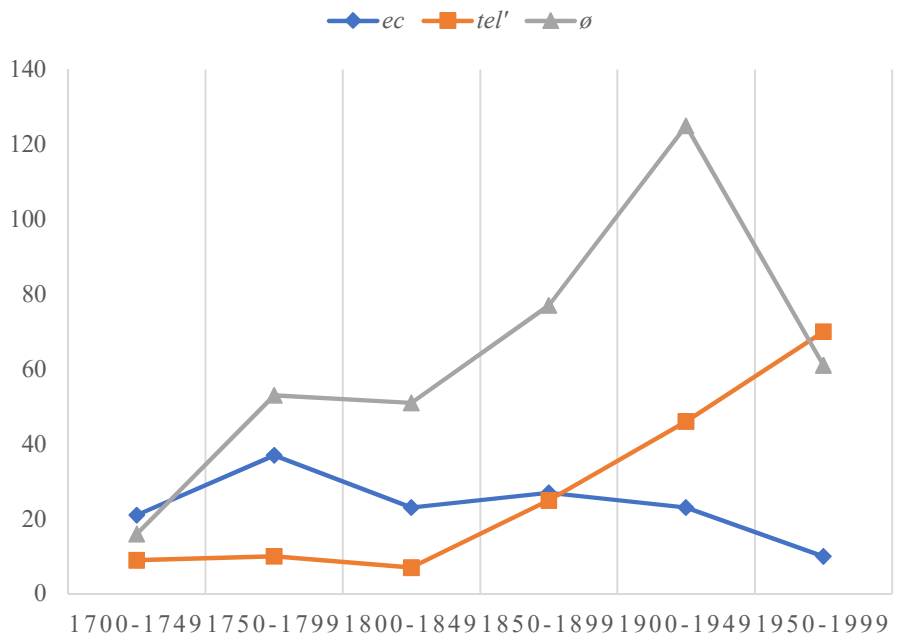


Figure 29 shows that most compounds in *-ec* occur for the first time in the RNC in the second half of the 18th century, which is significant, considering that the number of tokens available for that time span is much smaller compared to the other time spans. The production of new compounds in *-ec* seems to decrease over time and, especially starting from the 20th century, we find very few occurrences of new compounds in *-ec*, compared to the other two constructions. Most compounds in *-tel'* occur for the first time in the second half of the 20th century. An important factor to consider is that, as shown in Section

4.3.6, *-tel'* compounds denote Instruments much more often than *-ec* compounds, and this might lead us to think that the creation of new compounds in *-tel'* during the 20th century could be due to the need of denoting new technical devices, such as *skorosšivatel'* 'loose-leaf binder' (1927), *metalloiskatel'* 'metal detector' (1939), *vozduchopodogrevatel'* 'air heater' (1964), etc.

(2) «*Na zavode sel'skochozjajstvennych mašin izgotovlena opytnaja partija* ***vozduchopodogrevatelej*** *na židkom toplive proizvoditel'nost'ju 400 tys. kkal/č. Oni prednaznačeny dlja suški zernobobovych, počatkov kukuruzy, podsolnečnika, l'novolokna, kormovych bobov, a takže dlja dosuški sena.* ***Vozduchopodogrevateli*** *mogut rabotat' kak ot traktora, tak i ot èlektroseti*.» [obobščennyj. Korotkie korrespondencii // «Technika - molodeži», 1964]

'At the factory of agricultural machines a prototype lot of liquid propellant **air heaters** with a productivity of 400 thousand kcal/h has been manufactured. They are intended to dry legumes, corn cobs, sunflowers, flax fibers, broad beans and hay. **Air heaters** can work both from tractors and from an electric grid.'

Example (2) shows the first occurrence of the compound *vozduchopodogrevatel'* 'air heater' in the RNC (1964). This compound clearly denotes a referent that is new to the audience to whom the text is addressed. Accordingly, the text includes explanations concerning the functioning and intended applications of this instrument.

On the contrary, *-ec* compounds rarely denote Instruments, and almost never occur in technical discourse (cf. Section 4.5.2). Rather, their semantics is often associated with religious concepts (e.g. *bogomolec* 'pilgrim, devotee', *ikonoborec* 'iconoclast', *psalmopevec* 'psalmist', etc.), which explains why most of these compounds occur for the first time in earlier texts.[45]

The suffixless construction, which is the most frequent in the RNC, seems to be able to create all types of nouns, being less specialized by comparison to the construction in *-tel'*, and more neutral with respect to *-ec* (mainly in terms of distribution across genres

[45] Most compounds in *-ec* related to religious concepts are borrowed from Old Church Slavonic, where, in turn, they appear as the result of a calquing process from Ancient Greek (see Efimova 2006).

and registers; cf. Sections 4.5.2 and 4.6). Most compounds of this type show their first occurrence in the first half of the 20th century, but their production surpasses the production of *-ec* and *-tel'* compounds also in other time spans, particularly starting from the second half of the 18th century. The considerable growth in the number of first occurrences for the suffixless construction during the first half of the 20th century might be seen as the result of the new naming needs during the post-revolutionary epoch.

This period of disruption had significant consequences on the Russian lexicon, which was enriched by the introduction of terms denoting new professions, but also, more generally, new referents related to the social changes of that time. Remarkably, two of the most common constructions employed to denote professionals in different sectors, i.e. [X-*vod*] (e.g. *gruppovod*[46] 'group leader', first occurrence: 1934) and [X-*ved*] (e.g. *obščestvoved* 'expert on social sciences', first occurrence: 1920–1929), belong to the suffixless type. These two constructions alone make up the 23.4% of the suffixless group (cf. Appendix 3). Among the most frequent constructions within the suffixless group, we also find [X-*ed*] (e.g. *kuroed* 'bribe-taker', first occurrence: 1928–1940), [X-*ljub*] (e.g. *svobodoljub* 'freedom lover', first occurrence: 1923), [X-*voz*] (e.g. *bombovoz* 'bomber', first occurrence: 1926), and [X-*chod*] (e.g. *vezdechod* 'off-road vehicle', first occurrence: 1932–1942) (cf. Appendix 3). These constructions show a high number of first occurrences in the post-revolutionary period.

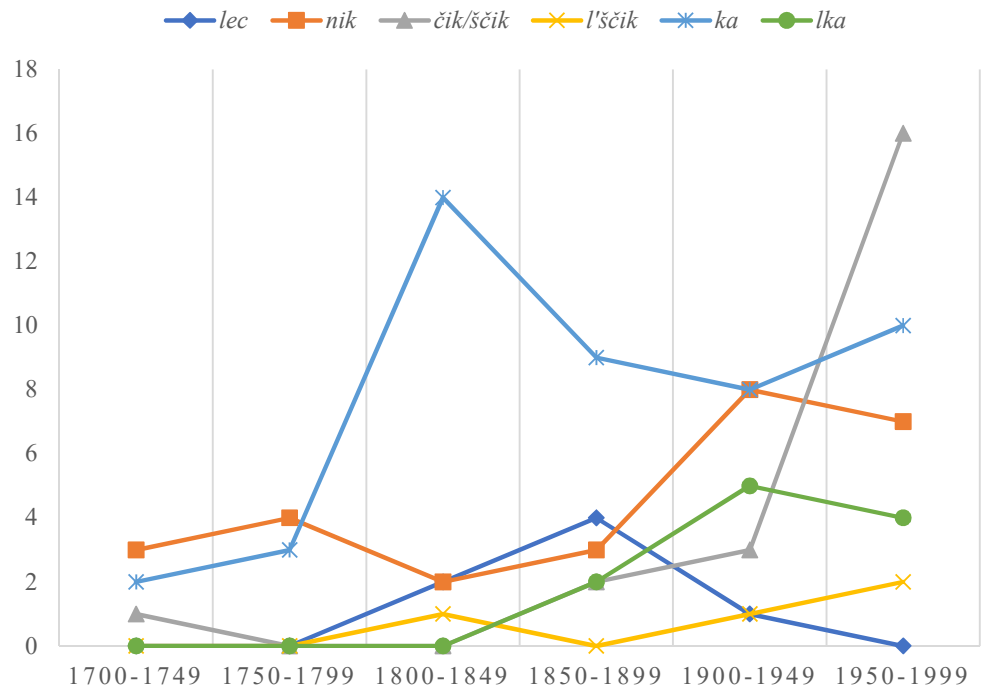
Conversely, in the period following the Second World War, we see a drop in the number of first occurrences for the suffixless construction. Note, however, that this drop is not that significant if compared to the number of first occurrences in the second half of the 19th century. It is the period in between that shows a significant difference with respect to all the others, which reflects the increased word-formation activity of those years, and has to be understood as related to the huge changes brought about by the Revolution. Remarkably, what continues to grow in the second half of the 20th century is the need to name new instruments that are the result of technological progress. As we have seen before, the construction in *-tel'*, which shows higher percentages of instrument nouns with respect to *-ec* and the suffixless construction (cf. Table 15 in Chapter 4), is the construction that shows a constant growth during the 20th century. Noticeably, although

[46] The term *gruppovod* appears in the dictionary of Sovietisms by Mokienko & Nikitina (1998) with the meaning of 'leader of a group in any work or activity (in the first years of the Soviet authority)' ("*rukovoditel' gruppy na kakich-l. rabotach, zanijatijach (v pervye gody sovetskoj vlasti)*").

the suffixless construction shows a consistent drop in the number of first occurrences during the second half of the 20th century, the number of new instrument nouns formed with this construction does not diminish, but increases from 21 to 23 first occurrences (from 16.8% to 37.7%).

Figure 30 shows the number of first occurrences in the RNC for each of the remaining constructions under examination (e.g. *-lec*, *-nik*, *-ščik/čik*, *-l'ščik*, *-ka*, and *-lka*). The graph suggests that most of these constructions (*-nik*, *-ščik/čik*, *-l'ščik*, and *-lka*) produce more new compounds during the 20th century, while only *-lec* and *-ka* show a greater number of first occurrences during the 19th century. However, the results corresponding to these constructions must be interpreted cautiously, due to their low frequencies in the word-formation database of the RNC.

Figure 30. First occurrences in the main corpus of the RNC for the minor constructions



Overall, the methodology applied in the present section to determine the diachronic productivity of the constructions under investigation appears as the solution that best suits the type of data at our disposal. The dictionary data which make up the word-formation database of the RNC can only witness possible changes in the productivity of the

examined constructions over time, but they cannot be employed to draw conclusions related to their synchronic productivity in Russian.

5.5. Productivity across genres and registers

The role of registers and textual types in determining the variation of productivity has been discussed in previous works (among others, see Baayen 1994a; Plag, Dalton-Puffer & Baayen 1999; Efthymiou, Fragaki & Markos 2011), where it is pointed out that the productivity of certain derivational affixes is subject to variation depending on the register (written vs. oral register) and on the textual type (e.g. fiction vs. non-fiction texts).

As already mentioned in Section 5.1, the paragraphs of the AG-80 devoted to Russian compounds formed with agent noun suffixes discuss the productivity of the different constructions as related to their employment in different areas of the language, and it is claimed that the productivity of most of these constructions is limited to technical terminology (*-tel'*, *-nik*, *-ščik/čik*, *-l'ščik*), and, in some cases, it also extends to everyday language (*-ka*, *-lka*). The construction in *-ec* is reported to be not very productive, i.e. its productivity is restricted to cases in which the verbal base plus the suffix already form an autonomous deverbal outside compounding (cf. Section 5.1).

Although exact calculations of productivity based on hapax measures cannot be applied to the data at our disposal for the reasons discussed throughout this chapter, the frequency-based analysis of the distribution of rival constructions across different textual genres presented in Section 4.5.2 seems to confirm that the constructions in *-tel'*, *-nik*, *-ščik/čik*, *-l'ščik*, and *-lka* are overrepresented in specialized (i.e. technical and official) texts with respect to the other constructions. The high frequencies of the suffixes *-ka* and *-lka* in fiction texts might also confirm the claims put forth in the AG-80, that is, that these constructions are also productive in everyday language. As already discussed in Section 4.5.2, the construction in *-ec* and the suffixless construction (whose productivity is scarcely discussed in the AG-80) are less represented in specialized texts.

The analysis shown in Section 4.5.3 has shown that the distribution of rival constructions is also affected by the type of register, i.e. written or spoken, and particularly that the constructions that show higher frequencies in written texts are *-ec*

and -*lec* (followed by -*tel'* and -*lka*), while in spoken texts the suffixless construction is overrepresented with respect to the others.

5.6. Restrictions on productivity

As has been shown in the previous sections, determining the productivity of rival word-formation constructions is not an easy task, especially if the available data and the corpora employed do not meet the requirements that would allow measuring productivity by adopting quantitative methods. The data from the word-formation database of the RNC have led to reaching reliable results only in terms of diachronic productivity (cf. Section 5.4), while they proved to be unsuitable for calculations of productivity that rely on hapax-based measures. The experiments described in Section 5.3, instead, have shown that such hapax-based approaches to productivity are adequate to determine the synchronic productivity of word-formation processes when data from large contemporary corpora are available.

In the present section, I address the question of productivity as related to some restrictions to which the constructions under examination are subject. Indeed, the productivity, intended here in a qualitative sense as the ability of a word-formation process to give rise to new words, of the rival constructions analyzed throughout this study seems to be dependent to a large extent on the nature of their components, and particularly on the availability of autonomous deverbals outside compounding (5.6.1) and on the type of verbal base included as the second element of the compound (5.6.2).

5.6.1. Availability of autonomous deverbals outside compounding

Remarkably, while some of the constructions examined here seem to be able to produce compounds almost only when the verbal base plus the agentive suffix form an existing deverbal noun (i.e. a noun that exists as an autonomous word outside the compound), other constructions do not seem to be subject to these restrictions. If this is true, it might be the case that the varying degrees of productivity of different constructions largely

depend on the restrictions to which they are subject. The possible situations are the following:

a) the verbal base plus the agentive suffix form a deverbal noun that is used outside the compound with the same meaning that it has inside the compound, e.g. *bytopisatel'* 'everyday-life writer' (*pisatel'* 'writer');
b) the verbal base plus the agentive suffix form a deverbal noun that is used outside the compound with a different meaning, e.g. *zubočistka* 'toothpick' (*čistka* 'cleaning process');
c) the verbal base plus the agentive suffix do not form a deverbal noun that exists outside compounding, e.g. *oruženosec* 'squire' (**nosec*).

In order to test this hypothesis, for each compound in the database, I checked whether the deverbal noun is attested in the RNC and, in case it is attested, I checked whether its meaning inside the compound corresponds to its meaning in dictionaries.[47]

Table 38. Absolute and relative frequencies of compounds including an existing deverbal

	Absolute frequency	**Relative frequency (%)**
-ec	**71**/143	49.6
-lec	**7**/7	100
-tel'	**172**/175	98.3
-nik	**26**/28	92.9
-ščik/čik	**20**/23	87
-l'ščik	**4**/4	100
-ka	**33**/47	70.2
-lka	**9**/11	81.8
-ø	**4**/393	1

Table 38 shows the absolute and relative frequencies of compounds including an existing deverbal, while Table 39 shows how many of these existing deverbals share the same meaning inside and outside compounding. The suffixless construction shows a peculiar situation, as the only four cases in which a deverbal exists outside compounding are cases

[47] For the dictionary search, I resorted to the site http://gufo.me/, which includes several dictionaries of modern Russian, among which are Ušakov (1935–1940), Ožegov & Švedova (1996), Kuznecov (1998), and Efremova (2000).

of prefixed verbal bases, i.e. *vypusk* 'release', *sbros* 'dumping', *spusk* 'descent', and *stok* 'flow', all showing a processual meaning, that is a different meaning with respect to the instrumental meaning they have in compounds.

Table 39. Absolute and relative frequencies of deverbals sharing the same meaning inside and outside compounding

	Absolute frequency	**Relative frequency (%)**
-ec	**42**/71	59.1
-lec	**7**/7	100
-tel'	**139**/172	80.8
-nik	**20**/26	76.9
-ščik/čik	**17**/20	85
-l'ščik	**4**/4	100
-ka	**10**/33	30.3
-lka	**8**/9	88.9
-ø	**0**/4	-

Let us first consider the low-frequency constructions in *-lec*, *-l'ščik* and *-lka*. As Table 38 shows, these constructions are almost always based on preexisting deverbals (all compounds in *-lec*, all compounds in *-l'ščik*, and 9 compounds out of 11 in *-lka*). Moreover, Table 39 shows that in almost all cases the meaning of the deverbal inside and outside compounding overlaps. If we also consider the analysis presented in Section 5.3, we can conclude that the productivity of these three constructions in modern Russian is restricted not only to cases in which the compounds are based on preexisting deverbal derivatives, but more specifically, it is restricted to compounds based on specific derivatives only. The most striking example is constituted by the construction in *-lec*, which in my database is exclusively represented by compounds of the type [X-*vladelec*] (i.e. owner of X), as previously discussed. Thus, we can conclude that these patterns of compounding in Russian are available for the creation of new words almost exclusively when they are based on preexisting derivatives that retain their meaning inside the compound.

The constructions in *-tel'*, *-nik* and *-ščik/čik* are also often based on preexisting deverbals, as can be seen in Table 38: this is the case for 98.3% of the compounds in *-tel'*, 92.9% of the compounds in *-nik*, and 87% of the compounds in *-ščik/čik*. In about

80% of these cases, the preexisting deverbal in these constructions retains its meaning (cf. Table 39).

Regarding the suffix *-tel'*, the tendency to give rise to compounds mostly based on preexisting deverbals has already been observed by Lychyk (1995: 149), who points out that *-tel'* compounds with personal meaning are mostly based on preexisting words, i.e. the verbal base plus the agentive suffix give rise to an autonomous word outside compounding, e.g. *vzjatkopolučatel'* 'bribe taker', where *polučatel'* 'recipient' is an existing word. Although Lychyk (1995) notes this tendency only as regards the suffix *-tel'*, this seems to be true also for *-nik* and *-ščik/čik*. Most of the times, indeed, these suffixes form compounds based on preexisting words, such as *kljatvoprestupnik* 'perjurer' (*prestupnik* 'criminal'), or *zvezdoletčik* 'astronaut' (*letčik* 'pilot').

The situation seems to be different for the remaining constructions, i.e. *-ec*, *-ka*, and the suffixless construction, which appear to be less subject to restrictions of this type. The figures in Table 38 show that the constructions in *-ec* and *-ka* are based on preexisting derivatives only in 49.6% and 70.2% of the cases, respectively. Even when the deverbal exists outside compounding, it does not always retain its meaning (59.1% of the cases for *-ec* and 30.3% of the cases for *-ka*). Thus, only 42 compounds in *-ec* out of 143 and 10 compounds in *-ka* out of 47 appear to be based on preexisting deverbals, while in most cases either the deverbal does not exist as an autonomous word (e.g. *kanatochodec* 'rope walker' and *myšelovka* 'mousetrap', where **chodec* and **lovka* do not exist as autonomous words), or it exists, but carries a different meaning (e.g. *pervoprochodec* 'earlier explorer' and *kofevarka* 'coffee machine', where *prochodec* is attested outside compounding only as the diminutive of *prochod* 'passage', and *varka* has the meaning of 'boiling process').

Turning to the suffixless construction, I have already pointed out that the only four cases in which a deverbal exists outside compounding are cases of prefixed verbal bases that have processual meanings outside compounding. Thus, we can say that the suffixless construction does not include any compound based on a preexisting deverbal.

Some scholars (cf. Bisetto & Melloni 2008; Melloni & Bisetto 2010; see also the discussion in Chapters 1 and 2) would describe these opposing tendencies in terms of different word-formation processes. Following this approach, the latter examples, showing a non-existent derivative or a derivative that carries a different meaning outside

compounding (e.g. *kanatochodec*, *myšelovka*, *pervoprochodec*, *kofevarka*, and others) would be described as parasynthetic compounds, in which the compounding process takes place at the same time as the addition of the suffix to the lexical bases. The other cases, more typical of the remaining constructions, showing an existing derivative that retains its meaning inside the compound (e.g. *vzjatkopolučatel'*, *kljatvoprestupnik*, *zvezdoletčik*, and others) would be interpreted differently in such a framework, i.e. as the result of a compounding process that takes place after a process of deverbal derivation. Thus, *vzjatkopolučatel'* would be formed by attaching the noun *vzjatka* 'bribe' to the noun *polučatel'* 'receiver' (see also the opposition of "bound nominalizations" and "autonomous nominalizations" proposed by Molinsky 1973).

As previously discussed (cf. Chapter 2, Section 2.3), in a constructionist approach we do not necessarily need to draw a line between the two situations. As I have shown, it is not always possible to determine precisely which one of the two word-formation processes is at play in some compounds, i.e. whether derivation takes place before compounding or whether the two processes are simultaneous, and the best solution to describe such cases is to resort to the concept of schema unification (cf. Chapter 2, Section 2.3).

However, it needs to be remarked that the existence of the deverbal noun outside compounding seems to affect the availability of some of the word-formation constructions in giving rise to new words in modern Russian, while other constructions do not seem to be much affected by this factor.

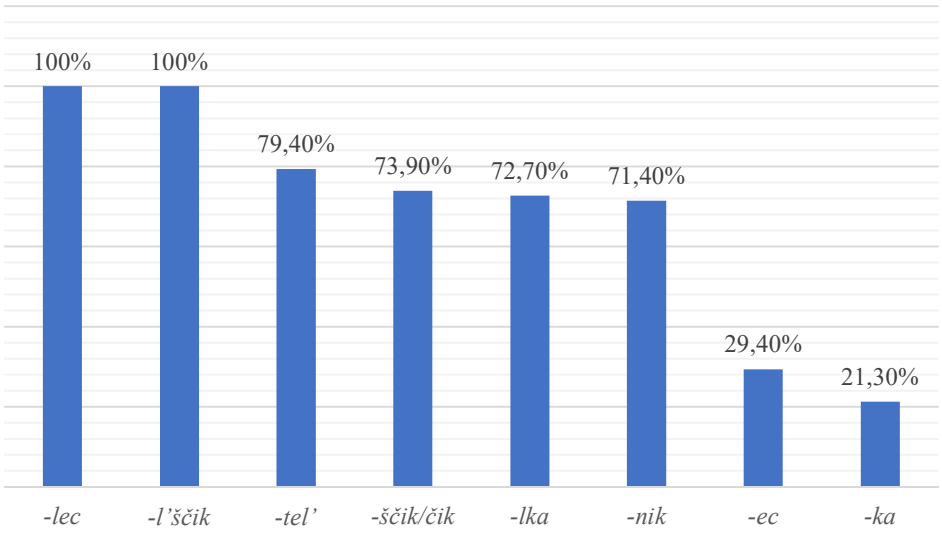
The graph reproduced in Figure 31 summarizes the findings of the analysis carried out in the present section and shows the percentage of compounds based on an existing derivative that retains its meaning inside compounding for each construction.

Interestingly, while the productivity of the constructions in *-lec*, *-l'ščik*, *-lka*, *-tel'*, *-nik* and *-ščik/čik* heavily depends on the existence of preexisting derivatives, the constructions in *-ec* and *-ka* undergo fewer restrictions in this respect. The suffixless construction is not represented in the graph because it does not show such cases.

The low percentage of compounds in *-ec* appearing in Figure 31 contradicts the claim in the AG-80 that the productivity of this construction is limited to cases in which the verbal base plus the suffix form an autonomous derivative outside compounding (§

559; cf. Sections 5.1 and 5.5). The data that I have shown, instead, demonstrate that this is only true in a minority of cases.

Figure 31. Percentage of compounds showing an existing deverbal that retains its meaning inside compounding



5.6.2. Variability of the verbal bases

Another type of restriction that can have an impact on the productivity of rival word-formation constructions is related to the variability of the verbal bases incorporated in compounds.[48] Indeed, as previously discussed (cf. Section 5.3), the productivity of certain constructions seems to be restricted to specific verbal bases, as in the case of the suffix *-lec*, whose productivity in compounding seems to be almost exclusively limited to the construction [X-*vladelec*] 'owner of X', based on the verb *vladet'* 'own, possess'. Thus, for each construction in the database, I checked the number of different verbs on which compounds are based. The results of this analysis are shown in Table 40.

[48] The productivity of compounds as restricted to specific bases has also been discussed in Kapatsinski & Vakareliyska (2013).

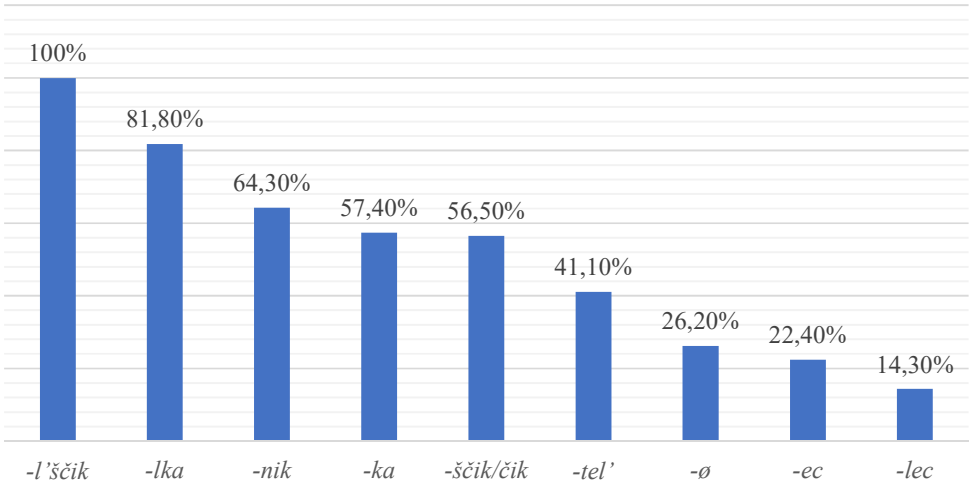
Table 40. Number of compounds and number of different verbal bases for rival constructions

	Number of compounds	Number of different verbal bases
-ec	143	32
-lec	7	1
-tel'	175	72
-nik	28	18
-ščik/čik	23	13
-l'ščik	4	4
-ka	47	27
-lka	11	9
-ø	393	103[49]

As Table 40 shows, for the construction in *-ec*, we find 143 compounds, but the number of different verbal bases included in these compounds only amounts to 32; for the construction in *-lec*, we have 7 compounds based on a single verbal stem, the construction in *-tel'* includes 175 compounds based on 72 different verbal stems, and so on.

Thus, to compare the degree of variability of the verbal bases among rival constructions, I divided the number of different verbal bases by the total number of compounds for each construction, and I got the percentages shown in Figure 32.

Figure 32. Percentage of verb-base variability



[49] The full list of base verbs for compounds belonging to the suffixless construction is found in Appendix 3.

The graph in Figure 32 shows different behavior patterns for rival constructions. While some of them show little variation in terms of the verbal bases embedded in compounds, others seem to allow for greater variation, and thus their productivity undergoes fewer restrictions in this sense.

As usual, the figures corresponding to the low-frequency constructions must be interpreted cautiously. The suffix *-l'ščik*, for instance, appears in first position, showing the greatest variation among rival constructions. This is due to the fact that the four compounds with the suffix *-l'ščik* found in the database are based on four different verbs, which brings to a 100% variation. However, the bottom-up analyses described in Section 5.3 have shown that, when the search is carried out in a large corpus, variation is much more limited, and the production of new compounds in *-l'ščik* is mostly restricted to a few specific constructions, the most common of which is [X-*platel'ščik*] 'payer of X'. The same is true for *-lka*, which infelicitously appears here in second position, and for *-lec*, whose low variation (confirmed by the bottom-up searches) is instead well represented in the graph.

Among the other constructions, those showing lower degrees of variation are the construction in *-ec* and the suffixless construction, which might suggest that the productivity of these constructions is mainly restricted to specific constructions.

Interestingly, it is often the case that constructions based on the same verbal base give rise both to compounds in *-ec* and to suffixless compounds. Thus, for example, the construction based on the verb *ljubit'* 'love' is well represented in both cases: [X-*ljubec*] and [X-*ljub*] 'one who loves X' produce 18 compounds each. This trend also confirms that these two constructions appear to behave very similarly and the competition between them is greater, as already discussed in Chapter 4 (cf. Sections 4.5 and 4.6).

What is crucial for the present discussion, however, is that a low base-verb variability indicates that it is probably specific constructions that are productive, rather than the high-order construction in its entirety. In other words, this means that the creation of new words in *-ljubec* or *-ljub* is boosted by the high frequencies of existing words belonging to the specific constructions [X-*ljubec*] and [X-*ljub*], rather than by the productivity of the high-order constructions in *-ec* and *-ø*. Thus, we might suppose that the most frequent words in *-ljubec* and *-ljub* have served as a basis for the creation of new words belonging to these constructions.

5.7. A comparison of partially specified schemas: two case studies

Starting from the observations put forward in Section 5.6.2, in the present section I compare the productivity of partially specified schemas, i.e. the constructions [X-*ljubec*] and [X-*ljub*] (cf. Section 5.7.1), and the constructions [X-*vodec*] and [X-*vod*] (cf. Section 5.7.2). These comparisons can be considered as a test to investigate the differences in productivity of the construction in *-ec* and the suffixless construction, which have repeatedly proven to be in stronger competition, being less specialized with respect to most of the other constructions (cf. Sections 4.5, 4.6, and 5.6). The analyses compare data retrieved from dictionaries (i.e. the RNC word-formation database used throughout this study) and data retrieved through a bottom-up search from the RNC and from the ruTenTen Corpus, and thus also aims at adducing more evidence to the discussion in Section 5.3, in which I have claimed that the nature of the data and of the corpora employed significantly affects calculations of productivity.

5.7.1. [X-ljubec] vs. [X-ljub]

The first case study compares the constructions [X-*ljubec*] and [X-*ljub*] 'one who loves X'. First, I used the RNC interface to check the number of tokens and *hapax legomena* corresponding to the compounds in *-ljubec* (e.g. *pravdoljubec* 'truth lover', from *pravda* 'truth' and *ljubit'* 'love') and *-ljub* (e.g. *žizneljub* 'life lover', from *žizn'* life' and *ljubit'* 'love') included in the RNC word-formation database. The results of this search are shown in Table 41.

Table 41. Type frequencies, token frequencies and *hapax legomena* of the constructions [X-*ljubec*] and [X-*ljub*] in the RNC word-formation database

	Type frequency	**Token frequency**	***Hapax legomena***
[X-*ljubec*]	18	1,194	0
[X-*ljub*]	18	565	3

Second, I carried out the searches **ljubec* and **ljub* in the main subcorpus of the RNC to extract the relevant data through a bottom-up approach, which yielded the results in Table 42.

Table 42. Type frequencies, token frequencies and *hapax legomena* of the constructions [X-*ljubec*] and [X-*ljub*] in the RNC (bottom-up approach)

	Type frequency	**Token frequency**	***Hapax legomena***
[X-*ljubec*]	83	1,518	40
[X-*ljub*]	82	1,280	44

Finally, I carried out the same searches in the ruTenTen Corpus. The results are shown in Table 43.

Table 43. Type frequencies, token frequencies and *hapax legomena* of the constructions [X-*ljubec*] and [X-*ljub*] in the ruTenTen Corpus

	Type frequency	**Token frequency**	***Hapax legomena***
[X-*ljubec*]	116	12,663	44
[X-*ljub*]	428	15,785	236

Now I can calculate the productivity rates of both constructions in the RNC and in the ruTenTen Corpus by using the formula in (1). The results are shown in Table 44. In all three cases, the productivity of the construction [X-*ljub*] is higher than the productivity of its rival [X-*ljubec*].

Table 44. Productivity rates of the constructions [X-*ljubec*] and [X-*ljub*] in the RNC (word-formation database and bottom-up approach) and in the ruTenTen Corpus

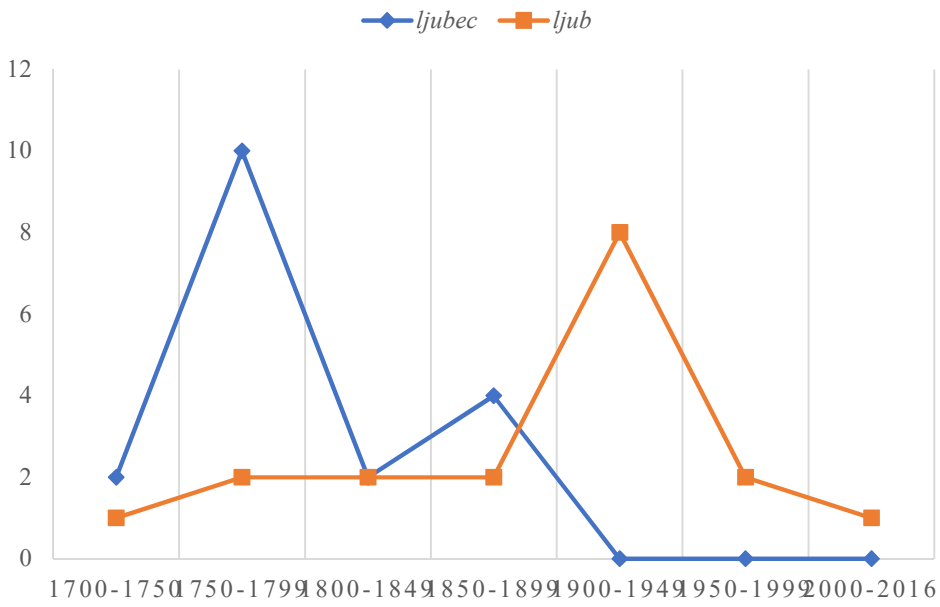
	Word-formation database of the RNC	**RNC (bottom-up approach)**	**ruTenTen Corpus (bottom-up approach)**
[X-*ljubec*]	0	0.026	0.003
[X-*ljub*]	0.005	0.034	0.014

If we also consider diachronic productivity, that is the number of new occurrences in different time spans (cf. Section 5.4), we see that the suffixless construction is becoming

more productive over time with respect to the construction in *-ec*. Figure 33 and Figure 34 show the diachronic productivity of the two constructions. Figure 33 shows the results based on data from the RNC word-formation database, while Figure 34 shows the results of a bottom-up approach. In both cases, the number of first occurrences in different time spans was checked in the RNC.[50]

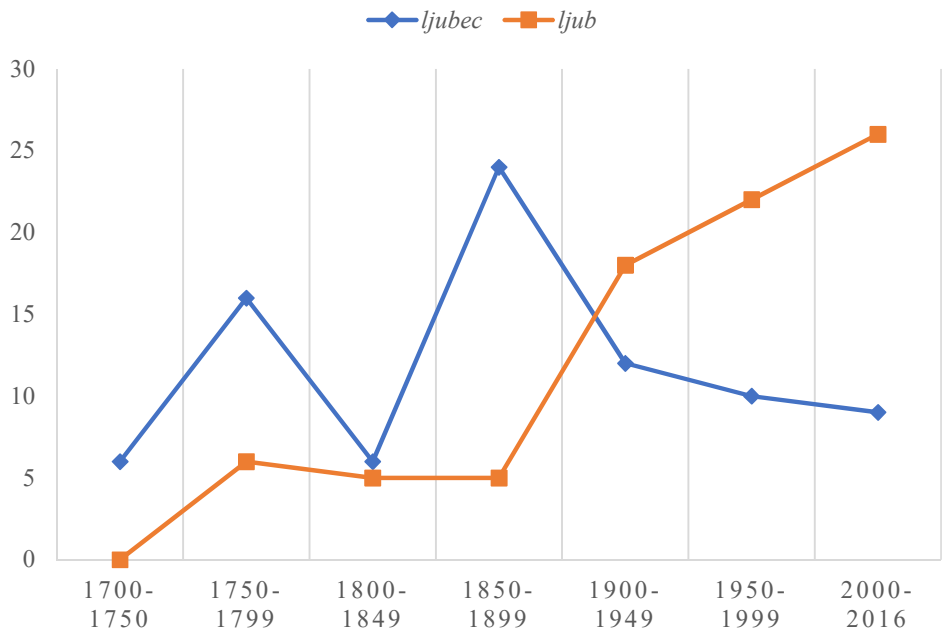
Notably, both analyses demonstrate that the construction with the suffix *-ec* shows greater productivity in earlier texts with respect to the suffixless construction, which, conversely, increases its productivity especially starting from the 20th century. The analysis carried out through a bottom-up approach gives us more reliable results for the last time span (as the data obtained also comprise compounds which are not yet included in dictionaries), and shows that the productivity of the suffixless construction seems to be still growing.

Figure 33. Diachronic productivity of the constructions [X-*ljubec*] and [X-*ljub*] in the RNC word-formation database



[50] As already discussed in Section 5.4, the ruTenTen Corpus does not allow carrying out diachronic investigations.

Figure 34. Diachronic productivity of the constructions [X-*ljubec*] and [X-*ljub*] in the RNC (bottom-up approach)



5.7.2. [X-vodec] vs. [X-vod]

Following the same procedure described in Section 5.7.1, I compared the constructions [X-*vodec*] (e.g. *polkovodec* 'commander', from *polk* 'regiment' and *vodit'* 'guide, lead') and [X-*vod*] (e.g. *èkskursovod* 'tour guide', from *èkskursija* 'tour' and *vodit'* 'guide, lead'). First, I used the RNC interface to check the number of tokens and *hapax legomena* corresponding to all compounds in -*vodec* and -*vod* included in the RNC word-formation database. The results are shown in Table 45.

Table 45. Type frequencies, token frequencies and *hapax legomena* of the constructions [X-*vodec*] and [X-*vod*] in the RNC word-formation database

	Type frequency	**Token frequency**	***Hapax legomena***
[X-*vodec*]	3	3,436	1
[X-*vod*]	50	4,469	5

Then, I carried out the searches **vodec* and **vod* in the main subcorpus of the RNC and extracted the relevant compounds through a bottom-up approach. The results are shown in Table 46.

Table 46. Type frequencies, token frequencies and *hapax legomena* of the constructions [X-*vodec*] and [X-*vod*] in the RNC (bottom-up approach)

	Type frequency	**Token frequency**	***Hapax legomena***
[X-*vodec*]	6	3,440	3
[X-*vod*]	97	8,284	29

Finally, I carried out the same searches in the ruTenTen Corpus, which yielded the results in Table 47.

Table 47. Type frequencies, token frequencies and *hapax legomena* of the constructions [X-*vodec*] and [X-*vod*] in the ruTenTen Corpus

	Type frequency	**Token frequency**	***Hapax legomena***
[X-*vodec*]	44	101,526	24
[X-*vod*]	770	551,443	296

Table 48 shows the productivity rates calculated using the formula in (1) based on the data shown in Table 45, Table 46, and Table 47.

Table 48. Productivity rates of the constructions [X-*vodec*] and [X-*vod*] in the RNC (word-formation database and bottom-up approach) and in the ruTenTen Corpus

	Word-formation database of the RNC	**RNC (bottom-up approach)**	**ruTenTen Corpus (bottom-up approach)**
[X-*vodec*]	0.0003	0.0009	0.0002
[X-*vod*]	0.0011	0.0035	0.0005

Again, we see that the suffixless construction shows higher productivity rates with respect to the construction in *-ec*, regardless of the corpus employed and of the approach adopted. The graphs in Figure 35 and Figure 36 show the diachronic productivity of the two constructions, i.e. the number of first occurrences of [X-*vodec*] and [X-*vod*] in different

time spans. The results adduce evidence of the higher productivity of the suffixless construction, which reaches its peak during the first half of the 20th century.

Figure 35. Diachronic productivity of the constructions [X-*vodec*] and [X-*vod*] in the RNC word-formation database

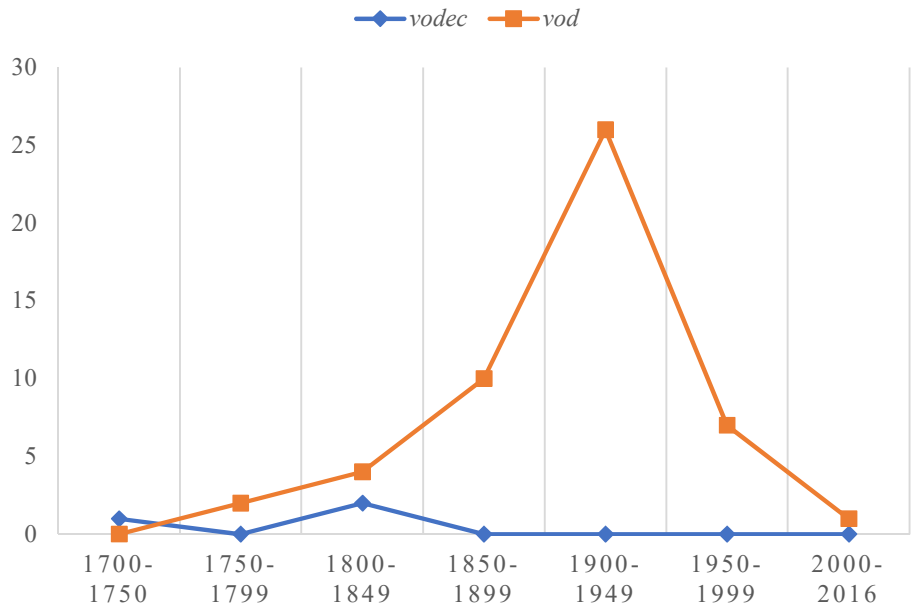
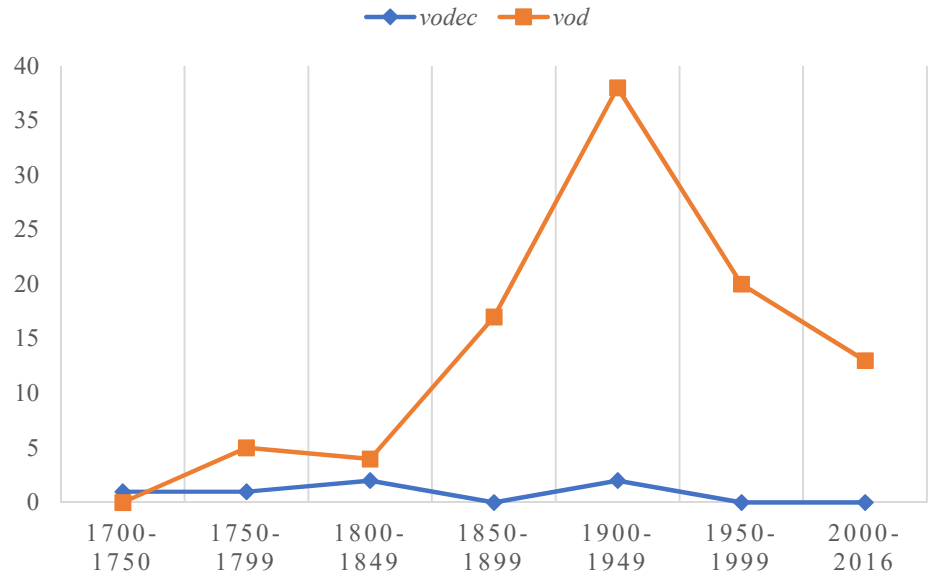


Figure 36. Diachronic productivity of the constructions [X-*vodec*] and [X-*vod*] in the RNC (bottom-up approach)



The highest productivity of the suffixless construction depends on a number of factors. Apart from what already pointed out in relation to the rivalry of *-ec* and the suffixless construction, that is, that the construction in *-ec* appears as more archaic and belonging to a higher register with respect to the suffixless construction (cf. Chapter 4, Section 4.6.2), we see here that other factors are at play. On the one hand, the construction [X-*vodec*] is almost exclusively used to form nouns denoting the leader of someone or something (i.e. one who leads someone/something), as in *polkovodec* 'commander' (from *polk* 'regiment' and *vodit'* 'guide, lead') and *flotovodec* 'naval commander' (from *flot* 'fleet' and *vodit'* 'guide, lead'), which are the most frequent compounds belonging to this construction. On the other hand, the construction [X-*vod*] inherits the polysemy of its base verb *vodit'* 'guide, lead, drive' (also 'breed, raise' when found in the prefixed form *razvodit'*) and forms nouns denoting:

a) the leader of someone or something, e.g. *èkskursovod* 'tour guide' (from *èkskursija* 'tour' and *vodit'* 'guide, lead');
b) the breader or raiser of something, e.g. *olenevod* 'reindeer breeder' (from *olen'* 'reindeer' and *razvodit'* 'breed');
c) the driver (and passionate) of a certain type of car, e.g. *tojotovod* 'Toyota driver/lover' (from *Tojota* 'Toyota' and *vodit'* 'drive');
d) in general, someone who is passionate or interested in something, e.g. *makovod* 'Mac user/passionate' (from *Mak* 'Mac' and *vodit'* 'guide, lead').

Thus, I argue that the larger polysemy brought about by the construction [X-*vod*] also contributes to its higher productivity with respect to [X-*vodec*].

5.8. Productivity and polysemy

Productive word-formation processes are often conceived as being correlated to semantic coherence, i.e. transparency and predictability (see, for instance, Aronoff 1976; Bauer 1983; Koefoed & van Marle 2000). However, as other scholars point out when discussing the status of the English suffix *-er* (cf. Kiefer 2000; Panther & Thornburg 2003), the polysemy and semantic vagueness of this suffix do not prevent it from being productive in English.

Panther & Thornburg (2003: 311) argue that, although the meaning of the suffix *-er* is not strictly predictable, its polysemy is semantically coherent, as it arises from mechanisms of semantic extension, i.e. metaphor and metonymy (cf. Section 3.3). The authors claim that it is exactly the capability of this suffix to exploit the mechanisms of metaphor and metonymy that enhances its productivity, and this becomes even more clear when its behavior is compared to two of its rival (and less productive) agent suffixes, i.e. *-ant/ent* (e.g. *servant*) and *-ist* (e.g. *specialist*), which show little polysemy compared to *-er* formations (Panther & Thornburg 2003: 314–315).

This assumption seems consistent with the findings of the analysis presented throughout Chapters 3, 4 and 5. The constructions showing limited polysemy patterns (cf. Section 3.6) also show limited degrees of productivity, that is, their productivity is restricted to specific constructions, often based on few verbal stems (e.g. the suffix *-lec*, whose productivity is limited to very few specific constructions). The constructions showing larger polysemy patterns, on the other hand, appear to be subject to fewer restrictions on their productivity.

The comparison of the constructions [X-*vodec*] and [X-*vod*] has added further evidence in support of this assumption, by showing that the more productive construction [X-*vod*] is more polysemous with respect to the less productive [X-*vodec*].

5.9. Summary

The present chapter has dealt with the question of productivity in word-formation, starting from the assumption that rival word-formation constructions show varying degrees of productivity in modern Russian.

Although the assessing of the productivity rates of rival constructions proved problematic, due to the nature of the data at my disposal (cf. Sections 5.2 and 5.3), I have shown that rival agentive constructions display varying degrees of productivity when the diachronic aspect of this phenomenon is taken into account (cf. Section 5.4). Moreover, the case studies on low-frequency constructions described in Section 5.3 have demonstrated that calculations of productivity strongly rely on the process of data retrieval and on the dimension of the corpus employed.

In Section 5.6, I have shown that the productivity of rival constructions varies according to restrictions that concern the nature of the compounds' constituents. The main findings in these respects can be summarized as follows:

i. the productivity of certain constructions is restricted to cases in which the verbal base plus the suffix form an existing deverbal outside compounding;

ii. the productivity of certain constructions is restricted to a small number of verbal bases.

The case studies presented in Section 5.7 comparing the productivity of the partially specified constructions [X-*ljubec*]–[X-*ljub*] and [X-*vodec*]–[X-*vod*] has provided further evidence showing that the strongest opposition occurring between the construction in *-ec* and the suffixless construction manifests itself in their diverging productivity patterns. These case studies have also supported the claim that measurements of productivity are significantly affected by the process of data retrieval and the type of corpora employed. Finally, the case study on [X-*vodec*]–[X-*vod*] has also demonstrated that productivity and polysemy are not mutually exclusive phenomena. Rather, polysemy appears to be one of the factors enhancing the productivity of word-formation constructions.

Concluding remarks

The goal of the present study has been to compare rival word-formation constructions giving rise to compound agent nouns in Russian. The results of this research contribute, on the one hand, to enriching studies on Russian word-formation, which often disregard the phenomenon of compounding. On the other hand, my research fits into theoretical discussions concerning issues such as the polysemy of word-formation constructions and questions of rivalry among constructions sharing the same functions. The corpus-based analyses carried out in Chapters 4 and 5 demonstrate the benefits of combining qualitative descriptions with quantitative investigations in studies dealing with word-formation, especially when it comes to determining productivity.

In Chapter 1, I have first introduced the phenomenon of compounding in general, by addressing questions such as the the definition and delimitation of the concept of "compound" (Section 1.1), and the classification of compounds (Section 1.2). Then, consistently with the topic of this dissertation, I have focused in particular on synthetic and parasynthetic compounds (Sections 1.3 and 1.4). Finally, I have discussed the state of the art of studies dealing with the phenomenon of compounding in Russian (Section 1.5), and I have then proposed an overview of the different compound types in contemporary Russian (Section 1.6).

In Chapter 2, I have discussed the constructionist approach to grammar (Section 2.1) and, particularly, to morphology and word-formation (Section 2.2). I have shown how the idea of a hierarchical lexicon, along with the notions of schema unification and embedded productivity, can be employed to account for synthetic compounds (2.3), as I have demonstrated for Russian synthetic compounds with agentive meanings (2.4). Through an exemplification based on the construction in *-ec*, I have shown that a constructionist approach to the analysis of compounds allows representing regularities and subregularities of the schemas embedded in a hierarchy of form-meaning pairs.

In Chapter 3, I have examined the origin, etymology and use of the suffixes employed in compounding to form agent nouns, as represented in the RNC word-formation database that I have used to collect my data, i.e. the suffixes *-ec*, *-lec*, *-tel'*, *-nik*, *-ščik/čik*, *-l'ščik*, *-ka*, and *-lka* (Section 3.1). Starting from the observation that these

suffixes bring about quite large polysemy patterns, I have addressed the question of polysemy in word-formation, as it is accounted for in different theoretical frameworks (Sections 3.2 to 3.5). In Section 3.6, I have shown how the polysemy of such suffixes can be represented by making use of the analytical tools of CxM, such as the hierarchical representation of the lexicon. In Section 3.7, I have briefly discussed the question of conversion/zero derivation in word-formation, and I have described the polysemy pattern of the suffixless construction, which is the result of the combination of compounding and conversion.

In Chapter 4, I have compared rival constructions forming compound agent nouns in Russian. First, I have briefly described the RNC, which I have used to check the data and retrieve examples (Section 4.1). Then, I have described the process of selection of the data, which has led to the compilation of a database containing 831 compounds (Section 4.2; cf. Appendices 1 and 2). The compounds selected have been analyzed according to six parameters, i.e. the word class and semantic role of the compound's non-verbal element, the transitivity and aspect of the compound's verbal element, the animacy of the compound's referent and the semantics of the compound (Section 4.3). The main findings, which have also been confirmed by the statistical tests carried out in Section 4.4, can be summarized as follows:

i. the low-frequency constructions in *-lec*, *-l'ščik* and *-lka* show no variation at all, no matter which parameter is taken into consideration. Each of these constructions is associated with one meaning only: *-lec* with Carriers of State, *-l'ščik* with Prototypical Agents, *-lka* with Instruments. These constructions are usually based on a restricted number of base verbs, and new compounds are mainly produced by analogy with a base model, e.g. [X-*vladelec*] 'owner of X';

ii. the constructions showing higher frequencies show greater variation, but the degree of variation depends on the parameter considered. The parameters that better discriminate the distribution of rival constructions are:

a) the aspect of the compound's verbal element, which brings out the different behavior of *-tel'* with respect to the other constructions, in that it is the only construction that allows a high percentage of perfective verbal bases. In Section 4.3.4, I have argued that the correlation between the construction in *-tel'* and

perfective bases could be sought in the fact that compounds in *-tel'* are often based on prefixed perfective verbs that bring about specialized meanings (which is also consistent with the semantic features that I have identified for compounds in *-tel'*) and that constitute the most basic elements in aspectual pairs in which the imperfective is secondary;

b) the semantics of the compound (which is also strictly related to the parameter of animacy). In these respects, I have noticed diverging tendencies among the high-frequency constructions in *-ec*, *-tel'* and *-ø*. Whereas the construction in *-ec* and the suffixless construction tend to be associated with meanings of Human Agents (both Prototypical and Non-Prototypical) and less frequently with instrumental meanings, the construction in *-tel'* shows a significantly higher percentage of nouns denoting Instruments. The remaining medium-frequency constructions in *-nik*, *-ščik*/*čik* and *-ka* behave more similarly to *-tel'* and display high percentages of instrumental meanings.

The analyses carried out in Section 4.5 have shown that diachronic and stylistic features are also relevant factors to understand the distribution of rival agentive constructions. By resorting to the metatextual information provided by the RNC, I have checked the diachronic distribution of each compound in my database, as well as their distribution across different textual genres and registers (written vs. spoken texts). The most striking finding of these analyses regards the distribution of the two constructions that had resulted more similar to one another based on semantic criteria, i.e. the construction in *-ec* and the suffixless construction:

i. the construction in *-ec* shows significant overrepresentation in earlier texts and until the first half of the 20th century, while displaying underrepresentation starting from the second half of the 20th century. Conversely, the suffixless construction shows significant underrepresentation in earlier texts, while it is overrepresented especially in texts belonging to the 20th century;

ii. both constructions show underrepresentation in technical and official texts, which confirms their lower degree of specialization with respect to the other constructions and their semantic similarity. However, whereas the construction in *-ec* is overrepresented in

scientific/educational texts (due to the fact that this genre includes religious texts), the suffixless construction is overrepresented in fiction texts;

iii. the two constructions display opposite distributions in terms of register. Whereas the construction in *-ec* is overrepresented in written texts and underrepresented in spoken texts, the suffixless construction shows slight underrepresentation in written texts and significant overrepresentation in spoken texts.

Concerning the other constructions, I have observed a general tendency for them to show overrepresentation or normal representation in later texts and in specialized discourses (e.g. technical and official/business texts). The high-frequency construction in *-tel'* behaves similarly to *-ec* in some respects (especially in that it also shows some degree of overrepresentation in earlier written texts), but it substantially differs from *-ec* for its overrepresentation in technical and official/business texts.

The analysis of 29 doublets proposed in Section 4.6 has corroborated the results discussed so far. When the opposition between the two compounds forming the doublet is not semantic, in most cases the constructions involved are *-ec* and the suffixless construction, and the two constructions mainly differ in terms of diachrony and register.

In Chapter 5, I have addressed the question of productivity in word-formation. Based on the premise that the different constructions analyzed (partially) cover the same functions, I have assumed that their current degrees of productivity must be different. However, determining the productivity of these constructions has proved complicated, due to the nature of the data at my disposal.

As I have discussed in Sections 5.2 and 5.3, synchronic productivity cannot be determined based on dictionary data, since this type of data can only prove the past productivity of a certain construction. In Section 5.3, I have demonstrated that a bottom-up approach to data retrieval and the use of big corpora affect significantly calculations of productivity and lead to more reliable results.

The case studies carried out in Section 5.3 and regarding the low-frequency constructions in *-lec*, *-l'ščik* and *-lka* have also demonstrated that most of the times productivity seems to be restricted to specific constructions. For instance, we can say that, within the construction [X-LV-V-*lec*]$_N$, there are few productive subconstructions, such as [X-*vladelec*] 'owner of X'. Thus, the construction [X-LV-V-*lec*]$_N$ appears to be

productive only when the element V is filled with specific verb bases, such as *vladet'* 'own, possess'.

In Section 5.4, I have determined the diachronic productivity of the rival constructions examined based on the first occurrence of each compound in the RNC. The most significant results in this respect regard the high-frequency constructions in *-ec*, *-tel'* and *-ø*, for which more data are available. While the productivity of the construction in *-ec* appears to decrease over time, the construction in *-tel'* and the suffixless construction seem to increase their productivity, especially starting from the 20th century. As I have discussed, the "drop-effect" of the productivity of the suffixless construction during the second half of the 20th century (cf. Figure 29) should be interpreted as the result of a stabilization of the number of new occurrences with that construction rather than an actual drop in productivity. It is the sharp increase in the number of new occurrences in the first half of the 20th century that is more significant and has to be interpreted as the result of the new naming needs in the post-revolutionary period.

Apart from diachrony, textual genres and registers also affect the productivity of rival constructions (Section 5.5). Although thorough hapax-based calculations of productivity could not be carried out due to the dictionary-based nature of the data, the analyses presented in Section 4.5 point to the existence of varying degrees of productivity for rival constructions depending on the textual genre and the register.

In Section 5.6, I have shown that productivity is also subject to restrictions depending on the nature of the compounds' constituents. Remarkably, most of the constructions considered appear to be productive only when the verbal base followed by the agentive suffix form an existing noun outside compounding (especially *-lec* and *-l'ščik*, followed by *-tel'*, *-ščik*/*čik*, *-lka*, and *-nik*), while other constructions do not show such restrictions (*-ec*, *-ka*, *-ø*). Moreover, the productivity of certain constructions is restricted to specific verbal bases, as already discussed for the construction in *-lec*, whose productivity appears to be limited to the occurrence of the verb *vladet'* 'own, possess'.

In Section 5.7, I have compared the productivity of two pairs of rival partially specified constructions, i.e. [X-*ljubec*]–[X-*ljub*] and [X-*vodec*]–[X-*vod*], which has allowed reinforcing some of the findings already discussed:

i. that the diachronic productivity of the rival constructions in *-ec* and *-ø* shows diverging tendencies, *-ø* becoming more productive than *-ec* over time;

ii. that the process of data retrieval and the type of corpora employed significantly affect measurements of productivity;

iii. that larger polysemy patterns can contribute to higher degrees of productivity.

The findings of this study open the way to possible future research on Russian compounds. First, other groups of rival constructions forming compounds (especially nominal and adjectival compounds) still need to be studied. The same analysis proposed in this dissertation could be applied to such other constructions, which would give us a more complete picture of the phenomenon of compounding in Russian. At the same time, the constructions that have been analyzed in this research could be further investigated by applying a consistent bottom-up approach to data retrieval. Of course, this task could be accomplished much more easily and quickly if compounds were annotated in corpora. In my view, the annotation of compounds in corpora is one of the most important and urgent tasks to be fulfilled in the near future, as this would result in the possibility of carrying out large data-oriented investigations on compounds that have not been realized so far.

References

Apresjan, Jurij D. 1974. Regular Polysemy. *Linguistics* 142: 5–32.

Arcodia, Giorgio F. 2014. Diachrony and the polysemy of derivational affixes. In: Franz Rainer, Francesco Gardani, Hans C. Luschützky & Wolfgang U. Dressler (eds.), *Morphology and Meaning. Selected papers from the 15th International Morphology Meeting, Vienna, February 2012*. Amsterdam/Philadelphia: John Benjamins. 127–140.

Aronoff, Mark. 1976. *Word Formation in Generative Grammar*. Cambridge, MA: The MIT Press.

Arppe, Antti. 2008. *Univariate, bivariate and multivariate methods in corpus-based lexicography. A study of synonymy*. PhD dissertation, University of Helsinki.

Baayen, Harald R. 1992. Quantitative aspects of morphological productivity. In: Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1991*. Amsterdam: Springer. 109–149.

Baayen, Harald R. 1993. On frequency, transparency and productivity. In: Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1992*. Dordrecht: Kluwer. 181–208.

Baayen, Harald R. 1994a. Derivational Productivity and Text Typology. *Journal of Quantitative Linguistics* 1: 16–34.

Baayen, Harald R. 1994b. Productivity in language production. *Language and Cognitive Processes* 9 (3): 447–469.

Baayen, Harald R. 2001. *Word frequency distributions*. Dordrecht: Kluwer.

Baayen, Harald R. 2008. *Analyzing Linguistic Data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, Harald R. & Rochelle Lieber. 1991. Productivity and English derivation: A corpus-based study. *Linguistics* 29 (5): 801–843.

Baayen, Harald R., Rochelle Lieber & Robert Schreuder. 1997. The morphological complexity of simplex nouns. *Linguistics* 35: 861–877.

Baayen, Harald R., Anna Endresen, Laura A. Janda, Anastasia Makarova & Tore Nesset. 2013. Making choices in Russian: pros and cons of statistical methods for rival forms. *Russian Linguistics* 37 (3): 253–291.

Barðdal, Jóhanna. 2008. *Productivity. Evidence from Case and Argument Structure in Icelandic*. Amsterdam/Philadelphia: John Benjamins.

Barðdal, Jóhanna, Elena Smirnova, Lotte Sommerer & Spike Gildea. 2015. (eds.) *Diachronic Construction Grammar*. Amsterdam/Philadelphia: John Benjamins.

Baroni, Marco & Stefan Evert. 2014. *The zipfR package for lexical statistics: A tutorial introduction*. See http://zipfr.r-forge.r-project.org/materials/zipfr-tutorial.pdf.

Barulin, Aleksandr N. 2017. O statuse i semantike tak nazyvaemogo "interfiksa" O/E v russkich kompozitach [On the status and semantics of the so-called interfix O/E in Russian compounds]. *Lingvistika i metodika prepodavanija inostrannych jazykov 9. Problemy opisanija jazyka* [Linguistics and methodology of foreign languages' teaching 9. Problems of language description]. 79–93.

Barulin, Aleksandr N. & Aleksandra Ju. Ajchenval'd. 1988. K grammatike sinteza form slova [Toward a grammar of the synthesis of word forms]. In: Vera I. Podlesskaja (ed.), *Sinchronija i diachronija v lingvističeskich issledovanijach* [Synchrony and diachrony in linguistic research], Volume 1. Moskva: Nauka. 36–44.

Basilio, Margarida. 2006. Metaphor and metonymy in word-formation. *DELTA: Revista de Documentaçao de Estudios em Lingüística Teórica e Aplicada* 22: 67–80.

Bauer, Laurie. 1983. *English word-formation*. Cambridge: Cambridge University Press.

Bauer, Laurie. 2001a. *Morphological productivity*. Cambridge: Cambridge University Press.

Bauer, Laurie. 2001b. Compounding. In: Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language Typology and Language Universals*. Berlin/New York: Mouton de Gruyter. 695–707.

Bauer, Laurie. 2003. *Introducing Linguistic Morphology*. Washington (DC): Georgetown University Press.

Bauer, Laurie. 2005. Productivity theories. In: Pavol Štekauer & Rochelle Lieber (eds.), *Handbook of Word-Formation*. Dordrecht: Springer. 315–334.

Bauer, Laurie. 2006. Competition in English Word Formation. In: Ans van Kemenade & Bettelou Los, *The Handbook of the History of English*. Padstow: Blackwell. 177–198.

Beard, Robert. 1995. *Lexeme morpheme base morphology*. Albany (NY): SUNY Press.

Bencini, Giulia M. L. 2013. Psycholinguistics. In: Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press. 379–396.

Benigni, Valentina & Francesca Masini. 2009. Compounds in Russian. *Lingue e Linguaggio* 8 (2): 171–193.

Bergen, Benjamin & Nancy Chang. 2013. Embodied Construction Grammar. In: Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press. 168–190.

Bilynsky, Michael. 2015. The productivity of deverbal categories and suffixal models within shared-root deverbal paradigms as reflected in the Oxford English Dictionary. *SKASE Journal of Theoretical Linguistics* 13 (2): 32–45.

Bisetto, Antonietta & Sergio Scalise. 2005. The classification of compounds. *Lingue e Linguaggio* 4 (2): 319–332.

Bisetto, Antonietta & Chiara Melloni. 2008. Parasynthetic compounding. *Lingue e Linguaggio* 7 (2): 233–259.

Boas, Hans C. 2010. (ed.) *Contrastive Studies in Construction Grammar*. Amsterdam/Philadelphia: John Benjamins.

Boas, Hans C. & Ivan A. Sag. 2012. Sign-based Construction Grammar. Stanford: CSLI.

Bogdanov, Aleksej V. 2011. *Semantika i sintaksis otglagol'nych ad''ektivov*. Dissertacija na soiskanie učenoj stepeni kandidata filologičeskich nauk, 10.02.19 Teorija

jazyka [The semantics and syntax of deverbal adjectives. PhD dissertation]. Moskovskij gosudarstvennyj universitet im. M. V. Lomonosova.

Booij, Geert. 1977. *Dutch Morphology: A Study of Word Formation in Generative Grammar*. Dordrecht: Foris.

Booij, Geert. 1986. Form and meaning in morphology: the case of Dutch 'agent nouns'. *Linguistics* 24: 503–518.

Booij, Geert. 1988. The relation between inheritance and argument structure: deverbal -er-nouns in Dutch. In: Martin Everaert, Arnold Evers, Rini Huybregts & Mieke Trommelen (eds.), *Morphology and Modularity. In Honour of Henk Schultink*. Dordrecht: Foris. 57–74.

Booij, Geert. 2002a. Constructional Idioms, Morphology and the Dutch Lexicon. *Journal of Germanic Linguistics* 14 (4): 301–329.

Booij, Geert. 2002b. Separable Complex Verbs in Dutch: A Case of Periphrastic Word Formation. In: Nicole Dehé, Ray Jackendoff, Andrew McIntyre & Silke Urban (eds.), *Verb-Particle Explorations*. Berlin: Mouton de Gruyter. 21–42.

Booij, Geert. 2005a. *The Grammar of Words. An Introduction to Linguistic Morphology*. Oxford/New York: Oxford University Press.

Booij, Geert. 2005b. Compounding and Derivation. Evidence for Construction Morphology. In: Wolfgang U. Dressler, Dieter Kastovsky, Oskar E. Pfeiffer & Franz Rainer (eds.), *Morphology and Its Demarcations: Selected Papers from the 11th Morphology Meeting, Vienna, February 2004*. Amsterdam/Philadelphia: John Benjamins. 109–132.

Booij, Geert. 2007. Polysemy and Construction Morphology. In: Fons Moerdijk, Ariane van Santen & Rob Tempelaars (eds.), *Leven met woorden*. Leiden: Instituut voor Nederlandse Lexicologie. 355–364.

Booij, Geert. 2009. Compounding and Construction Morphology. In: Rochelle Lieber & Pavol Štekauer, *The Oxford Handbook of Compounding*. New York: Oxford University Press. 201–216.

Booij, Geert. 2010. *Construction Morphology*. New York: Oxford University Press.

Booij, Geert. 2013. Morphology in Construction Grammar. In: Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press. 255–274.

Booij, Geert. 2015. Word-formation in Construction Grammar. In: Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-formation. An International Handbook of the Languages of Europe*. Volume 1. Berlin: De Gruyter, 188–202.

Breiman, Leo. 2001. Random Forests. *Machine Learning* 45 (1): 5–32.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & Harald R. Baayen. 2007. Predicting the dative alternation. In: Gerlof Bouma, Irene Krämer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*. Amsterdam: 7 Royal Netherlands Academy of Arts and Sciences. 69–94.

Briscoe, Edward, Ann Copestake & Valeria de Paiva. 1993. (eds.) *Inheritance, defaults and the lexicon*. Cambridge: Cambridge University Press.

Broccias, Cristiano. 2013. Cognitive Grammar. In: Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press. 191–210.

Bybee, Joan L. 1985. *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam/Philadelphia: John Benjamins.

Bybee, Joan L. 1995. Diachronic and typological properties of morphology and their implications for representation. In: Laurie B. Feldman (ed.), *Morphological Aspects of Language Processing*. Hillsdale (NJ): Erlbaum. 225–246.

Bybee, Joan L. 2006. From Usage to Grammar: The Mind's Response to Repetition. *Language* 82 (4): 711–733.

Bybee, Joan L. 2013. Usage-based Theory and Exemplar Representations of Constructions. In: Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press. 49–69.

Chappell, Hilary & William McGregor. 1996. (eds.) *The grammar of inalienability*. Berlin: Mouton de Gruyter.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: The MIT Press.

Chovanová, Iveta & Pavel Štichauer. 2014. Possessive compounds in Slavic and the Principle of Integrated Meronymy. In: Franz Rainer, Francesco Gardani, Hans C. Luschützky & Wolfgang U. Dressler (eds.), *Morphology and Meaning. Selected papers from the 15th International Morphology Meeting, Vienna, February 2012*. Amsterdam/Philadelphia: John Benjamins. 141–152.

Comrie, Bernard. 1989. *Language Universals and Linguistic Typology. Syntax and Morphology*. Chicago: The University of Chicago Press.

Corbin, Danielle. 1987. *Morphologie dérivationelle et structuration du lexique*. Tübingen: Niemeyer.

Croft, William A. 2001. *Radical Construction Grammar*. Oxford: Oxford University Press.

Croft, William A. 2013. Radical Construction Grammar. In: Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press. 211–232.

Croft, William A., Jóhanna Barðdal, Willem B. Hollmann, Violeta Sotirova & Chiaki Taoka. 2010. Revising Talmy's Typological Classification of Complex Events. In: Hans C. Boas, *Contrastive Studies in Construction Grammar*. Amsterdam/Philadelphia: John Benjamins. 201–235.

Croft, William A. & D. Alan Cruse. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Daniel, Michael. 2014. Against the Addressee of speech – Recipient metaphor: Evidence from East Caucasian. In: Silvia Luraghi & Heiko Narrog (eds.), *Perspectives on Semantic Roles.* Amsterdam/Philadelphia: John Benjamins. 205–240.

Deshors, Sandra C. & Stefan Th. Gries. 2014. A case for the multifactorial assessment of learner language. The uses of may and can in French-English interlanguage. In: Dylan Glynn and Justyna A. Robinson (eds.), *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy*. Amsterdam/Philadelphia: John Benjamins. 179–204.

Diessel, Holger. 2013. Construction Grammar and First Language Acquisition. In: Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press. 347–364.

Dirven, René. 1999. Conversion as a conceptual metonymy of event schemata. In: Klaus-Uwe Panther & Günter Radden (eds.), *Metonymy in language and thought*. Amsterdam/Philadelphia: John Benjamins. 275–287.

Di Sciullo, Anna M. 1992. Deverbal compounds and the external argument. In: Iggy M. Roca, *Thematic Structure. Its Role in Grammar*. Berlin/New York: Foris Publications. 65–72.

Di Sciullo, Anna M. 2005. Decomposing Compounds. *SKASE Journal of Theoretical Linguistics* 2 (3): 14–33.

Di Sciullo, Anna M. & Edwin Williams. 1987. *On the definition of word*. Cambridge, MA: The MIT Press.

Dressler, Wolfgang U. 2006. Compound Types. In: Gary Libben & Gonia Jarema (eds.), *The Representation and Processing of Compound Words*. New York: Oxford University Press. 23–44.

Efimova, Valerija S. 2006. *Staroslavjanskaja slovoobrazovatel'naja morfemika* [Old Slavic word-formation morphemics]. Moskva: Institut slavjanovedenija RAN.

Efremova, Tat'jana F. 2000. *Novyj slovar' russkogo jazyka. Tolkovo-slovoobrazovatel'nyj* [New dictionary of the Russian language. Explanatory and word-formational]. Moskva: Russkij jazyk. See https://gufo.me/dict/efremova.

Efthymiou, Angeliki, Georgia Fragaki & Angelos Markos. 2012. Productivity of verb-forming suffixes in Modern Greek: A corpus-based study. *Morphology* 22: 515–543.

Ellis, Nick C. 2013. Construction Grammar and Second Language Acquisition. In: Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press. 365–378.

Endresen, Anna. 2014. *Non-Standard Allomorphy in Russian Prefixes: Corpus, Experimental, and Statistical Exploration*. PhD dissertation, University of Tromsø: The Arctic University of Norway.

Evans, Roger & Gerald Gazdar. 1996. DATR: a language for lexical knowledge representation. *Computational Linguistics* 22: 167–216.

Fabb, Nigel. 1998. Compounding. In: Andrew Spencer & Arnold M. Zwicky (eds.), *The Handbook of Morphology*. Oxford/Malden: Blackwell. 66–83.

Fabiszak, Malgorzata, Anna Hebda, Iwona Kokorniak & Karolina Krawczak. 2014. The semasiological structure of Polish myśleć 'to think'. A study in verb-prefix semantics. In: Dylan Glynn and Justyna A. Robinson (eds.), *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy*. Amsterdam/Philadelphia: John Benjamins. 223–251.

Fanselow, Gisbert. 1981. *Zur Syntax und Semantik der Nominalkomposition.* Tübingen: Niemeyer.

Fedorova, Ljudmila L. 2004. Problemy leksikografičeskogo opisanija složnych slov [Problems of the lexicographic description of compounds]. In: Leonid P. Krysin (ed.), *Russkij jazyk segodnja 3. Problemy russkoj leksikografii* [The Russian language today 3. Problems of Russian Lexicography]. Moskva: Rossijskaja Akademija Nauk, Institut russkogo jazyka im. V. V. Vinogradova. 296–311.

Fedorova, Ljudmila L. 2007. Malen'kie komedii složnych slov [Little comedies of compound words]. In: Nina D. Artjunova (ed.), *Logičeskij analiz jazyka. Jazykovye mechanizmy komizma* [Logic analysis of the language. Comic mechanisms in the language]. Moskva: Rossijskaja Akademija Nauk, Institut jazykoznanija. 699–711.

Fedorova, Ljudmila L. 2008. Skrytyj dialogizm složnych slov [The hidden dialogism of compound words]. In: Igor' A. Šaronov (ed.), *Skrytye smysly v jazyke i kommunikacii* [Hidden meanings in language and communication]. Moskva: Rossijskij Gosudarstvennyj Gumanitarnyj Universitet, Institut lingvistiki. 101–110.

Fedorova, Ljudmila L. 2010. Notes on co-compounds and paired words. In: Robert J. Fourser (ed.), *Contemporary Korean Linguistics. International Perspectives: In Honor of Professor Sang-Oak Lee*. Seoul: Thaehaksa. 63–80.

Fedorova, Ljudmila L. 2012. O smechovoj prirode modeli složnych slov-prozvišč v slavjanskich i romanskich jazykach [On the humorous nature of compounds-nicknames in the Slavic and Romance languages]. In: Marina V. Pimenova (ed.), *Vostočnoslavjanskie jazyki i literatury v istoričeskom i kul'turnom kontekstach: kognitivnaja lingvistika i konceptual'nye issledovanija (sbornik naučnych statej)* [East-Slavic languages and literatures in historical and cultural contexts: cognitive linguistics and conceptual studies (a collection of scientific papers)]. Kiev: IJa im. Potebin NAN Ukrainy. 380–387.

Fedorova, Ljudmila L. 2013. Gusi-lebedi, soroki-vorony, cvety i pticy: zametki o parnych slovach [Geese-swans, magpies-ravens, flowers and birds: notes on paired words]. In: Aleksandr E. Kibrik (ed.), *Lingvističeskij bespredel 2. Sbornik naučnych trudov k jubileju A. I. Kuznecovoj* [Linguistic mayhem 2. A collection of scientific papers for the birthday of A. I. Kuznecova]. Moskva: Moskovskij Gosudarstvennyj Universitet im. M. V. Lomonosova, Filologičeskij fakul'tet. 318–333.

Fedorova, Ljudmila L. 2014a. Kognitivnye mechanizmy slovosloženija [Cognitive mechanisms of compounding]. *Kognitivnye issedovanija jazyka* 16: 75–87.

Fedorova, Ljudmila L. 2014b. Za tridevjat' zemel', za tridevjat' morej: primery "naivnoj arifmetiki" v strukturach jazyka [Far, far away (lit. beyond twenty-seven lands, beyond twenty-seven seas): examples of "naive arithmetics" in the structures of language]. In: Nina D. Artjunova (ed.), *Logičeskij analiz jazyka. Čislovoj kod v raznych jazykach i kul'turach* [Logic analysis of the language. The numeric code in different languages and cultures]. Moskva: Rossijskaja Akademija Nauk, Institut jazykoznanija. 488–499.

Fedorova, Ljudmila L. 2015. Složnye prilagatel'nye neotčuždaemoj prinadležnosti v russkom jazyke [Compound adjectives expressing inalienable possession in Russian]. *Moskovskij lingvističeskij žurnal* 17 (2): 61–74.

Fillmore, Charles. 1982. Frame Semantics. In: Linguistic Society of Korea (ed.), *Linguistics in the Morning Calm*. Seoul: Hanshin. 111–138.

Fillmore, Charles. 2013. Berkeley Construction Grammar. In: Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press. 111–132.

Fillmore, Charles & Collin Baker. 2010. A Frames Approach to Semantic Description. In: Bernd Heine & Heiko Narrog (eds.), *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press. 313–339.

Fillmore, Charles, Paul Kay & Mary Catherine O'Connor. 1988. Regularity and Idiomaticity in Grammatical Constructions: The Case of *Let Alone*. *Language* 64 (3): 501–538.

Gaeta, Livio. 2006. Lexical integrity as a constructional strategy. *Lingue e Linguaggio* 5 (1): 67–82.

Gaeta, Livio. 2010. Synthetic compounds: With special reference to German. In: Sergio Scalise & Irene Vogel (eds.), *Cross-Disciplinary Issues in Compounding*. Amsterdam/Philadelphia: John Benjamins. 219–236.

Gaeta, Livio & Davide Ricca. 2006. Productivity in Italian word formation: A variable-corpus approach. *Linguistics* 44 (1): 57–89.

Gaeta, Livio & Davide Ricca. 2015. Productivity. In: Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-formation. An international handbook of the languages of Europe*. Berlin: Mouton de Gruyter. 842–858.

Glynn, Dylan. 2014. The many uses of run. Corpus methods and Socio-Cognitive Semantics. In: Dylan Glynn and Justyna A. Robinson (eds.), *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy*. Amsterdam/Philadelphia: John Benjamins. 117–144.

Goldberg, Adele. 1995. *Constructions: A Construction Grammar Approach to Argument Structures*. Chicago: The University of Chicago Press.

Goldberg, Adele. 2006. *Constructions at Work*. Oxford: Oxford University Press.

Grandi, Nicola. 2006. Considerazioni sulla definizione e la classificazione dei composti. *Annali Online di Ferrara – Lettere* 1: 35–52.

Gries, Stefan Th. 2013. *Statistics for Linguistics with R. A Practical Introduction*. Berlin: Mouton de Gruyter.

Gries, Stefan Th. & Stefanie Wulff. 2005. Do Foreign Language Learners Also Have Constructions? Evidence from Priming, Sorting and Corpora. *Annual Review of Cognitive Linguistics* 3: 182–200.

Grimshaw, Jane. 1990. *Argument structure*. Cambridge, MA: MIT Press.

Grišina, Elena, Il'ja Itkin, Ol'ga Ljaševskaja & Marija Tagabileva. 2009. O zadačach i metodach slovoobrazovatel'noj razmetki v korpuse tekstov [On the tasks and methods of the word-formation annotation in a corpus of texts]. *Poljarnyj Vestnik* 12: 5–25.

Haiman, John. 1983. Iconic and economic motivation. *Language* 59 (4): 781–819.

Halle, Morris. 1973. Prolegomena to a Theory of Word-Formation. *Linguistic Inquiry* 4: 451–464.

Haspelmath, Martin. 2002. *Understanding morphology*. London: Arnold.

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

Hay, Jennifer. 2001. Lexical frequency in morphology: is everything relative? *Linguistics* 39: 1041–1070.

Hay, Jennifer & Harald R. Baayen. 2002. Parsing and productivity. In: Geert Booij & Jaap Van Marle (eds.), *Yearbook of Morphology 2001*. Dordrecht: Springer. 203–235.

Heine, Bernd. 1997. *Cognitive Foundations of Grammar*. New York/Oxford: Oxford University Press.

Hilpert, Martin. 2008. *Germanic Future Constructions. A Usage-based Approach to Language Change*. Amsterdam: John Benjamins.

Hilpert, Martin & Jan-Ola Östman. 2014. Introduction: Reflections on Constructions Across Grammars. In: Martin Hilpert & Jan-Ola Östman (eds.), *Reflections on*

Constructions Across Grammars. *Constructions and Frames* 6 (2). Special Issue. 137–142.

Hoffmann, Thomas & Graeme Trousdale. 2013. (eds.) *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press.

Insacco, Gioia. 2014. *Strutture argomentali e cicli lessicali delle nominalizzazioni italiane*. PhD dissertation. Università degli Studi Roma Tre.

Itkin, Il'ja B. 1996. Morfonogičeskie modeli russkogo otglagol'nogo slovoobrazovanija [Morphonological models of Russian deverbal word-formation]. *Rusistika Segodnja* 4: 17–44.

Itkin, Il'ja B. 2013. V poiskach nulevogo slovoobrazovatel'nogo suffiksa (Otglagol'nye suščestvitel'nye tipa *zvon*, *šum*, *šelest* v sovremennom russkom jazyke) [In search of the zero suffix in word-formation (Deverbal nouns of the type *zvon*, *šum*, *šelest* in contemporary Russian)]. *Russkij jazyk v naučnom osveščenii* 2 (26): 52–64.

Jackendoff, Ray. 1997. *The Architecture of the Language Faculty*. Cambridge, MA: The MIT Press.

Jackendoff, Ray. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Jackendoff, Ray. 2013. Constructions in the Parallel Architecture. In: Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press. 70–92.

Janda, Laura A. 2008. Metonymy via Perfectivization of Russian Verbs. *Slavica Helsingiensia* 35: 77–85.

Janda, Laura A. 2010. Russian word-formation in contrast with Czech and Norwegian. In: Alte Grønn & Irena Marijanovic (eds.), *Russian in Contrast*, Oslo Studies in Language 2 (2): 243–259.

Janda, Laura A. 2011. Metonymy in word-formation. *Cognitive Linguistics* 22 (2): 359–392. Dataset available at: https://opendata.uit.no/dataset.xhtml?persistentId=hdl:10037.1/10020.

Janda, Laura A. 2014. Metonymy and word-formation revisited. *Cognitive Linguistics* 25 (2): 341–349.

Janko-Trinickaja, Nadija A. 2001. *Slovoobrazovanie v sovremennom russkom jazyke* [Word-formation in contemporary Russian]. Moskva: Indrik.

Kang, Beom-mo. 2016. The alternative negative constructions in spoken and written Korean: Logistic regression analysis. *Corpus Linguistics and Linguistic Theory* (aop). See https://www.degruyter.com/downloadpdf/j/cllt.ahead-of-print/cllt-2016-0021/cllt-2016-0021.pdf.

Kapatsinki, Vsevolod & Cynthia Vakareliyska. 2013. [N[N]] compounds in Russian. A growing family of constructions. *Constructions and Frames* 5 (1): 69–87.

Kay, Paul. 1997. Construction Grammar. In: Paul Kay, *Words and the Grammar of Context*. Stanford: CSLI. 123–131.

Kay, Paul & Charles Fillmore. 1999. Grammatical Constructions and Linguistic Generalizations: The *What's X Doing Y?* Construction. *Language* 75 (1): 1–33.

Kiefer, Ferenc. 1993. Thematic roles and compounds. *Folia Linguistica* 27 (1–2): 45–55.

Kiefer, Ferenc. 2000. Regularity. In: Geert Booij, Christian Lehmann & Joachim Mugdan (eds.), *Morphologie: Ein Internationales Handbuch Zur Flexion und Wortbildung, Volume 1*. Berlin/New York: De Gruyter. 296–302.

Killbury, James, Wiebke Petersen & Christoph Rumpf. 2006. Inheritance-based models of the lexicon. In: Dieter Wunderlich (ed.), *Advances in the theory of the lexicon*. Berlin: Mouton de Gruyter. 429–480.

Kiparsky, Paul. 1982. From Cyclic Phonology to Lexical Phonology. In: Harry van der Hulst & Norval Smith (eds.), *The Structure of Phonological Representations 1*. 131–175.

Kittilä, Seppo, Katja Västi & Jussi Ylikoski. 2011. Introduction to case, animacy and semantic roles. In: Seppo Kittilä, Katja Västi & Jussi Ylikoski (eds.). *Case, Animacy and Semantic Roles*. Amsterdam/Philadelphia: John Benjamins. 1–28.

Klavan, Jane. 2014. A multifactorial corpus analysis of grammatical synonymy. The Estonian adessive and adposition peal 'on'. In: Dylan Glynn and Justyna A.

Robinson (eds.), *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy*. Amsterdam/Philadelphia: John Benjamins. 253–278.

Koch, Peter. 1999. On the cognitive bases of metonymies and certain types of word formation. In: Klaus-Uwe Panther & Günter Radden (eds.), *Metonymy in language and thought*. Amsterdam/Philadelphia: John Benjamins. 139–167.

Koenfoed, Geert & Jaap van Marle. 2000. Productivity. In: Geert Booij, Christian Lehmann & Joachim Mugdan (eds.), *Morphologie: Ein Internationales Handbuch Zur Flexion und Wortbildung, Volume 1*. Berlin/New York: De Gruyter. 303–311.

Kövecses, Zoltán. 2006. *Language, Mind and Culture: A practical introduction*. Oxford: Oxford University Press.

Kövecses, Zoltán. 2010. *Metaphor. A Practical Introduction*. Oxford: Oxford University Press.

Kuznecov, Sergej A. (ed.). 1998. *Bol'šoj tolkovyj slovar' russkogo jazyka* [Big explanatory dictionary of the Russian language]. Sankt-Petersburg: Norint. See http://gufo.me/kuznec_a.

Kuznecova, Ariadna I. & Tat'jana F. Efremova. 1986. *Slovar' morfem russkogo jazyka* [Dictionary of morphemes of the Russian language]. Moskva: Russkij Jazyk.

Lakoff, George. 1987. *Women, fire, and dangerous things. What categories reveal about the Mind*. Chicago/London: The University of Chicago Press.

Langacker, Ronald. 1987. *Foundations of Cognitive Grammar: Volume 1: Theoretical Prerequisites*. Stanford: Stanford University Press.

Langacker, Ronald. 1991. *Foundations of Cognitive Grammar: Volume 2: Descriptive Application*. Stanford: Stanford University Press.

Langacker, Ronald. 2008. *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.

Lascarides, Alex & Ann Copestake. 1999. Default representation in constraint-based frameworks. *Computational Linguistics* 25: 55–106.

Levshina, Natalia. 2015. *How to do Linguistics with R. Data exploration and statistical analysis*. Amsterdam/Philadelphia: John Benjamins.

Levshina, Natalia, Dirk Geeraerts & Dirk Speelman. 2014. Dutch causative constructions. Quantification of meaning and meaning of quantification. In: Dylan Glynn & Justyna A. Robinson (eds.), *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy*. Amsterdam/Philadelphia: John Benjamins. 205–221.

Lieber, Rochelle. 1983. Argument linking and compounds in English. *Linguistic Inquiry* 14: 251–286.

Lieber, Rochelle. 2004. *Morphology and lexical semantics*. Cambridge: Cambridge University Press.

Lieber, Rochelle & Pavol Štekauer. 2009. Introduction: Status and Definition of Compounding. In: Rochelle Lieber & Pavol Štekauer, *The Oxford Handbook of Compounding*. New York: Oxford University Press. 3–18.

Lohmann, Arne. 2013. Is tree hugging the way to go? Classification trees and random forests in linguistic study. *VIEWS* 22. See https://anglistik.univie.ac.at/fileadmin/user_upload/dep_anglist/weitere_Uploads/Views/VIEWS_22_2013_Lohmann.pdf.

Luraghi, Silvia. 1995. Prototypicality and agenthood in Indo-European. In: Henning Andersen (ed.), *Historical Linguistics 1993*. Amsterdam/Philadelphia: John Benjamins. 259–268.

Luraghi, Silvia & Heiko Narrog. 2014. Perspectives on semantic roles: An introduction. In: Silvia Luraghi & Heiko Narrog (eds.), *Perspectives on Semantic Roles.* Amsterdam/Philadelphia: John Benjamins. 1–22.

Luschützky, Hans C. 2011. Agent-noun polysemy in Slavic: some examples. *STUF, Akademie Verlag* 64 (1): 75–95.

Luschützky, Hans C. & Franz Rainer. 2011. Agent-noun polysemy in a cross-linguistic perspective. *STUF, Akademie Verlag* 64 (4): 287–338.

Luschützky, Hans C. & Franz Rainer. 2013. Instrument and place nouns: A typological and diachronic perspective. *Linguistics* 51 (6): 1301–1359.

Lychyk, Victor. 1995. Russian Agentive Noun Formation in the 1970s. *Canadian Slavonic Papers* / *Revue Canadienne des Slavistes* 37 (1/2): 137–161.

Marchand, Hans. 1960. *The Categories and Types of Present-Day English Word-Formation*. Wiesbaden: Otto Harrassowitz.

Marchand, Hans. 1967. Expansion, Transposition and Derivation. *La Linguistique* 1: 3–26.

Masini, Francesca. 2016. *Grammatica delle Costruzioni. Un'introduzione*. Roma: Carocci.

Meillet, Antoine. 1905. *Études sur l'étimologie et le vocabulaire du vieux slave*. *Partie II*. Paris: Bouillon.

Mel'čuk, Igor' A. 1975/1997. Opyt razrabotki fragmenta sistemy ponjatij i terminov dlja morfologii (k formalizacii jazyka lingvistiki) [The experience of elaborating a fragment of the system of concepts and terms for morphology (toward the formalization of the language of linguistics)]. *Semiotika i informatika* 35: 15–58.

Melloni, Chiara & Antonietta Bisetto. 2010. Parasynthetic compounds: Data and theory. In: Sergio Scalise & Irene Vogel (eds.), *Cross-Disciplinary Issues in Compounding*. Amsterdam/Philadelphia: John Benjamins. 199–218.

Michaelis, Laura A. & Josef Ruppenhofer. 2001. *Beyond Alternations: A Constructional Model of the German Applicative Pattern*. Stanford, CA: CSLI Publications.

Michaelis, Laura A. 2013. Sign-Based Construction Grammar. In: Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press. 133–152.

Modor, Carol L. 1992. *Productivity and categorization in morphological classes*. PhD dissertation. SUNY at Buffalo.

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.

Molinsky, Steven J. 1973. *Patterns of Ellipsis in Russian Compound Noun Formations*. The Hague/Paris: Mouton.

Naccarato, Chiara. 2016. A corpus-based quantitative approach to the study of morphological productivity in diachrony: The case of *samo*-compounds in

Russian. In: Hanna Christ, Daniel Klenovšak, Lukas Sönning & Valentin Werner (eds.), *A blend of MaLT: Selected contributions from the Methods and Linguistic Theories Symposium 2015*. Bamberg: University of Bamberg Press. 133–152.

Naccarato, Chiara & Erica Pinelli. Forth. Un approccio cognitivo alla formazione delle parole: il caso dei composti con suffissi d'agente in russo. In: Marina di Filippo & François Esvan (eds.), *Studi di linguistica slava*. Napoli: Il Torcoliere.

Newmeyer, Frederick J. 2010. On comparative concepts and descriptive categories: A reply to Haspelmath. *Language* 86 (3): 688–695.

Ohori, Toshio. 2005. Construction Grammar as a Conceptual Framework for Linguistic Typology: A Case from Reference Tracking. In: Mirjam Fried & Hans C. Boas (eds.), *Grammatical Constructions: Back to the Roots*. Amsterdam/Philadelphia: John Benjamins. 215–237.

Olsen, Susan. 2000. Composition. In: Geert Booij, Christian Lehmann & Joachim Mugdan (eds.), *Morphologie: Ein Internationales Handbuch Zur Flexion und Wortbildung, Volume 1*. Berlin/New York: De Gruyter. 897–916.

Olsen, Susan. 2001. Copulative compounds: A closer look at the interface between morphology and syntax. In: Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 2000*. Amsterdam: Springer. 279–320.

Ožegov, Sergej I. & Natal'ja Ju. Švedova. (eds.). 1996. *Tolkovyj slovar' russkogo jazyka* [Explanatory dictionary of the Russian language]. Moskva: Az''. See https://gufo.me/dict/ozhegov.

Palmer, Chris C. 2015. Measuring productivity diachronically: nominal suffixes in English letters, 1400–1600. *English Language & Linguistics* 19 (1): 107–129.

Panther, Klaus-Uwe & Linda L. Thornburg. 2003. The roles of metaphor and metonymy in English *-er* nominals. In: René Dirven & Ralph Pörings (eds.), *Metaphor and Metonymy in Comparison and Contrast*. Berlin/New York: Mouton de Gruyter. 279–319.

Pazel'skaja, Anna G. 2013. Inkorporacija v glagol'nych formach v russkom jazyke [Incorporation in verb forms in Russian]. In: Vladimir P. Selegej et al. (eds.),

Computational Linguistics and Intellectual Technologies, Papers from the Annual International Conference "Dialogue", Issue 12, Volume 1. 579–591.

Peirsman, Yves & Dick Geeraerts. 2006. Metonymy as a prototypical category. *Cognitive Linguistics* 17 (3): 269–316.

Plag, Ingo. 1999. *Morphological productivity: Structural constraints in English derivation*. Berlin: Mouton de Gruyter.

Plag, Ingo. 2002. The role of selectional restrictions, phonotactics and parsing in constraining suffix ordering in English. In: Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 2001*. Amsterdam: Springer. 285–314.

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

Plag, Ingo. 2006. Productivity. In: Bas Aarts & April McMahon (eds.), *The handbook of English linguistics*. Oxford: Blackwell. 537–556.

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

Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Pulvermüller, Friedemann, Bert Cappelle & Yury Shtyrov. 2013. Brain Basis of Meaning, Words, Constructions, and Grammar. In: Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press. 397–416.

Rachilina, Ekaterina V. (ed.). 2010. *Lingvistika konstrukcij* [Construction Grammar]. Moskva: Azbukovnik.

Rainer, Franz. 2011. The agent-instrument-place "polysemy" of the suffix -TOR in Romance. *STUF, Akademie Verlag* 64 (1): 8–32.

Rainer, Franz. 2014. Polysemy in derivation. In: Rochelle Lieber & Pavol Štekauer (eds.), *The Handbook of Derivational Morphology*. Oxford: Oxford University Press. 338–353.

Rainer, Franz. 2015. Agent and instrument nouns. In: Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-formation. An International Handbook of the Languages of Europe*. Volume 2. Berlin: De Gruyter. 1304–1316.

Ralli, Angela. 2010. Compounding versus derivation. In: Sergio Scalise & Irene Vogel (eds.), *Cross-Disciplinary Issues in Compounding*. Amsterdam/Philadelphia: John Benjamins. 57–73.

Robinson, Justyna A. 2014. Quantifying polysemy in Cognitive Sociolinguistics. In: Dylan Glynn & Justyna A. Robinson (eds.), *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy*. Amsterdam/Philadelphia: John Benjamins. 87–115.

Roeper, Thomas & Muffy E. A. Siegel. 1978. A Lexical Transformation for Verbal Compounds. *Linguistic Inquiry* 9 (2): 199–260.

Rosenberg, Maria. 2007. Agent Nouns, Productivity and Diachrony: An Analysis of $(VN/A)_{N/A}$ Compounds and *-eur* Derivations in French. In: Geert Booij, Luca Ducceschi, Bernard Fradin, Emilano Guevara, Angela Ralli & Sergio Scalise (eds.), *On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5)*, Fréjus, 15–18 September 2005. 359–378.

Russian National Corpus (RNC). See http://ruscorpora.ru.

Ryder, Mary E. 1999. *Bankers* and *blue-chippers*: An account of *-er* formations in present-day English. *English Language and Linguistics* 3: 269–297.

Scalise, Sergio. 1984. *Generative Morphology*. Dordrecht: Foris.

Scalise, Sergio. 1994. *Morfologia*, Bologna: Il Mulino.

Scalise, Sergio & Antonietta Bisetto. 2009. The Classification of Compounds. In: Rochelle Lieber & Pavol Štekauer, *The Oxford Handbook of Compounding*. New York: Oxford University Press. 34–53.

Scherer, Carmen. 2003. Diachronic word formation: "-er" nominals in German. In: *Proceedings of XVII International Congress of Linguists*. Prague, Czech Republic, July 24–29, 2003. Prague: Matfyzpress. 1–15.

Selkirk, Elisabeth O. 1982. *The Syntax of Words*. Cambridge, MA: The MIT Press.

Serrano Dolader, David. 1995. *Las formaciones parasintéticas en español*. Madrid: Arco Libros.

Shank, Christopher, Koen Plevoets & Hubert Cuyckens. 2014. A diachronic corpus-based multivariate analysis of "I think that" vs. "I think zero". In: Dylan Glynn & Justyna A. Robinson (eds.), *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy*. Amsterdam/Philadelphia: John Benjamins. 279–303.

Siegel, Dorothy C. 1974. *Topics in English Morphology*. PhD dissertation. Cambridge, MA: MIT.

Simone, Raffaele. 2000. Cycles lexicaux. *Studi Italiani di Linguistica Teorica e Applicata (SILTA)* 2: 259–287.

Simone, Raffaele. 2006. Classi di costruzioni. In: Nicola Grandi & Gabriele Iannaccaro (eds.), *Zhì. Scritti in onore di Emanuele Banfi in occasione del suo 60° compleanno*. Roma: Caissa Italia. 383–409.

Simone, Raffaele. 2007. Constructions and Categories in Verbal and Signed Languages. In: Elena Pizzuto, Paola Pietrandea & Raffaele Simone (eds.), *Verbal and Signed Languages: Comparing Structures, Constructs and Methodologies*. Berlin: Mouton de Gruyter: 199–250.

Speelman, Dirk. 2014. Logistic regression. A confirmatory technique for comparisons in corpus linguistics. In: Dylan Glynn & Justyna A. Robinson (eds.), *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy*. Amsterdam/Philadelphia: John Benjamins. 487–533.

Steels, Luc. 2013. Fluid Construction Grammar. In: Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press. 153–167.

Strobl, Carolin, James Malley & Gerhard Tutz. 2009. An Introduction to Recursive Partitioning: Rationale, Application and Characteristics of Classification and Regression Trees, Bagging and Random Forests. *Psychological Methods* 14 (4): 323–348.

Szymanek, Bogdan. 2009. IE, Slavonic: Polish. In: Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford Handbook of Compounding*. New York: Oxford University Press. 464–477.

Šanskij, Nikolaj M. 2010 [1968]. *Očerki po russkomu slovoobrazovaniju* [Essays on Russian word-formation]. Moskva: URSS.

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: Fabio Montermini, Gilles Boyé & Jesse Tseng (eds.), *Selected proceedings of the 6th Décembrettes*. Somerville: Cascadilla Proceedings Project. 138–147.

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: Sandra Augendre, Graziella Couasnon-Torlois, Déborah Lebon, Clément Michard, Gilles Boyé & Fabio Montermini (eds.), *Proceedings of the Décembrettes. 8th International Conference on Morphology*. Toulouse: Université de Toulouse. 319–326.

Švedova, Natal'ja Ju. (ed.) 1980. *Russkaja Grammatika* [Russian Grammar]. Moskva: Nauka. [AG-80]

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

Tagabileva, Marija & Julija Berezuckaja. 2010. Slovoobrazovatel'naja razmetka Nacional'nogo Korpusa russkogo jazyka: zadači i metody [Word-formation annotation of the Russian National Corpus: tasks and methods]. In: Andrej Kibrik et al. (eds.), *Computational Linguistics and Intellectual Technologies*, Papers from the Annual International Conference "Dialogue", Issue 9, Volume 16. Moskva: RGGU. 499–506.

Tagliamonte, Sali A. & Harald R. Baayen. 2012. Models, forests and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change* 24 (2): 135–178. See http://www.sfs.uni-tuebingen.de/~hbaayen/publications/TagliamonteBaayen2012.pdf.

Terkulov, Vjačeslav I. 2008a. *Kompozity russkogo jazyka v onomasiologičeskom aspekte*. Dissertacija na soiskanie učenoj stepeni doktora filologičeskich nauk, 10.02.02 Russkij jazyk. [Russian compounds from an onomasiological perspective, Dissertation for the title of doctor of philological sciences]. Gorlovskij gosudarstvennyj pedagogičeskij institut inostrannych jazykov.

Terkulov, Vjačeslav I. 2008b. Bachuvrichi v russkom jazyke [Bahuvrīhi in the Russian language]. *Vestnik Volgogradskogo gosudarstvennogo universiteta, Serija 2, Jazykoznanie* 2 (8): 185–190.

Terkulov, Vjačeslav I. 2013. Tipologija psevdouniverbalizacij v russkom jazyke [Typology of pseudo-univerbalization in the Russian language]. *Vestnik Volgogradskogo gosudarstvennogo universiteta, Serija 2, Jazykoznanie* 3 (19): 55–62.

The Sketch Engine. See https://www.sketchengine.co.uk/.

Tomasello, Michael. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

Townsend, Charles E. 1980. *Russian Word-formation*. Columbus: Slavica Publishers.

Traugott, Elizabeth. 2003. Constructions in Grammaticalization. In: Brian D. Joseph & Richard D. Janda (eds.), *The Handbook of Historical Linguistics*. Oxford/Malden: Blackwell.

Traugott, Elizabeth. 2008. The Grammaticalization of NP *of* NP Patterns. In: Alexander Bergs & Gabriele Diewald (eds.), *Constructions and Language Change*. Berlin: Mouton de Gruyter. 23–45.

Traugott, Elizabeth & Graeme Trousdale. 2010. Gradience, Gradualness and Grammaticalization: How do They Intersect? In: Elizabeth Traugott & Graeme Trousdale (eds.), *Gradience, Gradualness and Grammaticalization*. Amsterdam/Philadelphia: John Benjamins. 19–44.

Traugott, Elizabeth & Graeme Trousdale. 2014. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.

Uluchanov, Igor' S. 2008 [1996]. *Edinicy slovoobrazovatel'noj sistemy russkogo jazyka i ich leksičeskaja realizacija* [Units of the Russian word-formation system and their lexical realization]. Moskva: LKI.

Uluchanov, Igor' S. 2015 [2005]. *Motivacija v slovoobrazovatel'noj sisteme russkogo jazyka* [Motivation in the Russian word-formation system]. Moskva: LIBROKOM.

Ušakov, Dmitrij N. (ed.). 1935–1940. *Tolkovyj slovar' russkogo jazyka* [Explanatory dictionary of the Russian language]. Moskva: Gosudarstvennyj institut "Sovetskaja ènciklopedija". See https://gufo.me/dict/ushakov.

Vaillant, André. 1974. *Grammaire comparée des langues slaves, Tome IV, La formation des noms*. Paris: Klincksieck.

Vasilevskaja, Evgenija A. 1962. *Slovosloženie v russkom jazyke. Očerki i nabljudenija* [Compounding in Russian. Essays and observations]. Moskva: Učpedgiz.

Vasilevskaja, Evgenija A. 1968. *O russkom slovosloženii (slovosočetanie, slovosloženie i affiksacija). Učebnoe posobie* [On compounding in Russian (word combination, compounding, and affixation). A training manual]. Moskva: Moskovskij Poligrafičeskij Institut.

Warren, Beatrice. 1999. Aspects of referential metonymy. In: Klaus-Uwe Panther, Linda L. Thornburg & Antonio Barcelona (eds.), *Metonymy and metaphor in grammar*. Amsterdam/Philadelphia: John Benjamins. 121–138.

Worth, Dean. 1972. Morfonologija nulevoj affiksacii v russkom slovoobrazovanii [Morphonology of zero affixation in Russian word-formation]. *Voprosy Jazykoznanija* 6: 76–84.

Zemskaja, Elena A. 2005 [1992]. *Slovoobrazovanie kak dejatel'nost'* [Word-formation as an activity]. Moskva: Nauka.

Zemskaja, Elena A. 2011 [1973]. *Sovremennyj russkij jazyk. Slovoobrazovanie* [Contemporary Russian. Word-formation]. Moskva: Prosveščenie.

Appendix 1

Analysis of the compounds

		lemma	meaning	semantics	animacy	aspect	transitivity	pos	sr
-ec									
	1	*aviamosec*	aircraft carrier	instr	inan	ipfv	tr	n	th
	2	*basnopisec*	fable writer	ag	hum	ipfv	tr	n	th
	3	*bogoborec*	theomachist	ag	hum	ipfv	intr	n	th
	4	*bogomolec*	devotee	ag	hum	ipfv	intr	n	rec
	5	*bogonosec*	god-bearing	ag	hum	ipfv	tr	n	th
	6	*borzopisec*	careless writer	ag	hum	ipfv	tr	a	mnr
	7	*borzochodec*	fast runner	ag	hum	ipfv	intr	a	mnr
	8	*bronenosec*	armadillo	cos	an	ipfv	tr	n	th
	9	*bronenosec*	battleship	cos	inan	ipfv	tr	n	th
	10	*bykoboec*	bull fighter	ag	hum	ipfv	tr	n	th
	11	*vencenosec*	monarch	cos	hum	ipfv	tr	n	th
	12	*vertoletonosec*	helicopter carrier	instr	inan	ipfv	tr	n	th
	13	*verchovodec*	leader	ag	hum	ipfv	tr	a	mnr
	14	*vinoprodavec*	wine seller	ag	hum	ipfv	tr	n	th
	15	*vinotorgovec*	wine trader	ag	hum	ipfv	tr	n	th

15	*vinotorgovec*	wine trader	ag	hum	ipfv	tr	n	th
16	*vlastoderžec*	monarch	ag	hum	ipfv	tr	n	th
17	*vlastoljubec*	power lover	cos	hum	ipfv	tr	n	th
18	*vodonosec*	water carrier	ag	hum	ipfv	tr	n	th
19	*vol'nodumec*	free thinker	cos	hum	ipfv	tr	a	mnr
20	*vol'noljubec*	freedom lover	cos	hum	ipfv	tr	a	th
21	*vsevidec*	all-seeing	cos	hum	ipfv	tr	pro	th
22	*gorododeržec*	city ruler	ag	hum	ipfv	tr	n	th
23	*gradoderžec*	city ruler	ag	hum	ipfv	tr	n	th
24	*gromoveržec*	thunder thrower	ag	hum	ipfv	tr	n	th
25	*dvoeborec*	biathlon competitor	ag	hum	ipfv	intr	num	mnr
26	*desjatiborec*	decathlon competitor	ag	hum	ipfv	intr	num	mnr
27	*drevotočec*	cossid miller	ag	an	ipfv	tr	n	th
28	*duchoborec*	spirit warrior	ag	hum	ipfv	intr	n	pur
29	*duchovidec*	clairvoyant	cos	hum	ipfv	tr	n	loc
30	*dušegubec*	murderer, torturer	ag	hum	ipfv	tr	n	th
31	*dušelovec*	soul catcher	ag	hum	ipfv	tr	n	th
32	*edinoborec*	one-to-one fighter	ag	hum	ipfv	intr	a	mnr
33	*edinoderžec*	monarch	ag	hum	ipfv	tr	a	mnr
34	*žgutikonosec*	flagellate	cos	an	ipfv	tr	n	th
35	*žezlonosec*	crosier bearer	cos	hum	ipfv	tr	n	th

36	*ženoljubec*	ladies' man	cos	hum	ipfv	tr	n	th
37	*živopisec*	painter	ag	hum	ipfv	tr	a	th
38	*žizneljubec*	life lover	cos	hum	ipfv	tr	n	th
39	*zaimodavec*	creditor	ag	hum	ipfv	tr	n	th
40	*zajmodavec*	creditor	ag	hum	ipfv	tr	n	th
41	*zvezdonosec*	military decoration bearer	cos	hum	ipfv	tr	n	th
42	*zemlepašec*	farmer	ag	hum	ipfv	tr	n	th
43	*zemleprochodec*	explorer	ag	hum	ipfv	tr	n	path
44	*zmeeborec*	snake defeater	ag	hum	ipfv	intr	n	th
45	*zmeenosec*	serpent-bearer (constellation)	cos	hum	ipfv	tr	n	th
46	*znamenosec*	flag-bearer	cos	hum	ipfv	tr	n	th
47	*ikonoborec*	iconoclast	ag	hum	ipfv	intr	n	th
48	*ikonopisec*	icon painter	ag	hum	ipfv	tr	n	th
49	*kanatochodec*	rope-walker	ag	hum	ipfv	intr	n	path
50	*kitoboec*	whaler	instr	inan	ipfv	tr	n	th
51	*knigoderžec*	book-bearer	ag	hum	ipfv	tr	n	th
52	*knigoprodavec*	book seller	ag	hum	ipfv	tr	n	th
53	*knigotorgovec*	book trader	ag	hum	ipfv	tr	n	th
54	*kon'kobežec*	skater	ag	hum	ipfv	intr	n	instr
55	*kop'enosec*	spare-bearer	cos	hum	ipfv	tr	n	th
56	*korystoljubec*	profit-seeker	cos	hum	ipfv	tr	n	th

57	*krestonosec*	Crusader	cos	hum	ipfv	tr	n	th
58	*krovopivec*	oppressor	ag	hum	ipfv	tr	n	th
59	*krovopiec*	oppressor	ag	hum	ipfv	tr	n	th
60	*krjučkotvorec*	pettifogger	ag	hum	ipfv	tr	n	th
61	*lesotorgovec*	wood trader	ag	hum	ipfv	tr	n	th
62	*letopisec*	chronicler	ag	hum	ipfv	tr	n	th
63	*lichoimec*	bribe-taker	ag	hum	ipfv	tr	n	th
64	*lozochodec*	dowser	ag	hum	ipfv	intr	n	instr
65	*lyžebežec*	skier	ag	hum	ipfv	intr	n	instr
66	*medalenosec*	medal-bearer	cos	hum	ipfv	tr	n	th
67	*mečenosec*	sword-bearer	cos	hum	ipfv	tr	n	th
68	*mečenosec*	swordtail	cos	an	ipfv	tr	n	th
69	*mzdoimec*	bribe-taker	ag	hum	ipfv	tr	n	th
70	*minonosec*	torpedo boat	instr	inan	ipfv	tr	n	th
71	*miroderžec*	god	ag	hum	ipfv	tr	n	th
72	*miroljubec*	peace lover	cos	hum	ipfv	tr	n	th
73	*mirotvorec*	peace maker	ag	hum	ipfv	tr	n	th
74	*mnogoborec*	multisport athlete	ag	hum	ipfv	intr	num	mnr
75	*molotoboec*	hammerer	ag	hum	ipfv	tr	n	instr
76	*morechodec*	seaman	ag	hum	ipfv	intr	n	path
77	*muželožec*	homosexual	ag	hum	ipfv	intr	n	com

78	*narkotorgovec*	drug dealer	ag	hum	ipfv	tr	n	th
79	*narodoljubec*	people lover	cos	hum	ipfv	tr	n	th
80	*odnodumec*	person always thinking about the same	cos	hum	ipfv	tr	a	th
81	*ordenonosec*	order-bearer	cos	hum	ipfv	tr	n	th
82	*oruženosec*	sword-bearer	cos	hum	ipfv	tr	n	th
83	*očevidec*	eye witness	cos	hum	ipfv	tr	n	instr
84	*pervoprochodec*	earliest explorer	ag	hum	ipfv	tr	a	time
85	*pesnopevec*	writer/singer of religious songs	ag	hum	ipfv	tr	n	th
86	*pesnotvorec*	song writer	ag	hum	ipfv	tr	n	th
87	*pis'monosec*	postman	ag	hum	ipfv	tr	n	th
88	*polkovodec*	commander	ag	hum	ipfv	tr	n	th
89	*porfironosec*	monarch	cos	hum	ipfv	tr	n	th
90	*pravdoborec*	freedom fighter	ag	hum	ipfv	intr	n	pur
91	*pravdoljubec*	truth lover	cos	hum	ipfv	tr	n	th
92	*psalmopevec*	psalmist	ag	hum	ipfv	tr	n	th
93	*pjatiborec*	pentathlete	ag	hum	ipfv	intr	num	mnr
94	*rabotorgovec*	slave trader	ag	hum	ipfv	tr	n	th
95	*raketonosec*	rocket carrier	instr	inan	ipfv	tr	n	th
96	*ratoborec*	battle fighter	ag	hum	ipfv	intr	n	loc
97	*rogonosec*	cuckold	cos	hum	ipfv	tr	n	th

98	*rudoznatec*	expert on minerals	cos	hum	ipfv	tr	n	th
99	*rybotorgovec*	fish trader	ag	hum	ipfv	tr	n	th
100	*samovidec*	eye witness	cos	hum	ipfv	tr	pro	mnr
101	*samoderžec*	autocrat	ag	hum	ipfv	tr	pro	mnr
102	*samoljubec*	self-lover	cos	hum	ipfv	tr	pro	th
103	*samopisec*	self-recorder	instr	inan	ipfv	tr	pro	mnr
104	*samoupravec*	tyrant	ag	hum	ipfv	tr	pro	mnr
105	*svetoderžec*	lord of the world	ag	hum	ipfv	tr	n	th
106	*sveščenosec*	sacristan	cos	hum	ipfv	tr	n	th
107	*svobodoljubec*	freedom-lover	cos	hum	ipfv	tr	n	th
108	*sebjaljubec*	self-lover	cos	hum	ipfv	tr	pro	th
109	*serdcevedec*	god	cos	hum	ipfv	tr	n	th
110	*serdcevidec*	god	cos	hum	ipfv	tr	n	th
111	*skoropisec*	shorthand typist	ag	hum	ipfv	tr	a	mnr
112	*skotoložec*	zoophile	ag	hum	ipfv	tr	n	com
113	*skototorgovec*	cattle trader	ag	hum	ipfv	tr	n	th
114	*slavoljubec*	glory lover	cos	hum	ipfv	tr	n	th
115	*sladkopevec*	melodist	ag	hum	ipfv	tr	a	mnr
116	*slastoljubec*	pleasure lover	cos	hum	ipfv	tr	n	th
117	*slovotvorec*	word creator	ag	hum	ipfv	tr	n	th
118	*snovidec*	dream seer	cos	hum	ipfv	tr	n	th

119	*srebroljubec*	avid person	cos	hum	ipfv	tr	n	th
120	*stenopisec*	wall painter	ag	hum	ipfv	tr	n	loc
121	*strannoljubec*	foreigners lover	cos	hum	ipfv	tr	n	th
122	*strannopriimec*	shelter owner	ag	hum	ipfv	tr	n	th
123	*strastoterpec*	martyr	ag	hum	ipfv	tr	n	th
124	*suchojadec*	dry food eater	ag	hum	ipfv	tr	a	th
125	*tajnovidec*	clairvoyant	cos	hum	ipfv	tr	n	th
126	*tiranoborec*	tyranny fighter	ag	hum	ipfv	intr	n	th
127	*torpedonosec*	torpedo boat/bomber	instr	inan	ipfv	tr	n	th
128	*trudoljubec*	hard worker	cos	hum	ipfv	tr	n	th
129	*tunejadec*	parasite	ag	hum	ipfv	tr	a	mnr
130	*fakelonosec*	Olympic torch carrier	cos	hum	ipfv	tr	n	th
131	*flotovodec*	naval commander	ag	hum	ipfv	tr	n	th
132	*chlebopašec*	ploughman	ag	hum	ipfv	tr	n	th
133	*chlebotorgovec*	corn merchant	ag	hum	ipfv	tr	n	th
134	*chorugvenosec*	gonfalon bearer	cos	hum	ipfv	tr	n	th
135	*christoprodavec*	traitor	ag	hum	ipfv	tr	n	th
136	*čaetorgovec*	tea dealer	ag	hum	ipfv	tr	n	th
137	*čelovekoljubec*	people lover	cos	hum	ipfv	tr	n	th
138	*čestoljubec*	ambitious man	cos	hum	ipfv	tr	n	th
139	*čudotvorec*	saint who makes miracles	ag	hum	ipfv	tr	n	th

140	*čudotvorec*	miracle man	ag	hum	ipfv	tr	n	th
141	*ščitonosec*	shield bearer	cos	hum	ipfv	tr	n	th
142	*jazykotvorec*	word creator	ag	hum	ipfv	tr	n	th
143	*jasnovidec*	clairvoyant	cos	hum	ipfv	tr	a	mnr
-lec								
144	*gruzovladelec*	owner of goods	cos	hum	ipfv	intr	n	th
145	*domovladelec*	landlord	cos	hum	ipfv	intr	n	th
146	*zavodovladelec*	factory owner	cos	hum	ipfv	intr	n	th
147	*zemlevladelec*	land owner	cos	hum	ipfv	intr	n	th
148	*rabovladelec*	slave owner	cos	hum	ipfv	intr	n	th
149	*sudovladelec*	ship owner	cos	hum	ipfv	intr	n	th
150	*tovarovladelec*	owner of goods	cos	hum	ipfv	intr	n	th
-tel'								
151	*aviastroitel'*	aircraft builder	ag	hum	ipfv	tr	n	th
152	*avtoljubitel'*	car enthusiast	cos	hum	ipfv	tr	n	th
153	*avtomobilestroitel'*	car builder	ag	hum	ipfv	tr	n	th
154	*avtoproizvoditel'*	car producer	ag	hum	ipfv	tr	n	th
155	*avtostroitel'*	car builder	ag	hum	ipfv	tr	n	th
156	*arendodatel'*	landlord	ag	hum	pfv	tr	n	mnr

157	*balansoderžatel'*	accountant	ag	hum	ipfv	tr	n	th
158	*bacillonositel'*	bacilli carrier	cos	hum	ipfv	tr	n	th
159	*betonosmesitel'*	concrete mixer	instr	inan	pfv	tr	n	th
160	*blagodetel'*	benefactor	ag	hum	ipfv	tr	n	th
161	*blagoželatel'*	benevolent person	cos	hum	ipfv	tr	n	th
162	*blagotvoritel'*	benefactor	ag	hum	ipfv	tr	n	th
163	*blagoustroitel'*	benefactor	ag	hum	pfv	tr	n	th
164	*bogoiskatel'*	god seeker	ag	hum	ipfv	tr	n	th
165	*bogostroitel'*	god builder	ag	hum	ipfv	tr	n	th
166	*bomboderžatel'*	bomb rack	instr	inan	ipfv	tr	n	th
167	*bombometatel'*	bomber	ag	hum	ipfv	tr	n	th
168	*bumagoderžatel'*	paper holder	instr	inan	ipfv	tr	n	th
169	*bumagomaratel'*	scribbler	ag	hum	ipfv	tr	n	th
170	*bytopisatel'*	life writer	ag	hum	ipfv	tr	n	top
171	*vagonooprokidyvatel'*	wagon tippler	instr	inan	ipfv	tr	n	th
172	*vagonostroitel'*	carriage builder	ag	hum	ipfv	tr	n	th
173	*vekseledatel'*	bill of exchange drawer	ag	hum	pfv	tr	n	th
174	*vekselederžatel'*	bill of exchange owner	ag	hum	ipfv	tr	n	th
175	*veroučitel'*	dogma interpreter	ag	hum	ipfv	tr	n	th
176	*vertoletostroitel'*	helicopter builder	ag	hum	ipfv	tr	n	th
177	*vzjatkodatel'*	briber	ag	hum	pfv	tr	n	th

178	*vzjatkopolučatel'*	bribe taker	ag	hum	ipfv	tr	n	th
179	*vidoiskatel'*	view finder	instr	inan	ipfv	tr	n	th
180	*virusonositel'*	virus carrier	cos	hum	ipfv	tr	n	th
181	*vodonagrevatel'*	water heater	instr	inan	ipfv	tr	n	th
182	*vodootdelitel'*	water separator	instr	inan	pfv	tr	n	th
183	*vodopol'zovatel'*	water user	ag	hum	ipfv	intr	n	th
184	*vodopotrebitel'*	water consumer	ag	hum	pfv	tr	n	th
185	*vozduchonagrevatel'*	air heater	instr	inan	ipfv	tr	n	th
186	*vozduchoočistitel'*	air purifier	instr	inan	pfv	tr	n	th
187	*vozduchoplavatel'*	aeronaut	ag	hum	ipfv	intr	n	path
188	*vozduchopodogrevatel'*	air heater	instr	inan	ipfv	tr	n	th
189	*vol'noslušatel'*	auditor	cos	hum	ipfv	tr	a	mnr
190	*vsederžitel'*	the almighty	ag	hum	ipfv	tr	pro	th
191	*vygodopriobretatel'*	beneficiary	ag	hum	ipfv	tr	n	th
192	*gidroraspredelitel'*	hidraulic control valve	instr	inan	pfv	tr	n	th
193	*glubokorychlitel'*	subsurface plow	instr	inan	ipfv	tr	a	mnr
194	*gornospasatel'*	mine rescuer	ag	hum	ipfv	tr	a	loc
195	*gorovoschoditel'*	alpinist	ag	hum	ipfv	intr	n	dir
196	*gradopravitel'*	city ruler	ag	hum	ipfv	intr	n	th
197	*gradostroitel'*	city builder	ag	hum	ipfv	tr	n	th
198	*grobokopatel'*	grave digger	ag	hum	ipfv	tr	n	th

199	*gromkogovoritel'*	loudspeaker	instr	inan	ipfv	tr	a	mnr
200	*gruzootpravitel'*	consignor of goods	ag	hum	pfv	tr	n	th
201	*gruzopolučatel'*	consignee	ag	hum	ipfv	tr	n	th
202	*deloproizvoditel'*	secretary	ag	hum	ipfv	tr	n	th
203	*dobrodetel'*	benefactor	ag	hum	ipfv	tr	n	th
204	*dobroželatel'*	benevolent person	cos	hum	ipfv	tr	n	th
205	*dolgožitel'*	long liver	cos	hum	ipfv	intr	a	time
206	*domopravitel'*	house ruler	ag	hum	ipfv	intr	n	th
207	*domostroitel'*	house builder	ag	hum	ipfv	tr	n	th
208	*estestvoispytatel'*	naturalist	cos	hum	pfv	tr	n	th
209	*zajmoderžatel'*	loan holder	ag	hum	ipfv	tr	n	th
210	*zakonodatel'*	legislator	ag	hum	pfv	tr	n	th
211	*zakonoučitel'*	teacher of god's law	ag	hum	ipfv	tr	n	th
212	*zalogodatel'*	depositor	ag	hum	pfv	tr	n	th
213	*zalogoderžatel'*	pawnbroker	ag	hum	ipfv	tr	n	th
214	*zvukonositel'*	music tape	cos	inan	ipfv	tr	n	th
215	*zvukooformitel'*	sound operator	ag	hum	pfv	tr	n	th
216	*zvukopodražatel'*	sound imitator	ag	hum	ipfv	tr	n	th
217	*zvukosnimatel'*	pickup	instr	inan	ipfv	tr	n	th
218	*zvukoulovitel'*	acoustic mirror	instr	inan	pfv	tr	n	th
219	*zvukousilitel'*	loud-hailer	instr	inan	pfv	tr	n	th

220	*zemlepol'zovatel'*	land tenant	ag	hum	ipfv	intr	n	th
221	*zemleustroitel'*	land organizer	ag	hum	pfv	tr	n	th
222	*zloželatel'*	malevolent person	cos	hum	ipfv	tr	n	th
223	*zlopychatel'*	malevolent person	cos	hum	ipfv	intr	n	th
224	*zolotoiskatel'*	gold digger	ag	hum	ipfv	tr	n	th
225	*igloderžatel'*	needle holder	instr	inan	ipfv	tr	n	th
226	*iskrogasitel'*	spark arrester	instr	inan	ipfv	tr	n	th
227	*kanavokopatel'*	trench digger	instr	inan	ipfv	tr	n	th
228	*kvartironanimatel'*	tenant	ag	hum	ipfv	tr	n	th
229	*kinoljubitel'*	amateur filmmaker	cos	hum	ipfv	tr	n	th
230	*kladoiskatel'*	treasure seeker	ag	hum	ipfv	tr	n	th
231	*knigoizdatel'*	publisher	ag	hum	pfv	tr	n	th
232	*knigorasprostranitel'*	book spreader	ag	hum	pfv	tr	n	th
233	*kožzamenitel'*	leatherette	instr	inan	pfv	tr	n	th
234	*kop'emetatel'*	javelin thrower	ag	hum	ipfv	tr	n	th
235	*korablestroitel'*	ship builder	ag	hum	ipfv	tr	n	th
236	*krovezamenitel'*	blood substitute	instr	inan	pfv	tr	n	th
237	*krovosmesitel'*	incestuous person	ag	hum	pfv	tr	n	th
238	*lesozagotovitel'*	timber cutting worker	ag	hum	pfv	tr	n	th
239	*lesopol'zovatel'*	wood user	ag	hum	ipfv	intr	n	th
240	*lozoiskatel'*	dowser	ag	hum	ipfv	tr	n	instr

241	*maslootražatel'*	oil deflector	instr	inan	ipfv	tr	n	th
242	*mašinostroitel'*	machine builder	ag	hum	ipfv	tr	n	th
243	*mestobljustitel'*	locum tenens	ag	hum	ipfv	tr	n	th
244	*metalloiskatel'*	metal detector	instr	inan	ipfv	tr	n	th
245	*metrostroitel'*	underground builder	ag	hum	ipfv	tr	n	th
246	*minoiskatel'*	mine detector	instr	inan	ipfv	tr	n	th
247	*moreplavatel'*	sailor	ag	hum	ipfv	intr	n	path
248	*mostostroitel'*	bridge builder	ag	hum	ipfv	tr	n	th
249	*motorostroitel'*	motor builder	ag	hum	ipfv	tr	n	th
250	*musorosžigatel'*	incinerator	instr	inan	ipfv	tr	n	th
251	*najmodatel'*	renter	ag	hum	pfv	tr	n	th
252	*nasledodatel'*	testator	ag	hum	pfv	tr	n	th
253	*nebožitel'*	celestial being	cos	hum	ipfv	intr	n	loc
254	*nebokoptitel'*	loafer	ag	hum	ipfv	tr	n	th
255	*nedobroželatel'*	malevolent person	cos	hum	ipfv	tr	n	th
256	*nedropol'zovatel'*	subsurface user	ag	hum	ipfv	intr	n	th
257	*nomeronabiratel'*	dial	instr	inan	ipfv	tr	n	th
258	*ognetušitel'*	fire extinguisher	instr	inan	ipfv	tr	n	th
259	*otkazopolučatel'*	legatee	ag	hum	ipfv	tr	n	th
260	*očkovtiratel'*	impostor	ag	hum	ipfv	tr	n	th
261	*paroochladitel'*	steam cooler	instr	inan	pfv	tr	n	th

262	*paroočistitel'*	fallow fields purifier	instr	inan	pfv	tr	n	th
263	*paroperegrevatel'*	superheater	instr	inan	ipfv	tr	n	th
264	*patentoobladatel'*	patent holder	cos	hum	ipfv	tr	n	th
265	*penkosnimatel'*	cream-skimmer	ag	hum	ipfv	tr	n	th
266	*penoobrazovatel'*	foaming agent	instr	inan	dvuvid	tr	n	th
267	*pervovoschoditel'*	first climber	ag	hum	ipfv	intr	a	mnr
268	*pervootkryvatel'*	pioneer	ag	hum	ipfv	tr	a	mnr
269	*plamegasitel'*	flame arrester	instr	inan	ipfv	tr	n	th
270	*poklažedatel'*	depositor	ag	hum	pfv	tr	n	th
271	*polotencesušitel'*	towel drier	instr	inan	ipfv	tr	n	th
272	*počvoobrazovatel'*	soil former	instr	inan	dvuvid	tr	n	th
273	*pravdoiskatel'*	truth searcher	ag	hum	ipfv	tr	n	th
274	*pravonarušitel'*	transgressor	ag	hum	pfv	tr	n	th
275	*pravoobladatel'*	right holder	cos	hum	ipfv	tr	n	th
276	*priborostroitel'*	instrument builder	ag	hum	ipfv	tr	n	th
277	*prirodopol'zovatel'*	environment user	ag	hum	ipfv	intr	n	th
278	*pristanoderžatel'*	resetter	ag	hum	ipfv	tr	n	th
279	*pustynnožitel'*	hermit	cos	hum	ipfv	intr	a	mnr
280	*putevoditel'*	guide	instr	inan	ipfv	tr	n	th
281	*pyleulovitel'*	dust collector	instr	inan	pfv	tr	n	th
282	*pjatnovyvoditel'*	spot remover	instr	inan	ipfv	tr	n	th

283	*rabotodatel'*	employer	ag	hum	pfv	tr	n	th
284	*radioljubitel'*	radio amateur	cos	hum	ipfv	tr	n	th
285	*radioslušatel'*	radio listener	cos	hum	ipfv	tr	n	th
286	*raketonositel'*	missile carrier	instr	inan	ipfv	tr	n	instr
287	*raketostroitel'*	missile builder	ag	hum	ipfv	tr	n	th
288	*reestroderžatel'*	register holder	ag	hum	ipfv	tr	n	th
289	*reklamodatel'*	advertiser	ag	hum	pfv	tr	n	th
290	*reklamoproizvoditel'*	advertising producer	ag	hum	ipfv	tr	n	th
291	*reklamorasprostranitel'*	advertising spreader	ag	hum	pfv	tr	n	th
292	*samoletostroitel'*	aircraft builder	ag	hum	ipfv	tr	n	th
293	*samoučitel'*	manual for self-tuition	instr	inan	ipfv	tr	pro	mnr
294	*sacharozamenitel'*	sugar substitute	instr	inan	pfv	tr	n	th
295	*svetovozvraščatel'*	retro-reflector	instr	inan	ipfv	tr	n	th
296	*skorosšivatel'*	loose-leaf binder	instr	inan	ipfv	tr	a	mnr
297	*snegoočistitel'*	snow plough	instr	inan	pfv	tr	n	th
298	*ssudodatel'*	lender	ag	hum	pfv	tr	n	th
299	*ssudopolučatel'*	bailee	ag	hum	ipfv	tr	n	th
300	*stankostroitel'*	machine-tool builder	ag	hum	ipfv	tr	n	th
301	*stekloobrazovatel'*	glass forming element	instr	inan	dvuvid	tr	n	th
302	*stekloočistitel'*	screen wiper	instr	inan	pfv	tr	n	th
303	*stichoslagatel'*	verse writer	ag	hum	ipfv	tr	n	th

304	*stogometatel'*	hay stacker	instr	inan	ipfv	tr	n	th
305	*sudovoditel'*	navigator	ag	hum	ipfv	tr	n	th
306	*sudostroitel'*	ship builder	ag	hum	ipfv	tr	n	th
307	*teplonositel'*	heat carrier	instr	inan	ipfv	tr	n	th
308	*termopreobrazovatel'*	heat transformer	instr	inan	dvuvid	tr	n	th
309	*tovaroproizvoditel'*	commodity producer	ag	hum	ipfv	tr	n	th
310	*traktorostroitel'*	tractor builder	ag	hum	ipfv	tr	n	th
311	*tranšeekopatel'*	trencher	instr	inan	ipfv	tr	n	th
312	*turbostroitel'*	turbine builder	ag	hum	ipfv	tr	n	th
313	*urokodatel'*	lesson-giver	ag	hum	pfv	tr	n	th
314	*fondoobrazovatel'*	fund former	ag	hum	dvuvid	tr	n	th
315	*fotoljubitel'*	photography amateur	cos	hum	ipfv	tr	n	th
316	*chlebozagotovitel'*	state grain purchaser	ag	hum	pfv	tr	n	th
317	*chramozdatel'*	temple builder	ag	hum	pfv	tr	n	th
318	*čekodatel'*	cheque drawer	ag	hum	pfv	tr	n	th
319	*čekoderžatel'*	drawn cheque owner	ag	hum	ipfv	tr	n	th
320	*členovreditel'*	guilty of mutilation	ag	hum	ipfv	tr	n	th
321	*špagoglotatel'*	sword swallower	ag	hum	ipfv	tr	n	th
322	*ščitoderžatel'*	shield holder	cos	inan	ipfv	tr	n	th
323	*ènergonositel'*	energy carrier	instr	inan	ipfv	tr	n	th
324	*ènergostroitel'*	energy builder	ag	hum	ipfv	tr	n	th

325	*jamokopatel'*	hole digger	instr	inan	ipfv	tr	n	th

-nik

326	*bogochul'nik*	blasphemer	ag	hum	ipfv	tr	n	th
327	*bortprovodnik*	air steward	ag	hum	ipfv	tr	n	dir
328	*verootstupnik*	faith abandonner	ag	hum	us	intr	n	th
329	*vozduchozabornik*	air intake	instr	inan	us	tr	n	th
330	*grechovodnik*	sinner	ag	hum	ipfv	tr	n	th
331	*gruzopod"emnik*	goods elevator	instr	inan	us	tr	n	th
332	*derevoobdeločnik*	wood worker	ag	hum	us	tr	n	th
333	*ženonenavistnik*	woman hater	cos	hum	ipfv	tr	n	th
334	*zakonoprestupnik*	criminal	ag	hum	us	tr	n	th
335	*idolopoklonnik*	idolater	ag	hum	us	intr	n	dir
336	*kalopriemnik*	colostomy bag	instr	inan	us	tr	n	th
337	*kljatvoprestupnik*	perjurer	ag	hum	pfv	tr	n	th
338	*močetočnik*	ureter	instr	inan	ipfv	intr	n	th
339	*musorosbornik*	garbage bin	instr	inan	us	tr	n	th
340	*nizkopoklonnik*	toady	ag	hum	us	intr	a	dir
341	*ognepoklonnik*	fire worshipper	ag	hum	us	intr	n	dir
342	*pravozaščitnik*	human rights activist	ag	hum	pfv	tr	n	th
343	*pravopreemnik*	assignee	ag	hum	us	tr	n	th

344	*prestolonaslednik*	successor to the throne	ag	hum	dvuvid	tr	n	th
345	*rukomojnik*	washstand	instr	inan	ipfv	tr	n	th
346	*solncepoklonnik*	sun worshipper	ag	hum	us	intr	n	dir
347	*steklopod"emnik*	window raiser	instr	inan	us	tr	n	th
348	*sudopod"emnik*	ship elevator	instr	inan	us	tr	n	th
349	*teploobmennik*	heat exchanger	instr	inan	us	intr	n	th
350	*tokopriemnik*	current collector	instr	inan	us	tr	n	th
351	*tokos"emnik*	current collector	instr	inan	us	tr	n	th
352	*fotopriemnik*	photodetector	instr	inan	us	tr	n	th
353	*čelovekonenavistnik*	misanthrope	cos	hum	ipfv	tr	n	th

-ščik/čik

354	*astroletčik*	astronaut	ag	hum	us	intr	n	path
355	*asfal'toukladčik*	asphalt paver	instr	inan	ipfv	tr	n	th
356	*benzozapravščik*	refueller	instr	inan	us	tr	n	instr
357	*gazosvarščik*	gas welder	ag	hum	us	tr	n	instr
358	*granatometčik*	grenadier	ag	hum	us	tr	n	th
359	*gruzoperevozčik*	cargo company	ag	inan	ipfv	tr	n	th
360	*zvezdoletčik*	astronaut	ag	hum	us	intr	n	path
361	*kabeleukladčik*	cable handler	instr	inan	ipfv	tr	n	th
362	*kvartiros"emščik*	tenant	ag	hum	us	tr	n	th

363	*kondensatootvodčik*	steam trap	instr	inan	ipfv	tr	n	th
364	*maslobojščik*	oil miller	ag	hum	ipfv	tr	n	th
365	*mostoukladčik*	bridge-layer	instr	inan	ipfv	tr	n	th
366	*puskonaladčik*	motion adjuster	ag	hum	pfv	tr	n	th
367	*puteobchodčik*	track man	ag	hum	ipfv	tr	n	th
368	*puteukladčik*	track layer	ag	hum	ipfv	tr	n	th
369	*slovolitčik*	type founder	ag	hum	ipfv	tr	n	th
370	*stalelitejščik*	steel founder	ag	hum	ipfv	tr	n	th
371	*sudosborščik*	shipwright	ag	hum	us	tr	n	th
372	*toplivozapravščik*	refueller	instr	inan	us	tr	n	th
373	*troubokladčik*	pipelaying crane	instr	inan	ipfv	tr	n	th
374	*troubokladčik*	pipe layer	ag	hum	ipfv	tr	n	th
375	*čelobitčik*	petitioner	ag	hum	ipfv	tr	n	instr
376	*čugunolitejščik*	iron founder	ag	hum	ipfv	tr	n	th

-l'ščik

377	*l'notrebil'ščik*	flax puller	ag	hum	ipfv	tr	n	th
378	*nalogoplatel'ščik*	tax payer	ag	hum	ipfv	tr	n	th
379	*staleplavil'ščik*	steel maker	ag	hum	ipfv	tr	n	th
380	*šelkomotal'ščik*	cocoon unwinder	ag	hum	ipfv	tr	n	th

-ka

381	*belošvejka*	seamstress	ag	hum	ipfv	tr	n	th
382	*bumagomaraka*	paper-stainer	ag	hum	ipfv	tr	n	th
383	*vodokačka*	water-tower	instr	inan	ipfv	tr	n	th
384	*vodomerka*	pond skater	ag	an	ipfv	tr	n	th
385	*vozduchoduvka*	compressed air releaser	instr	inan	ipfv	tr	n	th
386	*vseznajka*	know-all	cos	hum	ipfv	tr	pro	th
387	*dušegrejka*	shrug	instr	inan	ipfv	tr	n	th
388	*dušegubka*	mobile gas chamber	instr	inan	ipfv	tr	n	th
389	*zemlerojka*	shrew	ag	an	ipfv	tr	n	th
390	*zolotošvejka*	gold threads master	ag	hum	ipfv	tr	n	instr
391	*zubočistka*	toothpick	instr	inan	ipfv	tr	n	th
392	*kožemjaka*	leather manufacturer	ag	hum	ipfv	tr	n	th
393	*kofevarka*	coffee-machine	instr	inan	ipfv	tr	n	th
394	*kofemolka*	coffee-grinder	instr	inan	ipfv	tr	n	th
395	*kraskoterka*	paint grinder	instr	inan	ipfv	tr	n	th
396	*kruporuška*	cereal transformer	instr	inan	ipfv	tr	n	th
397	*krysolovka*	rat trap	instr	inan	ipfv	tr	n	th
398	*krysolovka*	rat catcher	ag	an	ipfv	tr	n	th
399	*maloežka*	person who does not eat much	ag	hum	ipfv	tr	a	th
400	*maslobojka*	butter churn	instr	inan	ipfv	tr	n	th

401	*mnogoznajka*	know-all	cos	hum	ipfv	tr	num	th
402	*mucholovka*	fly trap	instr	inan	ipfv	tr	n	th
403	*mucholovka*	house centipede	ag	an	ipfv	tr	n	th
404	*mucholovka*	Venus flytrap (plant)	ag	inan	ipfv	tr	n	th
405	*myšelovka*	mouse trap	instr	inan	ipfv	tr	n	th
406	*mjasorubka*	meat grinder	instr	inan	ipfv	tr	n	th
407	*nosogrejka*	tobacco pipe	instr	inan	ipfv	tr	n	th
408	*ovoščečistka*	vegetable peeler	instr	inan	ipfv	tr	n	th
409	*polomojka*	scrubwoman	ag	hum	ipfv	tr	n	th
410	*poluznajka*	half-educated person	cos	hum	ipfv	tr	n	mnr
411	*posudomojka*	kitchen porter	ag	hum	ipfv	tr	n	th
412	*posudomojka*	dishwasher	instr	inan	ipfv	tr	n	th
413	*prosoruška*	millet-scourer	instr	inan	ipfv	tr	n	th
414	*pustolajka*	dog barking for no reason	ag	an	ipfv	tr	a	mnr
415	*samosbroska*	sliced bread dropper	instr	inan	pfv	tr	pro	mnr
416	*samosejka*	self-sown plant	ag	inan	ipfv	intr	pro	mnr
417	*samoskidka*	sliced bread dropper	instr	inan	us	tr	pro	mnr
418	*samoučka*	autodidact	ag	hum	ipfv	tr	pro	mnr
419	*skorovarka*	pressure cooker	instr	inan	ipfv	tr	a	mnr
420	*sladkoežka*	sweet tooth	ag	hum	ipfv	tr	a	th
421	*slastoežka*	sweet tooth	ag	hum	ipfv	tr	n	th

422	*solomorezka*	straw cutter	instr	inan	ipfv	tr	n	th
423	*sudomojka*	dishwasher	instr	inan	ipfv	tr	n	th
424	*syroežka*	russula (mushroom)	ag	inan	ipfv	tr	n	th
425	*telogrejka*	padded jacket	instr	inan	ipfv	tr	n	th
426	*uchovertka*	earwig	ag	an	ipfv	tr	n	th
427	*chleborezka*	bread slicer	instr	inan	ipfv	tr	n	th

-lka

428	*betonomešalka*	concrete mixer	instr	inan	ipfv	tr	n	th
429	*gazonokosilka*	lawnmower	instr	inan	ipfv	tr	n	th
430	*zemlečerpalka*	dredger	instr	inan	ipfv	tr	n	th
431	*zernosušilka*	grain dryer	instr	inan	ipfv	tr	n	th
432	*kamnedrobilka*	stone breaker	instr	inan	ipfv	tr	n	th
433	*kartofelesažalka*	potato planter	instr	inan	ipfv	tr	n	th
434	*kormodrobilka*	forage cutter	instr	inan	ipfv	tr	n	th
435	*senokosilka*	mowing machine	instr	inan	ipfv	tr	n	th
436	*snegotajalka*	snow melter	instr	inan	ipfv	tr	n	th
437	*snopovjazalka*	grain binder	instr	inan	ipfv	tr	n	th
438	*sokovyžimalka*	squeezer	instr	inan	ipfv	tr	n	th

-ø

439	*avtovoz*	car transporter	instr	inan	ipfv	tr	n	th
440	*atomochod*	atomic-powered vessel	instr	inan	ipfv	intr	a	instr
441	*bachčevod*	melon grower	ag	hum	ipfv	tr	n	th
442	*benzovoz*	tank truck	instr	inan	ipfv	tr	n	th
443	*bljudoliz*	lickspittle	ag	hum	ipfv	tr	n	th
444	*bogomaz*	icon painter	ag	hum	ipfv	tr	n	th
445	*bogomol*	mantis	ag	an	ipfv	intr	n	rec
446	*bolotochod*	swamp mobile	instr	inan	ipfv	intr	n	path
447	*bombovoz*	bomber	instr	inan	ipfv	tr	n	th
448	*bradobrej*	barber	ag	hum	ipfv	tr	n	th
449	*brakodel*	bungler	ag	hum	ipfv	tr	n	th
450	*bukvoed*	pedant	ag	hum	ipfv	tr	n	th
451	*vezdechod*	off-road vehicle	instr	inan	ipfv	intr	a	path
452	*verchovod*	leader	ag	hum	ipfv	tr	a	mnr
453	*verchogljad*	superficial person	ag	hum	ipfv	tr	a	dir
454	*vercholaz*	steeplejack	ag	hum	ipfv	intr	a	dir
455	*verchochvat*	superficial person	ag	hum	ipfv	tr	a	mnr
456	*vetroduv*	ventilator	instr	inan	ipfv	tr	n	th
457	*vetroduj*	ventilator	instr	inan	ipfv	tr	n	th
458	*vetromer*	anemometer	instr	inan	ipfv	tr	n	th

459	*vzjatkochvat*	bribe-taker	ag	hum	ipfv	tr	n	th
460	*vinoglot*	alchoolic	ag	hum	ipfv	tr	n	th
461	*vinodel*	wine maker	ag	hum	ipfv	tr	n	th
462	*vinokur*	distiller	ag	hum	ipfv	tr	n	th
463	*vodovoz*	water carrier	ag	hum	ipfv	tr	n	th
464	*vodovoz*	water carrier	instr	inan	ipfv	tr	n	th
465	*vodovypusk*	culvert	instr	inan	us	tr	n	th
466	*vodolaz*	diver	ag	hum	ipfv	intr	n	dir
467	*vodolaz*	Newfoundland dog	ag	an	ipfv	intr	n	dir
468	*vodolej*	Aquarius	ag	inan	ipfv	tr	n	th
469	*vodolej*	worker pouring water	ag	hum	ipfv	tr	n	th
470	*vodoliv*	worker pouring water	ag	hum	ipfv	tr	n	th
471	*vodoljub*	water scavenger beetle	cos	an	ipfv	tr	n	th
472	*vodomer*	water gauge	instr	inan	ipfv	tr	n	th
473	*vodonos*	water carrier	ag	hum	ipfv	tr	n	th
474	*vodosbros*	spillway	instr	inan	pfv	tr	n	th
475	*vodospusk*	floodgate	instr	inan	ipfv	tr	n	th
476	*vodostok*	drain pipe	instr	inan	us	intr	n	th
477	*vozduchoduv*	compressed air releaser	instr	inan	ipfv	tr	n	th
478	*volkodav*	wolfhound	ag	an	ipfv	tr	n	th
479	*volnolom*	breakwater	instr	inan	ipfv	tr	n	th

481	*volosočes*	hairdresser	ag	hum	ipfv	tr	n	th
482	*vol'nodum*	free thinker	cos	hum	ipfv	intr	a	mnr
483	*vostokoved*	orientalist	cos	hum	ipfv	tr	n	th
484	*vysotomer*	altimeter	instr	inan	ipfv	tr	n	th
485	*gazomer*	gas-meter	instr	inan	ipfv	tr	n	th
486	*golovorez*	daredevil	ag	hum	ipfv	tr	n	th
487	*golovotjap*	bungler	ag	hum	ipfv	tr	n	instr
488	*golubevod*	dove breeder	ag	hum	ipfv	tr	n	th
489	*goremyka*	poor wretch	cos	hum	ipfv	tr	n	th
490	*gorloder*	screamer	ag	hum	ipfv	tr	n	th
491	*gorlochvat*	boor, arrogant person	ag	hum	ipfv	tr	n	th
492	*gosudarstvoved*	political scientist	cos	hum	ipfv	tr	n	th
493	*granatomet*	grenade launcher	instr	inan	us	tr	n	th
494	*gromootvod*	lightning rod	instr	inan	ipfv	tr	n	th
495	*gruzinoved*	expert on Georgian	cos	hum	ipfv	tr	n	th
496	*gruntoved*	soil engineer	cos	hum	ipfv	tr	n	th
497	*gruppovod*	group leader	ag	hum	ipfv	tr	n	th
498	*gubošlep*	mumbler	ag	hum	us	intr	n	instr
499	*gurtoprav*	drover	ag	hum	ipfv	intr	n	th
500	*gusevod*	goose breeder	ag	hum	ipfv	tr	n	th
501	*dal'nomer*	rangefinder	instr	inan	ipfv	tr	n	th

501	*dal'nomer*	rangefinder	instr	inan	ipfv	tr	n	th
502	*darmoglot*	sponger	ag	hum	ipfv	tr	a	mnr
503	*darmoed*	sponger	ag	hum	ipfv	tr	a	mnr
504	*dobrochot*	benevolent person	cos	hum	ipfv	tr	n	th
505	*dolgodum*	slow-witted person	cos	hum	ipfv	intr	a	mnr
506	*domovod*	housewife	ag	hum	ipfv	tr	n	th
507	*domosed*	homebody	cos	hum	ipfv	intr	n	loc
508	*domouprav*	house manager	ag	hum	ipfv	intr	n	th
509	*drevogryz*	powderpost beetle	ag	an	ipfv	tr	n	th
510	*drevolaz*	poison dart frog	ag	an	ipfv	intr	n	dir
511	*drovokol*	woodchopper	ag	hum	ipfv	tr	n	th
512	*drovorub*	woodcutter	ag	hum	ipfv	tr	n	th
513	*drovosek*	woodcutter	ag	hum	ipfv	tr	n	th
514	*dubolom*	stupid person	ag	hum	ipfv	tr	n	th
515	*dubotolk*	stupid person	ag	hum	us	tr	n	th
516	*durolom*	stupid person	ag	hum	ipfv	tr	a	mnr
517	*duropljas*	stupid person	ag	hum	ipfv	tr	a	mnr
518	*duchobor*	spirit warrior	ag	hum	ipfv	intr	n	pur
519	*dušegub*	murderer, torturer	ag	hum	ipfv	tr	n	th
520	*dymokur*	bonfire against mosquitoes	instr	inan	ipfv	tr	n	th
521	*dymochod*	flue	instr	inan	ipfv	intr	n	th

522	*estestvoved*	expert on natural science	cos	hum	ipfv	tr	n	th
523	*ženoljub*	ladies' man	cos	hum	ipfv	tr	n	th
524	*živoglot*	exploiter	ag	hum	ipfv	tr	a	th
525	*živoglot*	snaketooth fish	ag	an	ipfv	tr	a	th
526	*živoder*	knacker	ag	hum	ipfv	tr	a	th
527	*životnovod*	cattle-breeder	ag	hum	ipfv	tr	n	th
528	*žizneljub*	life lover	cos	hum	ipfv	tr	n	th
529	*zakonoved*	jurist	cos	hum	ipfv	tr	n	th
530	*zvezdolet*	spaceship	instr	inan	us	intr	n	path
531	*zvezdochvat*	pickpocket	ag	hum	ipfv	tr	n	th
532	*zvezdočet*	astrologer	ag	hum	ipfv	tr	n	th
533	*zvezdočet*	stargazer (fish)	ag	an	ipfv	tr	n	th
534	*zveroboj*	ochotnik	ag	hum	ipfv	tr	n	th
535	*zveroboj*	St. John's wort (plant)	ag	inan	ipfv	tr	n	th
536	*zveroboj*	vodka made from St. John's wort	ag	inan	ipfv	tr	n	th
537	*zverovod*	fur farm operator	ag	hum	ipfv	tr	n	th
538	*zverolov*	hunter	ag	hum	ipfv	tr	n	th
539	*zemleved*	expert on physical geography	cos	hum	ipfv	tr	n	th
540	*zemlevoz*	worker transporting earth	ag	hum	ipfv	tr	n	th
541	*zemlekop*	digger	ag	hum	ipfv	tr	n	th
542	*zemlemer*	land surveyor	ag	hum	ipfv	tr	n	th

543	*zemlerob*	farmer	ag	hum	ipfv	tr	n	th
544	*zemlesos*	dredger	instr	inan	ipfv	tr	n	th
545	*zmeelov*	snake hunter	ag	hum	ipfv	tr	n	th
546	*zmeejad*	short-toed snake eagle	ag	an	ipfv	tr	n	th
547	*zuboder*	dentist	ag	hum	ipfv	tr	n	th
548	*zuboskal*	scoffer	ag	hum	ipfv	tr	n	th
549	*iskusstvoved*	art critic	cos	hum	ipfv	tr	n	th
550	*kavkazoved*	expert on Caucasian languages	cos	hum	ipfv	tr	n	th
551	*kaznokrad*	embezzler of public funds	ag	hum	ipfv	tr	n	th
552	*kamenotes*	stonemason	ag	hum	ipfv	tr	n	th
553	*kamnerez*	stone carver	ag	hum	ipfv	tr	n	th
554	*kamnetes*	stone carver	ag	hum	ipfv	tr	n	th
555	*karakulevod*	karakul breeder	ag	hum	ipfv	tr	n	th
556	*kartofelevod*	potato breeder	ag	hum	ipfv	tr	n	th
557	*kaučukonos*	rubber-bearing plant	cos	inan	ipfv	tr	n	th
558	*kaševar*	cook (in military or working areas)	ag	hum	ipfv	tr	n	th
559	*kinoved*	film expert	cos	hum	ipfv	tr	n	th
560	*kitaeved*	sinologist	cos	hum	ipfv	tr	n	th
561	*kitoboj*	whale hunter	ag	hum	ipfv	tr	n	th
562	*kitolov*	whaleman	ag	hum	ipfv	tr	n	th
563	*kleevar*	worker preparing glue	ag	hum	ipfv	tr	n	th

564	*knigoved*	bibliologist	cos	hum	ipfv	tr	n	th
565	*knigoed*	bookworm	ag	an	ipfv	tr	n	th
566	*knigoed*	pedant	ag	hum	ipfv	tr	n	th
567	*knigoljub*	bibliophile	cos	hum	ipfv	tr	n	th
568	*knutoboj*	executioner	ag	hum	ipfv	tr	n	instr
569	*kožeder*	knacker	ag	hum	ipfv	tr	n	th
570	*kožeed*	skin beetle	ag	an	ipfv	tr	n	th
571	*kozovod*	goat breeder	ag	hum	ipfv	tr	n	th
572	*kozodoj*	nightjar (bird)	ag	an	ipfv	tr	n	th
573	*konevod*	horse breeder	ag	hum	ipfv	tr	n	th
574	*konoval*	horse doctor	ag	hum	ipfv	tr	n	th
575	*konovod*	horse holder	ag	hum	ipfv	tr	n	th
576	*konovjaz'*	rope to tie horses	instr	inan	ipfv	tr	n	th
577	*konogon*	horse driver	ag	hum	ipfv	tr	n	th
578	*konokrad*	horse thief	ag	hum	ipfv	tr	n	th
579	*konoplevod*	hemp grower	ag	hum	ipfv	tr	n	th
580	*kontejnerovoz*	container vehicle	instr	inan	ipfv	tr	n	th
581	*korneed*	longhorn beetle	ag	an	ipfv	tr	n	th
582	*koroed*	bark beetle	ag	an	ipfv	tr	n	th
583	*kosmolet*	spaceship	instr	inan	us	intr	n	path
584	*kostolom*	illness causing bones to break	ag	inan	ipfv	tr	n	th

585	*kostolom*	person who breaks bones in a fight	ag	hum	ipfv	tr	n	th
586	*kostoprav*	chiropractor	ag	hum	ipfv	tr	n	th
587	*kostorez*	bone carver	ag	hum	ipfv	tr	n	th
588	*kostrožog*	fire lighter	ag	hum	ipfv	tr	n	th
589	*kotločist*	boiler cleaner	ag	hum	ipfv	tr	n	th
590	*krabolov*	crabber	ag	hum	ipfv	tr	n	th
591	*krabolov*	crabber	instr	inan	ipfv	tr	n	th
592	*kraeved*	expert on a certain region	cos	hum	ipfv	tr	n	th
593	*krovosos*	leech	ag	hum	ipfv	tr	n	th
594	*krovosos*	exploiter	ag	an	ipfv	tr	n	th
595	*krolikovod*	rabbit breeder	ag	hum	ipfv	tr	n	th
596	*krochovod*	hairsplitter	ag	hum	ipfv	tr	n	th
597	*kružkovod*	study group leader	ag	hum	ipfv	tr	n	th
598	*krysolov*	rat catcher	ag	hum	ipfv	tr	n	th
599	*krjučkotvor*	pettifogger	ag	hum	ipfv	tr	n	th
600	*kuklovod*	puppeteer	ag	hum	ipfv	tr	n	th
601	*kukuruzovod*	corn grower	ag	hum	ipfv	tr	n	th
602	*kurovod*	hen breeder	ag	hum	ipfv	tr	n	th
603	*kuroed*	bribe-taker	ag	hum	ipfv	tr	n	th
604	*kurochvat*	bribe-taker	ag	hum	ipfv	tr	n	th
605	*kurocap*	bribe-taker	ag	hum	us	tr	n	th

606	*kurocap*	hen thief	ag	hum	us	tr	n	th
607	*kuroščup*	woman seducer	ag	hum	ipfv	tr	n	th
608	*lakovar*	worker preparing varnishes	ag	hum	ipfv	tr	n	th
609	*legkover*	credulous person	cos	hum	ipfv	tr	a	mnr
610	*legkodum*	light-minded person	cos	hum	ipfv	tr	a	mnr
611	*ledorez*	ice cutter	instr	inan	ipfv	tr	n	th
612	*lermontoved*	expert on Lermontov	cos	hum	ipfv	tr	n	th
613	*lesovod*	silviculturist	ag	hum	ipfv	tr	n	th
614	*lesovoz*	timber ship/lorry	instr	inan	ipfv	tr	n	th
615	*lesorub*	wood-cutter	ag	hum	ipfv	tr	n	th
616	*lisogon*	fox hunting dog	ag	an	ipfv	tr	n	th
617	*listoed*	leaf beetle	ag	an	ipfv	tr	n	th
618	*literaturoved*	literary critic	cos	hum	ipfv	tr	n	th
619	*lichodej*	malefactor	ag	hum	ipfv	tr	n	th
620	*licemer*	hypocrite	ag	hum	ipfv	tr	n	th
621	*lobotrjas*	loafer	ag	hum	ipfv	tr	n	th
622	*luboed*	peppergrass beetle	ag	an	ipfv	tr	n	th
623	*lugovod*	plant grower	ag	hum	ipfv	tr	n	th
624	*lunochod*	Lunochod	instr	inan	ipfv	intr	n	path
625	*lykoder*	dealing with tree barking	ag	hum	ipfv	tr	n	th
626	*l'novod*	flax grower	ag	hum	ipfv	tr	n	th

627	*ljudoed*	cannibal	ag	hum	ipfv	tr	n	th
628	*malover*	sceptic	cos	hum	ipfv	tr	a	mnr
629	*maslodel*	oiler	instr	inan	ipfv	tr	n	th
630	*materialoved*	expert on material science	cos	hum	ipfv	tr	n	th
631	*mašinoved*	expert on machines	cos	hum	ipfv	tr	n	th
632	*medovar*	mead maker	ag	hum	ipfv	tr	n	th
633	*medonos*	bee plant	ag	inan	ipfv	tr	n	th
634	*medosos*	honeyeater (bird)	ag	an	ipfv	tr	n	th
635	*merzlotoved*	expert on geocryology	cos	hum	ipfv	tr	n	th
636	*metalloved*	expert on physical metallurgy	cos	hum	ipfv	tr	n	th
637	*miroed*	exploiter	ag	hum	ipfv	tr	n	th
638	*molodoženy*	recently married	ag	hum	dvuvid	intr	a	time
639	*molokosos*	greenhorn	ag	hum	ipfv	tr	n	th
640	*mongoloved*	expert on mongolistics	cos	hum	ipfv	tr	n	th
641	*morechod*	seaman	ag	hum	ipfv	intr	n	path
642	*muzeeved*	expert on museology	cos	hum	ipfv	tr	n	th
643	*muzykoved*	expert on music	cos	hum	ipfv	tr	n	th
644	*mukoed*	grain beetle	ag	hum	ipfv	tr	n	th
645	*mukomol*	miller	ag	hum	ipfv	tr	n	th
646	*mukosej*	worker sifting flour	ag	hum	ipfv	tr	n	th
647	*murav'ed*	ant eater	ag	an	ipfv	tr	n	th

648	*musorovoz*	dustcart	instr	inan	ipfv	tr	n	th
649	*muchomor*	amanita (poisoning mushroom)	ag	inan	ipfv	tr	n	th
650	*mylovar*	worker preparing soap	ag	hum	ipfv	tr	n	th
651	*myšelov*	mice hunter	ag	an	ipfv	tr	n	th
652	*mjasoed*	meat eater	ag	hum	ipfv	tr	n	th
653	*narodoved*	expert on ethnography	cos	hum	ipfv	tr	n	th
654	*neftevoz*	oiler/oil tanker	instr	inan	ipfv	tr	n	th
655	*novosel*	new settler	ag	hum	ipfv	intr	a	time
656	*obščestvoved*	expert on social sciences	cos	hum	ipfv	tr	n	th
657	*ovoščevod*	vegetable grower	ag	hum	ipfv	tr	n	th
658	*ovcevod*	sheepman	ag	hum	ipfv	tr	n	th
659	*ognemet*	flame thrower	instr	inan	ipfv	tr	n	th
660	*odnodum*	person always thinking about the same	cos	hum	ipfv	intr	pro	th
661	*odnoljub*	one-woman man	cos	hum	ipfv	tr	num	th
662	*olenevod*	deer breeder	ag	hum	ipfv	tr	n	th
663	*osoed*	European honey buzzard	ag	an	ipfv	tr	n	th
664	*ochotoved*	expert on hunting economy	cos	hum	ipfv	tr	n	th
665	*panelevoz*	panel transporter	instr	inan	ipfv	tr	n	th
666	*parovoz*	steam locomotive	instr	inan	ipfv	tr	n	instr
667	*parochod*	steamboat	instr	inan	ipfv	intr	n	instr

668	*peskožil*	lugworm	cos	an	ipfv	intr	n	loc
669	*pešechod*	pedestrian	ag	hum	ipfv	intr	a	mnr
670	*pivovar*	brewer	ag	hum	ipfv	tr	n	th
671	*planetochod*	rover	instr	inan	ipfv	intr	n	path
672	*plodovod*	fruit grower	ag	hum	ipfv	tr	n	th
673	*plotovod*	raftman	ag	hum	ipfv	tr	n	th
674	*plotovod*	raft convoy	instr	inan	ipfv	tr	n	th
675	*plotogon*	raftman	ag	hum	ipfv	tr	n	th
676	*polevod*	field-crop grower	ag	hum	ipfv	tr	n	th
677	*poloter*	floor polisher	instr	inan	ipfv	tr	n	th
678	*posudomoj*	dishwasher	ag	hum	ipfv	tr	n	th
679	*počvoved*	soil scientist	cos	hum	ipfv	tr	n	th
680	*počerkoved*	expert on calligraphy	cos	hum	ipfv	tr	n	th
681	*pravdoljub*	truth lover	cos	hum	ipfv	tr	n	th
682	*pravoved*	jurist	cos	hum	ipfv	tr	n	th
683	*preljubodej*	adulterer	ag	hum	ipfv	tr	n	th
684	*prirodoved*	natural historian	cos	hum	ipfv	tr	n	th
685	*pticevod*	poultry breeder	ag	hum	ipfv	tr	n	th
686	*pticeed*	tarantula	ag	an	ipfv	tr	n	th
687	*pticelov*	bird catcher	ag	hum	ipfv	tr	n	th
688	*pustobolt*	windbag	ag	hum	ipfv	intr	a	mnr

689	*pustobrech*	dog barking with no reason	ag	an	us	intr	a	mnr
690	*pustobrech*	windbag	ag	hum	us	intr	a	mnr
691	*pustozvon*	windbag	ag	hum	ipfv	tr	a	mnr
692	*pustopljas*	frivolous person	ag	hum	ipfv	tr	a	mnr
693	*puškinoved*	expert on Puškin	cos	hum	ipfv	tr	n	th
694	*pčelovod*	beekeeper	ag	hum	ipfv	tr	n	th
695	*pylesos*	vacuum cleaner	instr	inan	ipfv	tr	n	th
696	*rakoed*	crab-eating racoon	ag	an	ipfv	tr	n	th
697	*rastenievod*	plant grower	ag	hum	ipfv	tr	n	th
698	*religioved*	expert on religious studies	cos	hum	ipfv	tr	n	th
699	*repolov*	linnet (bird)	ag	an	ipfv	tr	n	th
700	*risovod*	rice grower	ag	hum	ipfv	tr	n	th
701	*rifmoplet*	rhymer	ag	hum	ipfv	tr	n	th
702	*rotozej*	scatterbrain	cos	hum	ipfv	intr	n	th
703	*rudovoz*	ore ship	instr	inan	ipfv	tr	n	th
704	*rudokop*	miner	ag	hum	us	tr	n	th
705	*rukosuj*	clumsy person	ag	hum	ipfv	tr	n	th
706	*rybovod*	fish breeder	ag	hum	ipfv	tr	n	th
707	*rybolov*	fisherman	ag	hum	ipfv	tr	n	th
708	*rybochod*	fish-pass	instr	inan	ipfv	intr	n	th
709	*sadovod*	fruit grower	ag	hum	ipfv	tr	n	th

710	*samovar*	samovar	instr	inan	ipfv	tr	pro	mnr
711	*samogudy*	samogudy	instr	inan	ipfv	intr	pro	mnr
712	*samodur*	despot	ag	hum	ipfv	intr	pro	mnr
713	*samokat*	push scooter/bicycle	instr	inan	ipfv	tr	pro	mnr
714	*samolet*	airplane	instr	inan	us	intr	pro	mnr
715	*samolov*	trap	instr	inan	ipfv	tr	pro	mnr
716	*samoljub*	self-lover	cos	hum	ipfv	tr	pro	th
717	*samosad*	home-grown tobacco	ag	inan	ipfv	intr	pro	mnr
718	*samosval*	dump body truck	instr	inan	us	tr	pro	mnr
719	*samosev*	self-sown plant	ag	inan	ipfv	intr	pro	mnr
720	*samosej*	self-sown plant	ag	inan	ipfv	intr	pro	mnr
721	*samostrel*	soldier deliberately wounding himself	ag	hum	ipfv	tr	pro	mnr
722	*samostrel*	ancient bow/bad gun	instr	inan	ipfv	tr	pro	mnr
723	*samochval*	boaster	ag	hum	ipfv	tr	pro	th
724	*samochod*	self-traction machine	instr	inan	ipfv	intr	pro	mnr
725	*sacharovar*	expert on sugar production	ag	hum	ipfv	tr	n	th
726	*sveklovod*	beet grower	ag	hum	ipfv	tr	n	th
727	*svinoboj*	pig slaughterer	ag	hum	ipfv	tr	n	th
728	*svinovod*	pig breeder	ag	hum	ipfv	tr	n	th
729	*svinopas*	swineherd	ag	hum	ipfv	tr	n	th
730	*svobodoljub*	freedom lover	cos	hum	ipfv	tr	n	th

731	*sebjaljub*	self-lover	cos	hum	ipfv	tr	pro	th
732	*sekundomer*	stopwatch	instr	inan	ipfv	tr	n	th
733	*semenovod*	seed grower	ag	hum	ipfv	tr	n	th
734	*serdceved*	god	cos	hum	ipfv	tr	n	th
735	*serdceed*	lady-killer	ag	hum	ipfv	tr	n	th
736	*skalolaz*	rock-climber	ag	hum	ipfv	intr	n	path
737	*skirdoprav*	packer	ag	hum	ipfv	tr	n	th
738	*skorodum*	quick-thinking person	cos	hum	ipfv	intr	a	mnr
739	*skorochvat*	clever person	ag	hum	ipfv	tr	a	mnr
740	*skorochod*	fast runner	ag	hum	ipfv	intr	a	mnr
741	*skotovod*	cattle-breeder	ag	hum	ipfv	tr	n	th
742	*skotogon*	cattle driver	ag	hum	ipfv	tr	n	th
743	*slavjanoved*	slavist	cos	hum	ipfv	tr	n	th
744	*slavjanoljub*	slavophile	cos	hum	ipfv	tr	n	th
745	*slastoljub*	pleasure lover	cos	hum	ipfv	tr	n	th
746	*smolovar*	tar extractor	ag	hum	ipfv	tr	n	th
747	*smolokur*	tar extractor	ag	hum	ipfv	tr	n	th
748	*snegopach*	snow breaker	instr	inan	ipfv	tr	n	th
749	*snegochod*	snow mobile	instr	inan	ipfv	intr	n	path
750	*sobakoved*	expert on cynology	cos	hum	ipfv	tr	n	th
751	*sobakovod*	dog breeder	ag	hum	ipfv	tr	n	th

752	*solevar*	salt extractor	ag	hum	ipfv	tr	n	th
753	*spinogryz*	sprog	ag	hum	ipfv	tr	n	th
754	*stalevar*	steel melter	ag	hum	ipfv	tr	n	th
755	*starodum*	old-fashioned person	cos	hum	ipfv	intr	a	th
756	*starožil*	old resident	cos	hum	ipfv	intr	a	time
757	*steklovar*	glass melter	ag	hum	ipfv	tr	n	th
758	*stekloduv*	glass blower	ag	hum	ipfv	tr	n	th
759	*steklorez*	glass cutter	ag	hum	ipfv	tr	n	th
760	*steklorez*	glass cutter	instr	inan	ipfv	tr	n	th
761	*stenolaz*	wallcreeper (bird)	ag	an	ipfv	intr	n	path
762	*stichoplet*	bad rhymer	ag	hum	ipfv	tr	n	th
763	*suknoval*	cloth fuller	ag	hum	ipfv	tr	n	th
764	*suknodel*	cloth producer	ag	hum	ipfv	tr	n	th
765	*suchovej*	hot wind	ag	inan	ipfv	tr	n	mnr
766	*sčetovod*	accountant	ag	hum	ipfv	tr	n	th
767	*syrovar*	cheese maker	ag	hum	ipfv	tr	n	th
768	*syrodel*	cheese maker	ag	hum	ipfv	tr	n	th
769	*syroed*	raw vegan	ag	hum	ipfv	tr	a	th
770	*tabakovod*	tobacco grower	ag	hum	ipfv	tr	n	th
771	*tabakur*	one who smokes or snuffs tobacco	ag	hum	ipfv	tr	n	th
772	*teatroved*	expert on theatre	cos	hum	ipfv	tr	n	th

773	*teneljub*	mole-rat	cos	an	ipfv	tr	n	th
774	*teplovoz*	diesel locomotive	instr	inan	ipfv	tr	n	instr
775	*teploljub*	heat lover	cos	an	ipfv	tr	n	th
776	*teploljub*	heat lover	cos	inan	ipfv	tr	n	th
777	*teplomer*	calorimeter	instr	inan	ipfv	tr	n	th
778	*teplochod*	motor ship	instr	inan	ipfv	intr	n	instr
779	*testomes*	worker kneading dough	ag	hum	ipfv	tr	n	th
780	*tigrolov*	tiger hunter	ag	hum	ipfv	tr	n	th
781	*tichochod*	lazy-bones	ag	hum	ipfv	intr	a	mnr
782	*tichochod*	sloth	ag	an	ipfv	intr	a	mnr
783	*tovaroved*	expert on merchandising	cos	hum	ipfv	tr	n	th
784	*tonkoprjad*	butterfly	ag	an	ipfv	tr	a	mnr
785	*trubovoz*	truck used to transport pipes	instr	inan	ipfv	tr	n	th
786	*trubočist*	chimney sweep	ag	hum	ipfv	tr	n	th
787	*trudoljub*	hard worker	cos	hum	ipfv	tr	n	th
788	*tugodum*	slowcoach	cos	hum	ipfv	intr	a	mnr
789	*tjaželovoz*	carthorse	ag	an	ipfv	tr	a	th
790	*tjaželodum*	slow-witted person	cos	hum	ipfv	tr	a	mnr
791	*tjažkodum*	slow-witted person	cos	hum	ipfv	tr	a	mnr
792	*uglevoz*	ship for coal transport	instr	inan	ipfv	tr	n	th
793	*ugležog*	charcoal burner	ag	hum	ipfv	tr	n	th

794	*uglekop*	miner	ag	hum	ipfv	tr	n	th
795	*chlebopek*	baker	ag	hum	ipfv	tr	n	th
796	*chleborez*	bread cutter	ag	hum	ipfv	tr	n	th
797	*chleborob*	grain grower	ag	hum	ipfv	tr	n	th
798	*chlopkovod*	cotton grower	ag	hum	ipfv	tr	n	th
799	*chlopkorob*	cotton grower	ag	hum	ipfv	tr	n	th
800	*chmelevod*	hop grower	ag	hum	ipfv	tr	n	th
801	*cvetovod*	floriculturist	ag	hum	ipfv	tr	n	th
802	*cvetoljub*	flower lover	cos	hum	ipfv	tr	n	th
803	*cvetonos*	flower stalk	cos	inan	ipfv	tr	n	th
804	*citrusovod*	citriculturist	ag	hum	ipfv	tr	n	th
805	*čaevod*	tea grower	ag	hum	ipfv	tr	n	th
806	*čarodej*	magician	ag	hum	ipfv	tr	n	th
807	*častotomer*	frequency counter	instr	inan	ipfv	tr	n	th
808	*čelovekoljub*	humanist	cos	hum	ipfv	tr	n	th
809	*červovod*	(silk)worm breeder	ag	hum	ipfv	tr	n	th
810	*čertogon*	devil's-bit (plant)	ag	inan	ipfv	tr	n	th
811	*čertolom*	rude person	ag	hum	ipfv	tr	n	th
812	*čertopoloch*	thistle	cos	inan	ipfv	tr	n	th
813	*čestoljub*	ambitious man	cos	hum	ipfv	tr	n	th
814	*čechoved*	expert on Čechov	cos	hum	ipfv	tr	n	th

815	*čistopljuj*	cleany person	ag	hum	ipfv	intr	a	mnr
816	*čudodej*	miracle maker	ag	hum	ipfv	tr	n	th
817	*šapoval*	hat-maker	ag	hum	ipfv	tr	n	th
818	*šekspiroved*	expert on Shakespeare	cos	hum	ipfv	tr	n	th
819	*šelkovod*	silkworm breeder	ag	hum	ipfv	tr	n	th
820	*šelkoprjad*	silkworm	ag	an	ipfv	tr	n	th
821	*šerstobit*	wool beater	ag	hum	ipfv	tr	n	th
822	*šerstoboj*	wool beater	instr	inan	ipfv	tr	n	th
823	*škuroder*	exploiter	ag	hum	ipfv	tr	n	th
824	*èkskursovod*	tour guide	ag	hum	ipfv	tr	n	th
825	*èlektrovoz*	electric locomotive	instr	inan	ipfv	tr	a	instr
826	*èlektrochod*	motor ship	instr	inan	ipfv	intr	a	instr
827	*èfironos*	oil-bearing plant	cos	inan	ipfv	tr	n	th
828	*jazykoved*	linguist	cos	hum	ipfv	tr	n	th
829	*jajcevod*	oviduct	instr	inan	ipfv	tr	n	th
830	*jajceed*	egg eater (insect)	ag	an	ipfv	tr	n	th
831	*japonoved*	nipponist	cos	hum	ipfv	tr	n	th

Appendix 2

Number of tokens and first occurrence of each compound in the RNC

		lemma	meaning	tokens RNC	first occurrence RNC
-ec					
	1	*avianosec*	aircraft carrier	406	1939
	2	*basnopisec*	fable writer	133	1811
	3	*bogoborec*	theomachist	75	1811-1818
	4	*bogomolec*	devotee	1413	1682-1709
	5	*bogonosec*	god-bearing	189	1733
	6	*borzopisec*	careless writer	89	1825-1826
	7	*borzochodec*	fast runner	3	1938
	8	*bronenosec*	armadillo	2068	1860
	9	*bronenosec*	battleship	-	1860
	10	*bykoboec*	bull fighter	10	1911
	11	*vencenosec*	monarch	323	1788-1822
	12	*vertoletonosec*	helicopter carrier	12	1975
	13	*verchovodec*	leader	1	1843-1847
	14	*vinoprodavec*	wine seller	7	1829

15	*vinotorgovec*	wine trader	98	1846-1849
16	*vlastoderžec*	monarch	3	1891
17	*vlastoljubec*	power lover	85	1769
18	*vodonosec*	water carrier	6	1826
19	*vol'nodumec*	free thinker	398	1779-1790
20	*vol'noljubec*	freedom lover	7	1869-1870
21	*vsevidec*	all-seeing	8	1720
22	*gorododeržec*	city ruler	1	1922
23	*gradoderžec*	city ruler	1	1918
24	*gromoveržec*	thunder thrower	220	1793
25	*dvoeborec*	biathlon competitor	7	2002
26	*desjatiborec*	decathlon competitor	46	1919
27	*drevotočec*	cossid miller	63	1864-1874
28	*duchoborec*	spirit warrior	90	1824
29	*duchovidec*	clairvoyant	68	1827
30	*dušegubec*	murderer, torturer	146	1766
31	*dušelovec*	soul catcher	2	1960-1970
32	*edinoborec*	one-to-one fighter	24	1803-1818
33	*edinoderžec*	monarch	10	1941
34	*žgutikonosec*	flagellate	20	1886
35	*žezlonosec*	crosier bearer	29	1888

36	*ženoljubec*	ladies' man	22	1750
37	*živopisec*	painter	3088	1702
38	*žizneljubec*	life lover	20	1877
39	*zaimodavec*	creditor	310	1714
40	*zajmodavec*	creditor	26	1995
41	*zvezdonosec*	military decoration bearer	31	1814
42	*zemlepašec*	farmer	201	1766
43	*zemleprochodec*	explorer	131	1891
44	*zmeeborec*	snake defeater	15	1907
45	*zmeenosec*	serpent-bearer (constellation)	86	1871
46	*znamenosec*	flag-bearer	212	1811-1818
47	*ikonoborec*	iconoclast	66	1790
48	*ikonopisec*	icon painter	560	1726
49	*kanatochodec*	rope-walker	132	1901
50	*kitoboec*	whaler	14	1936
51	*knigoderžec*	book-bearer	3	1894
52	*knigoprodavec*	book seller	507	1748
53	*knigotorgovec*	book trader	82	1826-1855
54	*kon'kobežec*	skater	303	1878
55	*kop'enosec*	spare-bearer	23	1895
56	*korystoljubec*	profit-seeker	88	1765

57	*krestonosec*	Crusader	439	1806-1818
58	*krovopivec*	oppressor	127	1854
59	*krovopiec*	oppressor	291	1779-1790
60	*krjučkotvorec*	pettifogger	29	1772
61	*lesotorgovec*	wood trader	41	1880
62	*letopisec*	chronicler	2154	1733
63	*lichoimec*	bribe-taker	112	1751-1752
64	*lozochodec*	dowser	3	1997
65	*lyžebežec*	skier	33	1902
66	*medalenosec*	medal-bearer	6	1941-1943
67	*mečenosec*	sword-bearer	95	1750
68	*mečenosec*	swordtail	-	1750
69	*mzdoimec*	bribe-taker	138	1761
70	*minonosec*	torpedo boat	1958	1860
71	*miroderžec*	god	18	1757
72	*miroljubec*	peace lover	17	1750
73	*mirotvorec*	peace maker	535	1790
74	*mnogoborec*	multisport athlete	22	1913
75	*molotoboec*	hammerer	185	1880
76	*morechodec*	sailor	167	1766-1794
77	*muželožec*	homosexual	11	1761

78	*narkotorgovec*	drug dealer	81	1997
79	*narodoljubec*	people lover	62	1852-1874
80	*odnodumec*	person always thinking about the same	3	2001
81	*ordenonosec*	order-bearer	274	1920-1943
82	*oruženosec*	sword-bearer	475	1709
83	*očevidec*	eye witness	2319	1770
84	*pervoprochodec*	earliest explorer	366	1915
85	*pesnopevec*	writer/singer of religious songs	91	1726
86	*pesnotvorec*	song writer	71	1734
87	*pis'monosec*	postman	69	1766-1768
88	*polkovodec*	commander	3299	1682-1709
89	*porfironosec*	monarch	3	1788-1822
90	*pravdoborec*	freedom fighter	1	1988-1998
91	*pravdoljubec*	truth lover	112	1808-1822
92	*psalmopevec*	psalmist	137	1717
93	*pjatiborec*	pentathlete	24	1957
94	*rabotorgovec*	slave trader	85	1890
95	*raketonosec*	rocket carrier	59	1970
96	*ratoborec*	battle fighter	42	1785-1786
97	*rogonosec*	cuckold	159	1765
98	*rudoznatec*	expert on minerals	37	1859

99	*rybotorgovec*	fish trader	15	1911
100	*samovidec*	eye witness	71	1733
101	*samoderžec*	autocrat	1181	1682-1709
102	*samoljubec*	self-lover	25	1765
103	*samopisec*	self-recorder	95	1933-1937
104	*samoupravec*	tyrant	28	1862
105	*svetoderžec*	lord of the world	1	1922
106	*sveščenosec*	sacristan	30	1816-1820
107	*svobodoljubec*	freedom-lover	28	1718
108	*sebjaljubec*	self-lover	75	1829
109	*serdcevedec*	god	80	1757
110	*serdcevidec*	god	9	1717
111	*skoropisec*	shorthand typist	18	1750
112	*skotoložec*	zoophile	4	1988
113	*skototorgovec*	cattle trader	3	1913
114	*slavoljubec*	glory lover	5	1861
115	*sladkopevec*	melodist	28	1832
116	*slastoljubec*	pleasure lover	147	1764
117	*slovotvorec*	word creator	9	1928
118	*snovidec*	dream seer	46	1862-1875
119	*srebroljubec*	avid person	53	1757

120	*stenopisec*	wall painter	4	1922
121	*strannoljubec*	foreigners lover	2	1775
122	*strannopriimec*	shelter owner	35	1751
123	*strastoterpec*	martyr	179	1750
124	*suchojadec*	dry food eater	3	1849
125	*tajnovidec*	clairvoyant	31	1776
126	*tiranoborec*	tyranny fighter	9	1940-1954
127	*torpedonosec*	torpedo boat/bomber	40	1927
128	*trudoljubec*	hard worker	30	1760-1775
129	*tunejadec*	parasite	520	1722
130	*fakelonosec*	Olympic torch carrier	41	1874
131	*flotovodec*	naval commander	136	1826
132	*chlebopašec*	ploughman	231	1767
133	*chlebotorgovec*	corn merchant	58	1869
134	*chorugvenosec*	gonfalon bearer	43	1881-1889
135	*christoprodavec*	traitor	114	1835
136	*čaetorgovec*	tea dealer	18	1891
137	*čelovekoljubec*	people lover	87	1717
138	*čestoljubec*	ambitious man	329	1766-1794
139	*čudotvorec*	saint who makes miracles	1147	1682-1709
140	*čudotvorec*	miracle man	-	1682-1709

141	*ščitonosec*	shield bearer	23	1817
142	*jazykotvorec*	word creator	7	1970-1972
143	*jasnovidec*	clairvoyant	109	1846

-lec

144	*gruzovladelec*	owner of goods	69	1914
145	*domovladelec*	landlord	889	1830-1870
146	*zavodovladelec*	factory owner	37	1861
147	*zemlevladelec*	land owner	1544	1841
148	*rabovladelec*	slave owner	200	1862
149	*sudovladelec*	ship owner	199	1888
150	*tovarovladelec*	owner of goods	19	1894-1899

-tel'

151	*aviastroitel'*	aircraft builder	63	1932-1942
152	*avtoljubitel'*	car enthusiast	342	1951
153	*avtomobilestroitel'*	car builder	67	1951
154	*avtoproizvoditel'*	car producer	221	1996
155	*avtostroitel'*	car builder	25	1970
156	*arendodatel'*	landlord	229	1989
157	*balansoderžatel'*	accountant	14	2001

158	*bacillonositel'*	bacilli carrier	12	1938-1939
159	*betonosmesitel'*	concrete mixer	8	2003
160	*blagodetel'*	benefactor	3237	1709
161	*blagoželatel'*	benevolent person	59	1862-1875
162	*blagotvoritel'*	benefactor	426	1766-1777
163	*blagoustroitel'*	benefactor	9	1869
164	*bogoiskatel'*	god seeker	60	1908
165	*bogostroitel'*	god builder	18	1908
166	*bomboderžatel'*	bomb rack	13	1930
167	*bombometatel'*	bomber	43	1906
168	*bumagoderžatel'*	paper holder	3	1928-1931
169	*bumagomaratel'*	scribbler	17	1769
170	*bytopisatel'*	life writer	174	1803-1818
171	*vagonooprokidyvatel'*	wagon tippler	4	1977
172	*vagonostroitel'*	carriage builder	6	1974
173	*vekseledatel'*	bill of exchange drawer	36	1879
174	*vekselederžatel'*	bill of exchange owner	39	1886
175	*veroučitel'*	dogma interpreter	58	1828
176	*vertoletostroitel'*	helicopter builder	5	1975
177	*vzjatkodatel'*	briber	62	1869-1872
178	*vzjatkopolučatel'*	bribe taker	11	1996

179	*vidoiskatel'*	view finder	153	1926
180	*virusonositel'*	virus carrier	32	1989
181	*vodonagrevatel'*	water heater	48	1952
182	*vodootdelitel'*	water separator	4	1999
183	*vodopol'zovatel'*	water user	92	1980-1984
184	*vodopotrebitel'*	water consumer	12	1995
185	*vozduchonagrevatel'*	air heater	12	1959
186	*vozduchoočistitel'*	air purifier	35	1872-1876
187	*vozduchoplavatel'*	aeronaut	296	1780-1914
188	*vozduchopodogrevatel'*	air heater	28	1964
189	*vol'noslušatel'*	auditor	113	1860
190	*vsederžitel'*	the almighty	320	1720
191	*vygodopriobretatel'*	beneficiary	103	1995
192	*gidroraspredelitel'*	hidraulic control valve	5	2003
193	*glubokorychlitel'*	subsurface plow	1	1983
194	*gornospasatel'*	mine rescuer	21	1970-1975
195	*gorovoschoditel'*	alpinist	8	1956
196	*gradopravitel'*	city ruler	49	1733
197	*gradostroitel'*	city builder	87	1926
198	*grobokopatel'*	grave digger	63	1968-1969
199	*gromkogovoritel'*	loudspeaker	640	1920-1943

200	*gruzootpravitel'*	consignor of goods	94	1885-1886
201	*gruzopolučatel'*	consignee	49	1903
202	*deloproizvoditel'*	secretary	429	1809
203	*dobrodetel'*	benefactor	8309	1716
204	*dobroželatel'*	benevolent person	544	1719
205	*dolgožitel'*	long liver	267	1958-1973
206	*domopravitel'*	house ruler	65	1726
207	*domostroitel'*	house builder	34	1718
208	*estestvoispytatel'*	naturalist	720	1730
209	*zajmoderžatel'*	loan holder	1	1934
210	*zakonodatel'*	legislator	2065	1758
211	*zakonoučitel'*	teacher of god's law	241	1765
212	*zalogodatel'*	depositor	187	1858
213	*zalogoderžatel'*	pawnbroker	163	1875
214	*zvukonositel'*	music tape	4	1998
215	*zvukooformitel'*	sound operator	9	1950-1959
216	*zvukopodražatel'*	sound imitator	3	1933
217	*zvukosnimatel'*	pickup	19	1962
218	*zvukoulovitel'*	acoustic mirror	10	1940-1973
219	*zvukousilitel'*	loud-hailer	3	1933
220	*zemlepol'zovatel'*	land tenant	88	1925

221	*zemleustroitel'*	land organizer	46	1904
222	*zloželatel'*	malevolent person	14	1780-1814
223	*zlopychatel'*	malevolent person	99	1868
224	*zolotoiskatel'*	gold digger	210	1855
225	*igloderžatel'*	needle holder	16	1917
226	*iskrogasitel'*	spark arrester	7	1870
227	*kanavokopatel'*	trench digger	20	1950
228	*kvartironanimatel'*	tenant	23	1901
229	*kinoljubitel'*	amateur filmmaker	40	1951
230	*kladoiskatel'*	treasure seeker	129	1861
231	*knigoizdatel'*	publisher	55	1891
232	*knigorasprostranitel'*	book spreader	5	1999
233	*kožzamenitel'*	leatherette	63	1962
234	*kop'emetatel'*	javelin thrower	8	1924
235	*korablestroitel'*	ship builder	120	1822
236	*krovezamenitel'*	blood substitute	33	1965
237	*krovosmesitel'*	incestuous person	21	1870
238	*lesozagotovitel'*	timber cutting worker	110	1928
239	*lesopol'zovatel'*	wood user	145	1980-1984
240	*lozoiskatel'*	dowser	8	2003
241	*maslootražatel'*	oil deflector	2	2004

242	*mašinostroitel'*	machine builder	178	1907
243	*mestobljustitel'*	locum tenens	208	1809-1820
244	*metalloiskatel'*	metal detector	109	1939
245	*metrostroitel'*	underground builder	9	1977
246	*minoiskatel'*	mine detector	101	1941-1945
247	*moreplavatel'*	sailor	846	1726
248	*mostostroitel'*	bridge builder	23	1959
249	*motorostroitel'*	motor builder	99	1969
250	*musorosžigatel'*	incinerator	3	1997
251	*najmodatel'*	renter	66	1978
252	*nasledodatel'*	testator	115	1914
253	*nebožitel'*	celestial being	306	1794-1795
254	*nebokoptitel'*	loafer	12	1903-1905
255	*nedobroželatel'*	malevolent person	645	1759
256	*nedropol'zovatel'*	subsurface user	93	1992
257	*nomeronabiratel'*	dial	5	1977
258	*ognetušitel'*	fire extinguisher	273	1902
259	*otkazopolučatel'*	legatee	20	1993
260	*očkovtiratel'*	impostor	27	1923
261	*paroochladitel'*	steam cooler	2	1999
262	*paroočistitel'*	fallow fields purifier	7	2002

263	*paroperegrevatel'*	superheater	67	1912
264	*patentoobladatel'*	patent holder	61	1992
265	*penkosnimatel'*	cream-skimmer	119	1852-1874
266	*penoobrazovatel'*	foaming agent	10	1965
267	*pervovoschoditel'*	first climber	10	1934
268	*pervootkryvatel'*	pioneer	481	1927
269	*plamegasitel'*	flame arrester	26	1938
270	*poklažedatel'*	depositor	62	1995
271	*polotencesušitel'*	towel drier	28	1997-2008
272	*počvoobrazovatel'*	soil former	9	1889
273	*pravdoiskatel'*	truth searcher	46	1913-1914
274	*pravonarušitel'*	transgressor	289	1920-1943
275	*pravoobladatel'*	right holder	235	1992
276	*priborostroitel'*	instrument builder	15	1965
277	*prirodopol'zovatel'*	environment user	19	1990
278	*pristanoderžatel'*	resetter	20	1829
279	*pustynnožitel'*	hermit	91	1760-1775
280	*putevoditel'*	guide	1002	1757
281	*pyleulovitel'*	dust collector	30	1963
282	*pjatnovyvoditel'*	spot remover	41	1965
283	*rabotodatel'*	employer	1863	1886

284	*radioljubitel'*	radio amateur	1428	1925
285	*radioslušatel'*	radio listener	265	1926
286	*raketonositel'*	missile carrier	47	1958
287	*raketostroitel'*	missile builder	15	1959
288	*reestroderžatel'*	register holder	25	1995
289	*reklamodatel'*	advertiser	472	1925-1926
290	*reklamoproizvoditel'*	advertising producer	7	1997
291	*reklamorasprostranitel'*	advertising spreader	8	2001
292	*samoletostroitel'*	aircraft builder	20	1950
293	*samoučitel'*	manual for self-tuition	153	1833
294	*sacharozamenitel'*	sugar substitute	18	1998-1999
295	*svetovozvraščatel'*	retro-reflector	12	1978
296	*skorosšivatel'*	loose-leaf binder	56	1927
297	*snegoočistitel'*	snow plough	69	1912
298	*ssudodatel'*	lender	28	1904
299	*ssudopolučatel'*	bailee	28	1931
300	*stankostroitel'*	machine-tool builder	41	1947
301	*stekloobrazovatel'*	glass forming element	17	1968
302	*stekloočistitel'*	screen wiper	54	1961
303	*stichoslagatel'*	verse writer	8	1875-1881
304	*stogometatel'*	hay stacker	2	1959

305	*sudovoditel'*	navigator	95	1932-1935
306	*sudostroitel'*	ship builder	157	1851
307	*teplonositel'*	heat carrier	228	1965
308	*termopreobrazovatel'*	heat transformer	2	2004
309	*tovaroproizvoditel'*	commodity producer	259	1894
310	*traktorostroitel'*	tractor builder	25	1951
311	*tranšeekopatel'*	trencher	6	1960-1969
312	*turbostroitel'*	turbine builder	12	1974
313	*urokodatel'*	lesson-giver	2	1971
314	*fondoobrazovatel'*	fund former	2	1993
315	*fotoljubitel'*	photography amateur	87	1933-1937
316	*chlebozagotovitel'*	state grain purchaser	12	1927
317	*chramozdatel'*	temple builder	9	1880
318	*čekodatel'*	cheque drawer	10	1993
319	*čekoderžatel'*	drawn cheque owner	9	1993
320	*členovreditel'*	guilty of mutilation	11	1854-1858
321	*špagoglotatel'*	sword swallower	20	1908
322	*ščitoderžatel'*	shield holder	5	1990
323	*ènergonositel'*	energy carrier	459	1987
324	*ènergostroitel'*	energy builder	7	1968-1991
325	*jamokopatel'*	hole digger	4	1869

-nik

326	*bogochul'nik*	blasphemer	122	1772
327	*bortprovodnik*	air steward	26	1980
328	*verootstupnik*	faith abandonner	27	1877
329	*vozduchozabornik*	air intake	137	1962
330	*grechovodnik*	sinner	148	1782
331	*gruzopod"emnik*	goods elevator	6	1921
332	*derevoobdeločnik*	wood worker	38	1913
333	*ženonenavistnik*	woman hater	55	1892
334	*zakonoprestupnik*	criminal	14	1719
335	*idolopoklonnik*	idolater	164	1715
336	*kalopriemnik*	colostomy bag	3	2000
337	*kljatvoprestupnik*	perjurer	140	1715
338	*močetočnik*	ureter	51	1925
339	*musorosbornik*	garbage bin	21	1971
340	*nizkopoklonnik*	toady	35	1829
341	*ognepoklonnik*	fire worshipper	101	1809-1820
342	*pravozaščitnik*	human rights activist	658	1933
343	*pravopreemnik*	assignee	218	1903
344	*prestolonaslednik*	successor to the throne	58	1895

345	*rukomojnik*	washstand	537	1755
346	*solncepoklonnik*	sun worshipper	15	1910
347	*steklopod"emnik*	window raiser	70	1959
348	*sudopod"emnik*	ship elevator	5	1909
349	*teploobmennik*	heat exchanger	215	1939-1941
350	*tokopriemnik*	current collector	35	1953
351	*tokos"emnik*	current collector	5	1974
352	*fotopriemnik*	photodetector	23	1976
353	*čelovekonenavistnik*	misanthrope	75	1789

-ščik/čik

354	*astroletčik*	astronaut	21	1956
355	*asfal'toukladčik*	asphalt paver	28	1963
356	*benzozapravščik*	refueller	19	1937
357	*gazosvarščik*	gas welder	7	1956
358	*granatometčik*	grenadier	78	1930
359	*gruzoperevozčik*	cargo company	13	1998
360	*zvezdoletčik*	astronaut	52	1961
361	*kabeleukladčik*	cable handler	6	1968-1990
362	*kvartiros"emščik*	tenant	138	1957
363	*kondensatootvodčik*	steam trap	3	1999

364	*maslobojščik*	oil miller	3	1863-1864
365	*mostoukladčik*	bridge-layer	6	1999
366	*puskonaladčik*	motion adjuster	1	2003
367	*puteobchodčik*	track man	6	1952
368	*puteukladčik*	track layer	23	1949
369	*slovolitčik*	type founder	2	1858
370	*stalelitejščik*	steel founder	4	1951
371	*sudosborščik*	shipwright	4	1970
372	*toplivozapravščik*	refueller	17	1976
373	*troubukladčik*	pipelaying crane	35	1952
374	*troubukladčik*	pipe layer	-	1952
375	*čelobitčik*	petitioner	250	1682-1709
376	*čugunolitejščik*	iron founder	4	1959
377	*l'notrebil'ščik*	flax puller	3	1957
378	*nalogoplatel'ščik*	tax payer	1358	1830
379	*staleplavil'ščik*	steel maker	4	1951
380	*šelkomotal'ščik*	cocoon unwinder	1	1904-1956

-ka

381	*belošvejka*	seamstress	120	1846
382	*bumagomaraka*	paper-stainer	24	1836

383	*vodokačka*	water-tower	390	1878
384	*vodomerka*	pond skater	67	1923-1958
385	*vozduchoduvka*	compressed air releaser	21	1896
386	*vseznajka*	know-all	95	1843-1847
387	*dušegrejka*	shrug	119	1741-1752
388	*dušegubka*	mobile gas chamber	161	1833
389	*zemlerojka*	shrew	93	1847
390	*zolotošvejka*	gold threads master	24	1789
391	*zubočistka*	toothpick	281	1817
392	*kožemjaka*	leather manufacturer	55	1846-1857
393	*kofevarka*	coffee-machine	231	1964
394	*kofemolka*	coffee-grinder	82	1964
395	*kraskoterka*	paint grinder	1	1975
396	*kruporuška*	cereal transformer	30	1874
397	*krysolovka*	rat trap	31	1894
398	*krysolovka*	rat catcher	-	1894
399	*maloežka*	person who does not eat much	10	1966-1970
400	*maslobojka*	butter churn	111	1863
401	*mnogoznajka*	know-all	5	1923
402	*mucholovka*	fly trap	102	1847
403	*mucholovka*	house centipede	-	1847

404	*mucholovka*	Venus flytrap (plant)	-	1847
405	*myšelovka*	mouse trap	494	1831
406	*mjasorubka*	meat grinder	586	1910-1911
407	*nosogrejka*	tobacco pipe	55	1855-1858
408	*ovoščečistka*	vegetable peeler	2	2004
409	*polomojka*	scrubwoman	68	1842
410	*poluznajka*	half-educated person	17	1895-1900
411	*posudomojka*	kitchen porter	121	1961
412	*posudomojka*	dishwasher	-	1961
413	*prosoruška*	millet-scourer	6	1907
414	*pustolajka*	dog barking for no reason	4	1970
415	*samosbroska*	sliced bread dropper	4	1967
416	*samosejka*	self-sown plant	5	1982
417	*samoskidka*	sliced bread dropper	1	1937-1940
418	*samoučka*	autodidact	777	1782
419	*skorovarka*	pressure cooker	29	1951
420	*sladkoežka*	sweet tooth	63	1877
421	*slastoežka*	sweet tooth	5	1926
422	*solomorezka*	straw cutter	14	1915
423	*sudomojka*	dishwasher	189	1841
424	*syroežka*	russula (mushroom)	226	1822

425	*telogrejka*	padded jacket	982	1794-1795
426	*uchovertka*	earwig	22	1747
427	*chleborezka*	bread slicer	76	1935

-lka

428	*betonomešalka*	concrete mixer	131	1926
429	*gazonokosilka*	lawnmower	121	1959
430	*zemlečerpalka*	dredger	54	1910
431	*zernosušilka*	grain dryer	33	1867
432	*kamnedrobilka*	stone breaker	27	1926
433	*kartofelesažalka*	potato planter	5	1977
434	*kormodrobilka*	forage cutter	1	1960
435	*senokosilka*	mowing machine	67	1886-1887
436	*snegotajalka*	snow melter	20	1901
437	*snopovjazalka*	grain binder	25	1913-1914
438	*sokovyžimalka*	squeezer	113	1958

-ø

439	*avtovoz*	car transporter	18	1975
440	*atomochod*	atomic-powered vessel	157	1959
441	*bachčevod*	melon grower	3	1929

442	*benzovoz*	tank truck	81	1959
443	*bljudoliz*	lickspittle	23	1789
444	*bogomaz*	icon painter	99	1860-1861
445	*bogomol*	mantis	551	1741-1752
446	*bolotochod*	swamp mobile	6	1977
447	*bombovoz*	bomber	87	1926
448	*bradobrej*	barber	119	1825-1843
449	*brakodel*	bungler	61	1930-1963
450	*bukvoed*	pedant	52	1826-1855
451	*vezdechod*	off-road vehicle	661	1932-1942
452	*verchovod*	leader	34	1862
453	*verchogljad*	superficial person	59	1717
454	*vercholaz*	steeplejack	36	1861-1863
455	*verchochvat*	superficial person	3	1876
456	*vetroduv*	ventilator	3	1968-1991
457	*vetroduj*	ventilator	7	1928-1933
458	*vetromer*	anemometer	2	1929
459	*vzjatkochvat*	bribe-taker	2	1770
460	*vinoglot*	alchoolic	1	2001
461	*vinodel*	wine maker	231	1828
462	*vinokur*	distiller	341	1812

463	*vodovoz*	water carrier	251	1789
464	*vodovoz*	water carrier	-	1789
465	*vodovypusk*	culvert	9	1950
466	*vodolaz*	diver	1163	1795
467	*vodolaz*	Newfoundland dog	-	1795
468	*vodolej*	Aquarius	373	1747
469	*vodolej*	worker pouring water	-	1747
470	*vodoliv*	worker pouring water	100	1770
471	*vodoljub*	water scavenger beetle	33	1910
472	*vodomer*	water gauge	37	1782
473	*vodonos*	water carrier	107	1792
474	*vodosbros*	spillway	26	1950
475	*vodospusk*	floodgate	12	1862
476	*vodostok*	drain pipe	166	1870
477	*vozduchoduv*	compressed air releaser	6	1933
478	*volkodav*	wolfhound	750	1839
479	*volnolom*	breakwater	57	1873
480	*volopas*	Bootes (constellation)	44	1905
481	*volosočes*	hairdresser	8	1789
482	*vol'nodum*	free thinker	13	1784-1792
483	*vostokoved*	orientalist	199	1911-1916

484	*vysotomer*	altimeter	72	1925
485	*gazomer*	gas-meter	2	1997
486	*golovorez*	daredevil	480	1788
487	*golovotjap*	bungler	102	1869-1870
488	*golubevod*	dove breeder	13	1935
489	*goremyka*	poor wretch	306	1825
490	*gorloder*	screamer	33	1860
491	*gorlochvat*	boor, arrogant person	93	1819
492	*gosudarstvoved*	political scientist	47	1909
493	*granatomet*	grenade launcher	460	1933
494	*gromootvod*	ligthning rod	281	1843
495	*gruzinoved*	expert on Georgian	2	1976
496	*gruntoved*	soil engineer	4	1936
497	*gruppovod*	group leader	10	1934
498	*gubošlep*	mumbler	172	1861-1863
499	*gurtoprav*	drover	13	1857
500	*gusevod*	goose breeder	1	1928-1935
501	*dal'nomer*	rangefinder	128	1914
502	*darmoglot*	sponger	1	1987-1999
503	*darmoed*	sponger	435	1790
504	*dobrochot*	benevolent person	249	1733

505	*dolgodum*	slow-witted person	3	1909
506	*domovod*	housewife	4	1787
507	*domosed*	homebody	183	1792
508	*domouprav*	house manager	77	1933-1965
509	*drevogryz*	powderpost beetle	3	2002
510	*drevolaz*	poison dart frog	4	1887
511	*drovokol*	woodchopper	12	1849
512	*drovorub*	woodcutter	30	1723
513	*drovosek*	woodcutter	708	1721
514	*dubolom*	stupid person	147	1888
515	*dubotolk*	stupid person	8	1863
516	*durolom*	stupid person	26	1869
517	*duropljas*	stupid person	4	1924-1932
518	*duchobor*	spirit warrior	283	1890-1899
519	*dušegub*	murderer, torturer	345	1833-1834
520	*dymokur*	bonfire against mosquitoes	66	1841
521	*dymochod*	flue	258	1905
522	*estestvoved*	expert on natural science	10	1894-1899
523	*ženoljub*	ladies' man	37	1882
524	*živoglot*	exploiter	100	1856-1857
525	*živoglot*	snaketooth fish	-	1856-1857

526	*živoder*	knacker	259	1787
527	*životnovod*	cattle-breeder	308	1909
528	*žizneljub*	life lover	81	1906-1913
529	*zakonoved*	jurist	38	1836
530	*zvezdolet*	spaceship	551	1935
531	*zvezdochvat*	pickpocket	3	2000
532	*zvezdočet*	astrologer	432	1781
533	*zvezdočet*	stargazer (fish)	-	1781
534	*zveroboj*	ochotnik	300	1785
535	*zveroboj*	St. John's wort (plant)	-	1785
536	*zveroboj*	vodka made from St. John's wort	-	1785
537	*zverovod*	fur farm operator	23	1930-1953
538	*zverolov*	hunter	118	1803-1818
539	*zemleved*	expert on physical geography	1	1918-1921
540	*zemlevoz*	worker transporting earth	1	1987
541	*zemlekop*	digger	428	1812
542	*zemlemer*	land surveyor	679	1747
543	*zemlerob*	farmer	56	1908
544	*zemlesos*	dredger	19	1949
545	*zmeelov*	snake hunter	25	1966
546	*zmeejad*	short-toed snake eagle	2	2003

547	*zuboder*	dentist	28	1871
548	*zuboskal*	scoffer	118	1769
549	*iskusstvoved*	art critic	805	1926
550	*kavkazoved*	expert on Caucasian languages	9	1913
551	*kaznokrad*	embezzler of public funds	150	1826-1855
552	*kamenotes*	stonemason	102	1871-1908
553	*kamnerez*	stone carver	49	1936
554	*kamnetes*	stone carver	18	1871-1908
555	*karakulevod*	karakul breeder	3	1970
556	*kartofelevod*	potato breeder	17	1929
557	*kaučukonos*	rubber-bearing plant	31	1930
558	*kaševar*	cook (in military or working areas)	211	1815-1820
559	*kinoved*	film expert	82	1960
560	*kitaeved*	sinologist	42	1926
561	*kitoboj*	whale hunter	88	1855
562	*kitolov*	whaleman	23	1841
563	*kleevar*	worker preparing glue	2	1871-1908
564	*knigoved*	bibliologist	7	1902
565	*knigoed*	bookworm	10	1867
566	*knigoed*	pedant	-	1867
567	*knigoljub*	bibliophile	173	1912-1936

568	*knutoboj*	executioner	2	1952
569	*kožender*	knacker	4	1934-1939
570	*kožeed*	skin beetle	13	1768-1769
571	*kozovod*	goat breeder	3	1894-1899
572	*kozodoj*	nightjar (bird)	118	1870
573	*konevod*	horse breeder	29	1873
574	*konoval*	horse doctor	804	1682-1709
575	*konovod*	horse holder	357	1829
576	*konovjaz'*	rope to tie horses	268	1770-1811
577	*konogon*	horse driver	27	1914
578	*konokrad*	horse thief	454	1834
579	*konoplevod*	hemp grower	7	1931
580	*kontejnerovoz*	container vehicle	70	1971
581	*korneed*	longhorn beetle	1	1942
582	*koroed*	bark beetle	112	1909-1910
583	*kosmolet*	spaceship	39	1967
584	*kostolom*	illness causing bones to break	56	1825-1833
585	*kostolom*	person who breaks bones in a fight	-	1825-1833
586	*kostoprav*	chiropractor	72	1761
587	*kostorez*	bone carver	25	1936-1979
588	*kostrožog*	fire lighter	9	1988

589	*kotločist*	boiler cleaner	4	1920
590	*krabolov*	crabber	30	1931
591	*krabolov*	crabber	-	1931
592	*kraeved*	expert on a certain region	258	1909
593	*krovosos*	leech	140	1913-1932
594	*krovosos*	exploiter	-	1913-1932
595	*krolikovod*	rabbit breeder	12	1911
596	*krochovod*	hairsplitter	79	1831
597	*kružkovod*	study group leader	9	1928-1931
598	*krysolov*	rat catcher	82	1910
599	*krjučkotvor*	pettifogger	35	1788-1822
600	*kuklovod*	puppeteer	107	1937
601	*kukuruzovod*	corn grower	5	1958
602	*kurovod*	hen breeder	19	1899
603	*kuroed*	bribe-taker	56	1928-1940
604	*kurochvat*	bribe-taker	1	1898
605	*kurocap*	bribe-taker	34	1995-1996
606	*kurocap*	hen thief	-	1995-1996
607	*kuroščup*	woman seducer	12	1912
608	*lakovar*	worker preparing varnishes	1	1969
609	*legkover*	credulous person	6	1927-1944

610	*legkodum*	light-minded person	1	1958
611	*ledorez*	ice cutter	33	1873
612	*lermontoved*	expert on Lermontov	9	1980
613	*lesovod*	silviculturist	258	1876
614	*lesovoz*	timber ship/lorry	149	1930
615	*lesorub*	wood-cutter	401	1906
616	*lisogon*	fox hunting dog	3	1925
617	*listoed*	leaf beetle	10	1871-1908
618	*literaturoved*	literary critic	721	1920-1943
619	*lichodej*	malefactor	219	1769
620	*licemer*	hypocrite	597	17818
621	*lobotrjas*	loafer	210	1830-1832
622	*luboed*	peppergrass beetle	2	2001
623	*lugovod*	plant grower	1	2004
624	*lunochod*	Lunochod	123	1954-1979
625	*lykoder*	dealing with tree barking	4	1831
626	*l'novod*	flax grower	26	1926
627	*ljudoed*	cannibal	902	1780-1814
628	*malover*	sceptic	128	1816-1862
629	*maslodel*	oiler	14	1894
630	*materialoved*	expert on material science	19	1970

631	*mašinoved*	expert on machines	5	1956
632	*medovar*	mead maker	51	1766-1777
633	*medonos*	bee plant	43	1923
634	*medosos*	honeyeater (bird)	2	1899
635	*merzlotoved*	expert on geocryology	7	1951
636	*metalloved*	expert on physical metallurgy	14	1941-1945
637	*miroed*	exploiter	325	1829
638	*molodoženy*	recently married	633	1828
639	*molokosos*	greenhorn	415	1770
640	*mongoloved*	expert on mongolistics	5	1980
641	*morechod*	seaman	370	1830-1837
642	*muzeeved*	expert on museology	11	1928
643	*muzykoved*	expert on music	273	1937
644	*mukoed*	grain beetle	1	2010
645	*mukomol*	miller	104	1839
646	*mukosej*	worker sifting flour	6	1882
647	*murav'ed*	ant eater	70	1876
648	*musorovoz*	dustcart	105	1969
649	*muchomor*	amanita (poisoning mushroom)	409	1787
650	*mylovar*	worker preparing soap	30	1863
651	*myšelov*	mice hunter	6	1970

652	*mjasoed*	meat eater	358	1761
653	*narodoved*	expert on ethnography	1	2002
654	*neftevoz*	oiler/oil tanker	10	1969
655	*novosel*	new settler	547	1858-1889
656	*obščestvoved*	expert on social sciences	86	1920-1929
657	*ovoščevod*	vegetable grower	43	1940
658	*ovcevod*	sheepman	84	1841
659	*ognemet*	flame thrower	155	1919
660	*odnodum*	person always thinking about the same	15	1887-1894
661	*odnoljub*	one-woman man	89	1907
662	*olenevod*	deer breeder	243	1895-1896
663	*osoed*	European honey buzzard	6	1970-1996
664	*ochotoved*	expert on hunting economy	95	1956
665	*panelevoz*	panel transporter	6	1964
666	*parovoz*	steam locomotive	4595	1842-1843
667	*parochod*	steamboat	17667	1822
668	*peskožil*	lugworm	5	1926
669	*pešechod*	pedestrian	1815	1784
670	*pivovar*	brewer	314	1741
671	*planetochod*	rover	12	1975
672	*plodovod*	fruit grower	5	1910

673	*plotovod*	raftman	1	1959
674	*plotovod*	raft convoy	-	1959
675	*plotogon*	raftman	25	1919
676	*polevod*	field-crop grower	62	1923
677	*poloter*	floor polisher	204	1829
678	*posudomoj*	dishwasher	1	1961
679	*počvoved*	soil scientist	132	1892
680	*počerkoved*	expert on calligraphy	7	1933
681	*pravdoljub*	truth lover	48	1847
682	*pravoved*	jurist	425	1843
683	*preljubodej*	adulterer	157	1719
684	*prirodoved*	natural historian	18	1923
685	*pticevod*	poultry breeder	33	1888
686	*pticeed*	tarantula	9	1975
687	*pticelov*	bird catcher	174	1766-1794
688	*pustobolt*	windbag	6	1883
689	*pustobrech*	dog barking with no reason	85	1878
690	*pustobrech*	windbag	-	1878
691	*pustozvon*	windbag	61	1833
692	*pustopljas*	frivolous person	26	1869-1886
693	*puškinoved*	expert on Puškin	29	1912

694	*pčelovod*	beekeeper	271	1831-1832
695	*pylesos*	vacuum cleaner	925	1919-1922
696	*rakoed*	crab-eating racoon	3	1926
697	*rastenievod*	plant grower	19	1923-1925
698	*religioved*	expert on religious studies	29	1992
699	*repolov*	linnet (bird)	6	1921-1925
700	*risovod*	rice grower	15	1912-1936
701	*rifmoplet*	rhymer	32	1806-1809
702	*rotozej*	scatterbrain	226	1777-1778
703	*rudovoz*	ore ship	24	1956
704	*rudokop*	miner	296	1795
705	*rukosuj*	clumsy person	87	1869-1870
706	*rybovod*	fish breeder	23	1920-1928
707	*rybolov*	fisherman	1282	1717
708	*rybochod*	fish-pass	9	1951
709	*sadovod*	fruit grower	802	1856
710	*samovar*	samovar	6705	1770-1811
711	*samogudy*	samogudy	42	1832
712	*samodur*	despot	518	1736-1774
713	*samokat*	push scooter/bicycle	190	1854
714	*samolet*	airplane	27749	1787

715	*samolov*	trap	131	1763
716	*samoljub*	self-lover	8	1742
717	*samosad*	home-grown tobacco	135	1924
718	*samosval*	dump body truck	722	1949
719	*samosev*	self-sown plant	83	1867
720	*samosej*	self-sown plant	4	1908
721	*samostrel*	soldier deliberately wounding himself	197	1785
722	*samostrel*	ancient bow/bad gun	-	1785
723	*samochval*	boaster	277	1742
724	*samochod*	self-traction machine	82	1786
725	*sacharovar*	expert on sugar production	14	1863
726	*sveklovod*	beet grower	16	1952
727	*svinoboj*	pig slaughterer	5	1925-1926
728	*svinovod*	pig breeder	41	1909-1915
729	*svinopas*	swineherd	103	1794-1795
730	*svobodoljub*	freedom lover	4	1923
731	*sebjaljub*	self-lover	3	1985-1995
732	*sekundomer*	stopwatch	249	1889
733	*semenovod*	seed grower	3	1930
734	*serdceved*	god	49	1832
735	*serdceed*	lady-killer	96	1851

736	*skalolaz*	rock-climber	96	1933-1937
737	*skirdoprav*	packer	1	1970
738	*skorodum*	quick-thinking person	22	1822
739	*skorochvat*	clever person	12	1830
740	*skorochod*	fast runner	515	1752
741	*skotovod*	cattle-breeder	242	1773
742	*skotogon*	cattle driver	3	1884
743	*slavjanoved*	slavist	2	1934
744	*slavjanoljub*	slavophile	1	2002
745	*slastoljub*	pleasure lover	1	1786-1787
746	*smolovar*	tar extractor	2	1937
747	*smolokur*	tar extractor	344	1871-1874
748	*snegopach*	snow breaker	2	1910
749	*snegochod*	snow mobile	183	1970-1977
750	*sobakoved*	expert on cynology	7	1990
751	*sobakovod*	dog breeder	93	1907
752	*solevar*	salt extractor	12	1904
753	*spinogryz*	sprog	22	1988
754	*stalevar*	steel melter	282	1924-1925
755	*starodum*	old-fashioned person	269	1782
756	*starožil*	old resident	1244	1754-1758

757	*steklovar*	glass melter	13	1933
758	*stekloduv*	glass blower	129	1929
759	*steklorez*	glass cutter	16	1967
760	*steklorez*	glass cutter	-	1967
761	*stenolaz*	wallcreeper (bird)	4	1870
762	*stichoplet*	bad rhymer	42	1830
763	*suknoval*	cloth fuller	12	1871-1908
764	*suknodel*	cloth producer	3	1908
765	*suchovej*	hot wind	170	1862
766	*sčetovod*	accountant	461	1879-1881
767	*syrovar*	cheese maker	26	1874
768	*syrodel*	cheese maker	15	1968
769	*syroed*	raw vegan	6	1913
770	*tabakovod*	tobacco grower	14	1917
771	*tabakur*	one who smokes or snuffs tobacco	5	1912
772	*teatroved*	expert on theatre	109	1922
773	*teneljub*	mole-rat	1	1987
774	*teplovoz*	diesel locomotive	583	1925-1926
775	*teploljub*	heat lover	3	1936
776	*teploljub*	heat lover	-	1936
777	*teplomer*	calorimeter	5	1784

778	*teplochod*	motor ship	1221	1909
779	*testomes*	worker kneading dough	2	1991
780	*tigrolov*	tiger hunter	5	1934
781	*tichochod*	lazy-bones	63	1888
782	*tichochod*	sloth	-	1888
783	*tovaroved*	expert on merchandising	95	1923-1925
784	*tonkoprjad*	butterfly	1	1995-1996
785	*trubovoz*	truck used to transport pipes	14	1976
786	*trubočist*	chimney sweep	283	1777-1778
787	*trudoljub*	hard worker	68	1789
788	*tugodum*	slowcoach	65	1925-1928
789	*tjaželovoz*	carthorse	65	1905
790	*tjaželodum*	slow-witted person	14	1920
791	*tjažkodum*	slow-witted person	6	1862
792	*uglevoz*	ship for coal transport	17	1890
793	*ugležog*	charcoal burner	42	1896
794	*uglekop*	miner	90	1881
795	*chlebopek*	baker	132	1812
796	*chleborez*	bread cutter	59	1875-1883
797	*chleborob*	grain grower	326	1886
798	*chlopkovod*	cotton grower	21	1905

799	*chlopkorob*	cotton grower	53	1901
800	*chmelevod*	hop grower	1	1903
801	*cvetovod*	floriculturist	98	1855
802	*cvetoljub*	flower lover	4	1926
803	*cvetonos*	flower stalk	73	1940
804	*citrusovod*	citriculturist	1	1951
805	*čaevod*	tea grower	22	1966
806	*čarodej*	magician	736	1715
807	*častotomer*	frequency counter	7	1968
808	*čelovekoljub*	humanist	9	1846-1847
809	*červovod*	(silk)worm breeder	3	1932
810	*čertogon*	devil's-bit (plant)	15	1871-1874
811	*čertolom*	rude person	4	1862
812	*čertopoloch*	thistle	376	1787
813	*čestoljub*	ambitious man	2	1894
814	*čechoved*	expert on Čechov	10	1920-1938
815	*čistopljuj*	cleany person	93	1887-1894
816	*čudodej*	miracle maker	83	1833
817	*šapoval*	hat-maker	163	1845
818	*šekspiroved*	expert on Shakespeare	76	1938
819	*šelkovod*	silkworm breeder	17	1828-1838

820	*šelkoprjad*	silkworm	358	1871
821	*šerstobit*	wool beater	77	1847
822	*šerstoboj*	wool beater	5	1836
823	*škuroder*	exploiter	26	1928
824	*èkskursovod*	tour guide	568	1920-1943
825	*èlektrovoz*	electric locomotive	409	1910
826	*èlektrochod*	motor ship	41	1942-1943
827	*èfironos*	oil-bearing plant	9	1939
828	*jazykoved*	linguist	112	1849
829	*jajcevod*	oviduct	18	1970
830	*jajceed*	egg eater (insect)	3	1966
831	*japonoved*	nipponist	8	2007

Appendix 3

Base verbs of the suffixless construction

	Base verb	Meaning	Absolute Frequency
1	*(raz)vodit'*	guide, lead, drive (breed, raise)	52
2	*vedat'*	know, be expert on	40
3	*est'*	eat	22
4	*vozit'*	transport	18
5	*ljubit'*	love	18
6	*chodit'*	go, walk	17
7	*varit'*	boil, prepare	13
8	*lovit'*	catch	11
9	*merit'*	measure	10
10	*dumat'*	think	9
11	*bit'*	beat, kill	8
12	*rezat'*	cut	7
13	*drat'*	tear	6
14	*lazit'*	climb	6
15	*lomat'*	break	6
16	*sosat'*	suck, dredge	6
17	*chvatat'*	snatch	6
18	*gonjat'*	drive away	5
19	*delat'*	do, make	5
20	*nosit'*	bring, carry, wear	5
21	*glotat'*	swallow	4
22	*dejat'*	do (anc., poet.)	4
23	*dut'*	blow	4
24	*kurit'*	smoke, distill	4
25	*kopat'*	dig	3
26	*letat'*	fly	3
27	*lit'*	pour	3

28	*robit'*	work	3
29	*brechat'*	bark	2
30	*valjat'*	felt	2
31	*verit'*	believe	2
32	*gryzt'*	gnaw	2
33	*žeč'*	burn	2
34	*žit'*	live	2
35	*krast'*	steal	2
36	*metat'*	throw	2
37	*pasti*	tend	2
38	*plesti*	weave	2
39	*pljasat'*	dance	2
40	*pravit'*	govern	2
41	*prjast'*	spin	2
42	*rubit'*	cut	2
43	*sejat'sja*	sow	2
44	*streljat'*	shoot	2
45	*tesat'*	trim	2
46	*capat'*	snatch	2
47	*čistit'*	clean	2
48	*(s)čitat'*	read (count)	2
49	*boltat'*	chat	1
50	*borot'sja*	fight	1
51	*brat'*	take	1
52	*brit'*	shave	1
53	*valit'*	make fall	1
54	*vejat'*	blow	1
55	*vjazat'*	tie	1
56	*vpravljat'*	set	1
57	*vypuskat'*	let out	1
58	*gljadet'*	look	1
59	*gubit'*	kill	1

60	*gudet'*	buzz	1
61	*davit'*	knock down	1
62	*doit'*	milk	1
63	*durit'*	be obstinate	1
64	*ženit'sja*	marry	1
65	*zijat'*	gape	1
66	*zvonit'*	call	1
67	*katat'*	roll	1
68	*kolot'*	chop	1
69	*lizat'*	lick	1
70	*mazat'*	oil	1
71	*mesit'*	knead	1
72	*molit'sja*	pray	1
73	*molot'*	grind	1
74	*morit'*	exterminate	1
75	*mykat'*	live in misery	1
76	*myt'*	wash	1
77	*otvodit'*	reject	1
78	*pachat'*	till	1
79	*peč'*	bake	1
80	*plevat'*	spit	1
81	*pugat'*	scare	1
82	*sažat'sja*	get out from the land	1
83	*sbrosit'*	drop	1
84	*svalivat'*	dump	1
85	*selit'sja*	settle	1
86	*seč'*	cut	1
87	*sejat'*	sow	1
88	*sidet'*	sit	1
89	*skalit'*	show one’s teeth	1
90	*sovat'*	put in	1
91	*spuskat'*	let out	1

92	*stekat'*	flow	1
93	*tvorit'*	create	1
94	*teret'*	polish	1
95	*tolkat'*	push	1
96	*trjasti*	shake	1
97	*tjapat'*	hit	1
98	*upravljat'*	manage	1
99	*chotet'*	want	1
100	*chvalit'*	boast	1
101	*česat'*	comb	1
102	*šlepat'*	slap	1
103	*ščupat'*	touch, probe	1