

**REPRODUCIBLE MULTIWORD EXPRESSIONS
FROM A THEORETICAL AND EMPIRICAL
PERSPECTIVE**

VOLUME NINE

INTERCONTINENTAL DIALOGUE ON PHRASEOLOGY
University of Białystok, Poland

**REPRODUCIBLE MULTIWORD EXPRESSIONS
FROM A THEORETICAL AND EMPIRICAL
PERSPECTIVE**

VOLUME NINE



UNIVERSITY OF BIAŁYSTOK PUBLISHING HOUSE
BIAŁYSTOK 2020

IDP Series Editor:
Joanna Szerszunowicz

Editors:
Joanna Szerszunowicz
Martyna Awier

Reviewers:
Jasminka Delova-Siljanova
Mariangela Albano
Ai Inoue
Bożena Kochman-Haładaj
Vincenzo Lambertini
Christine Konecny
Alenka Vrbinc
Irina Zykova
Ewa Gorlewska

This publication has been peer reviewed.

© Copyright by University of Białystok, Białystok 2020

Typesetting: Stanisław Żukowski



The volume has been financed by the Philological Department
of the University of Białystok

ISBN 978-83-7431-689-7

University of Białystok Publishing House
15-328, Białystok, 20B Świerkowa Street
phone number: 857457120, e-mail: wydawnictwo@uwb.edu.pl,
<http://wydawnictwo.uwb.edu.pl>

Printed and bound by: Hot Art Przemysław Zaczek

LIST OF CONTENTS

<i>Joanna Szerszunowicz, Martyna Awier</i> Introduction	7
Part I. Multiaspectuality of reseach on phrasemes: kinds of units, their specificity and functions	
<i>María Auxiliadora Barrios Rodríguez</i> Lexical Functions and Pragmatic Functions: a Proposal for the Formalization of the Pragmatemes within the Meaning-Text Theory	15
<i>Tatsiana Maiko</i> What Can You Give in Italian that You Can't Give in Russian? A Contrastive Study of Constructions with the Light Verbs <i>dare</i> in Italian and <i>davat' / dat'</i> in Russian	33
<i>Antonio Pamies</i> Phrasal Verbs, Idiomaticity and the Fixedness Continuum	55
<i>Iris Vogel</i> Representing Gestural Idioms in the Lexicon	75
Part II. Dynamics of phraseological stock	
<i>Tetiana Anokhina</i> The Form Changes and Shifts of Meaning of the English MWUs ...	93
<i>Damien Villers</i> Phraseological Blunders: When New Phrasemes Are Born from Errors	109

Part III. Variability of fixed expressions

Chen Ting

Types and Degrees of Variation in English and Chinese Phraseological Units	129
---	-----

Tomáš Jelínek

Multi-word Lexical Units with Repetition of Lexemes in Czech and Identification of Their Variants	141
--	-----

Marie Kopřivová

Variability of Czech Verbal Phrasemes: Case Study of <i>dát</i> ('to give')	155
---	-----

Part IV. Reproducibility from a theoretical perspective

Wolfgang Eismann

Reproduzierbarkeit als unikales Kriterium zur Bestimmung von Phraseologismen. Zu den Verdiensten von Leonid Ivanovič Rojzenzon um die phraseologische Theorie	165
---	-----

Marina Gutovskaya

Phrasemes: Reasons for Reproducibility and Specificity of Sign Functions	177
---	-----

Part V. Identification and analyses of phrasemes across language varieties

Ewa Koziol-Chrzanowska

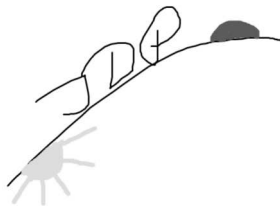
The Problems and (some) Solutions of Identifying Key Multi-word Expressions (MWEs). The Case Study of Polish Newspeak	189
--	-----

Katie Ní Loingsigh

Irish-language Idioms: an Overview and Analysis of Idioms in the Published Work of Peadar Ó Laoghaire	201
--	-----

Daniela Prutscher

Phraseme und Wortverbindungen in der deutschen Wissenschaftssprache	217
--	-----



Introduction

The recent decades have witnessed a growing interest in research on fixed expressions of various kind. Many modern approaches favour the broad concept of phraseology in which reproducibility of multiword units is viewed as a defining criterion. If this perspective is adopted, then phraseological stock is rich and varied, as it encompasses collocations, idioms, phrasemes, proverbs, sayings, winged words, formulae and many other types of units. Therefore, analyses of polylexical units are of great importance, since they focus on linguistically specific and – in many cases – culturally embedded language items commonly used in the proces of communication. Phraseological units appear across language varieties and are well represented in different forms of discourse, which is reflected in the present collected monograph that contains many studies are of theoreticall and empirical character.

The volume is composed of five parts which contain papers discussing various issues relevant both in terms of theoretical development of phraseology and empirical studies on fixed expressions. The first part offers four articles which attest the multiaspectual character of reseach on phrasemes – units which represent different types and perform various functions. It begins with María Auxiliadora Barrios Rodríguez's contribution titled *Lexical Functions and Pragmatic Functions: a proposal for the formalization of the pragmatemes within the Meaning-Text Theory*. The paper deals with pragmatically restricted phrasemes called pragmatemes. It offers a presentation of different proposals concerning the notion of pragmateme and proposes the definition of the term. The Author questions the adequacy of Lexical Functions in this respect, offering a new tool the Pragmatic Functions. It was inspired by Lexical Functions considered useful for the formalization of prag-

matemes. The areas which can benefit greatly as a result of such an approach are dictionaries used for Second Language Acquisition purposes and Natural Language Processing applications. From a theoretical perspective, the proposal may also serve for the creation of a new taxonomy of illocutionary verbs.

The paper titled *What Can You Give in Italian that You Can't Give in Russian? A Contrastive Study of Constructions with the Light Verbs "dare" in Italian and "davat'/dat'" in Russian* contributed by Tatsiana Maiko discusses light verb constructions in Russian and Italian. It should be emphasized that contrastive research on such language units in these two languages has been neglected. This study analyses constructions formed by the Italian light verb *dare* 'to give' and the Russian equivalent which is the light verb *davat'/dat'* 'to give', with a noun in the function of the direct object. The Author adopts Construction Grammar and the notion of a family of constructions with a view to determining relatedness of LVCs within each language. In the study, productive systematic metaphors that license extensions from the basic sense of the verbs are identified. The analysis allows for indicating convergences and divergences between the constructions of the two languages chosen for comparison.

In his paper titled *Phrasal Verbs, Idiomaticity and the Fixedness Continuum*, Antonio Pamies underscores the continuity of fixedness from the perspective of idiomaticity on the example of verbal construction. The Author claims that phrasal verbs share more properties with phrasemes than with syntactic combinations. Moreover, the scholar observes that their function resembles that of many Romance and Slavonic prefixed verbs. The study offers a discussion on units exhibiting different degrees of fixedness and idiomaticity that range from regular syntactic combinations to completely lexicalized verbs, with an intermediate zone in which several kinds of units are situated, including: light verb collocations, verbal idioms, separable phrasal verbs, inseparable phrasal verbs and compound verbs. The aim is to determine mechanisms on the example of the group of units selected for the analysis.

The onomasiological approach in phraseological research is adopted by Iris Vogel who focuses on idiomatic expressions which contain the names of body parts, described by the scholar as "the interface between nonverbal and verbal communication". Her paper *Representing Gestural Idioms in the Lexicon* presents the challenges of encoding gestural idioms in lexicography. The study, based on a thorough analysis of selected Japanese dictionaries, enabling the identification of patterns for comprehensive description of gestural idioms. The results are of importance not only for lexicography, but also for language learning.

The focal issue of the second part is the dynamic character of phraseological stock. Tetiana Anokhina's paper *The form changes and shifts of meaning of the English MWUs* proposes a particular approach to multiword units change, creation and rearrangement. The aim is to discuss form changes and shifts of meanings of multiword expressions on the example of American English units. The study is corpus-based: the Corpus of Contemporary American English (COCA) is used for excerption of research material. The analysed units are considered as secondary ones providing that their actual meaning is changed, modified or erased. The papers discusses semantic deviation viewed as shift and rearrangement demonstrating mimesis when the initial meaning of polylexical units can be modified to some extent or removed. Semantic shifts are either secondary or post-secondary. In the process of semantic deviation, a brand new conceptual unit appears as result of the deletion of the old meaning of a given figurative expression

The process of phraseogenesis is discussed by Damien Villers whose contribution *Phraseological blunders: when new phrasemes are born from errors* focuses of a particular mechanism. The paper presents a typology of phraseological "blunders", which comprise both errors and mistakes that accidentally modify the standard form, usage, or meaning of a phraseme. The Author describes their causes, such as language pathologies, poor linguistic skills, or mere absence of mind. The article gives an insight into the phenomenon of variation, as it is of importance for determining what is considered as an error. The vital criteria include: frequency ratio, communicational efficiency, and semantic coherence. Phraseological blunders as sources of new phrasemes are exemplified and accounted for with the memetic approach with a view to giving an insight into the "phraseme genesis" process.

The third part contains three papers which focus on the phenomenon of variance in phraseology. Chen Ting in her paper titled *Types and Degrees of Variation in English and Chinese Phraseological Units* presents different classifications of phraseological units in two languages chosen for the analysis: English and Chinese. The author discusses universal reasons for variation occurring in phraseological units and types such as lexical, constructional, grammatical and pragmatic kinds of variation. As English and Chinese phraseological stocks differ in the structure of unit, the Author states that typologically different languages exhibit specific tendencies in types of variation. It is observed that in general, phraseological variation is more typical of English than Chinese and grammatical variation is not characteristic of Chinese. The paper contains a proposal of a model of potential degrees and levels variation of word combinations, including fixed expressions.

Tomáš Jelínek's paper *Multi-word lexical units with repetition of lexemes in Czech and identification of their variants* is devoted to the phenomenon of variability of multi-word lexical units with repetition of lexemes such as *Bůh dal, Bůh vzal* 'the Lord has given, the Lord has taken away'. Based on a large corpus, four case studies are analysed to discuss the extent of variability of the expressions in question. Fixed units with repeated constituents are also discussed on the background of a base of multi-word expressions, with a special focus on their variability. The Author also undertakes to present the process of the automatic identification of multi-word expressions with repeated constituents in Czech texts.

Czech phraseology is also the object of studies in Marie Kopřivová's paper *Variability of Czech verbal phrasemes: Case study of dát ('to give')*. It is focused on the variability of Czech verbal idioms and its representation in a database of a multi-word expressions. The aim is to offer a detailed analysis of the verb *dát* 'to give', one of the most frequent verbal components of Czech verbal idioms. The study is based on SYN2015 which is a representative corpus of contemporary written Czech that contains fiction, non-fiction, and newspapers and magazines, allowing for an automatic annotation of multi-word units.

The fourth part, Reproducibility from a theoretical perspective, contains two papers which shed light on two issues of paramount importance. Wolfgang Eismann's paper titled *Reproduzierbarkeit als unikales Kriterium zur Bestimmung von Phraseologismen. Zu den Verdiensten von Leonid Ivanovič Rojzenzon um die phraseologische Theorie*. The Author observes that although some works on phraseology mention Leonid Ivanovič Rojzenzon (1920–1977), his merits in this field deserve their due appreciation. Rojzenzon was the first to view reproducibility (Russian term *vosproizvodimost*) as the defining criterion of phraseology. The aim of the paper is to show how Rojzenzon defines the term of reproducibility, how the scholar distinguishes reproducibility and quotability and what reasoning he uses to defend his conception of the dichotomic character of reproducibility against critical objections. Rojzenzon's contribution to phraseological studies is invaluable: in the '70s of the 20th century, thanks to him, Samara became a center of research, including that of Russian phraseology. Rojzenzon pioneered in phraseological research, for instance, he initiated studies on gestural phraseologisms.

In turn, Marina Gutovskaya's paper *Phrasemes: Reasons for Reproducibility and Specificity of Sign Functions* analyses mechanisms of phraseologization. The scholar undertakes to discuss reasons for reproducibility of nominative phrasemes on the material of the English and Russian phraseological units referring to speech and related phenomena. The material for the analysis was

excerpted from lexicographic works. The article examines two kinds of reproducible units, i.e. idiomatic phrasemes and non-idiomatic ones. The paper aims to determine the proportions of these two types of fixed expressions in the research corpus. The results of the study conducted by the Author enable presenting the functions of phrasemes.

The fifth part of the volume deals with identification and analyses of phrasemes across language varieties. Ewa Koziol-Chrzanowska's contribution titled *The problems and (some) solutions of identifying key multi-word expressions (MWEs). The case study of Polish Newspeak* has twofold aims: first, to determine problems concerning practical usage of methods of identifying key multiword expressions; second, to propose some solutions. The Author refers to works by Anna Wierzbicka and Raymond Williams: in the paper, the notion of key polylexical items is understood analogically to the key words as viewed by these scholars. The research material represents Polish Newspeak, i.e. the language of propaganda and its mass media in the totalitarian period in Poland. The Author attempts to discuss three difficult stages preceding the analysis proper which are as follows: preparation of an initial list of units supposed to be the key ones, the collection of searchable linguistic data and the choice of the criteria of selecting appropriate texts. The issues discussed in the paper are illustrated with well-chosen exemplification.

Katie Ní Loingsigh's contribution *The principal features and characteristics of Irish-language idioms* contains a presentation of the principal features and characteristics of Irish-language idioms (syntax and semantics), specifically idioms collected from the published work of Canon Peadar Ó Laoghaire (1838–1920), the primary Irish-language prose author of the late nineteenth and early twentieth century.

In a language, phraseological stocks differ across various kinds of discourse, registers and styles. *Phraseme und Wortverbindungen in der deutschen Wissenschaftssprache* contributed by Daniela Prutscher examines one of them, as her study discusses German scientific language from a phraseological perspective. One of the properties of this variety is special phraseology comprising specific phrases and exemplary formulations. The Author determines current collocations and text routines using a corpus of 500 scientific articles from the period from 2010 to 2017. The analysed texts come from three different the areas, mainly German Studies, Foreign Language Teaching and Medicine.

The papers included in this volume present a panoramic perspective on current research research in the field of broadly understood phraseology. The articles reflect the multispectuality of analyses of fixed expres-

sions, focusing on crucial issues which – although extensively studied – still remain to be further explored. Offering an insight into such problems as kinds of reproducible polylexical units, phraseological continuum, semantics and functions of multiword expressions makes the volume a relevant contribution to phraseological studies. The volume highlights the dynamic aspect of set phrases: changes, shifts and variability, at the same time showing the development of the reproducibility concept from a theoretical perspective.

Białystok, December 2020

Joanna Szerszunowicz
Martyna Awier

PART I

**MULTIASPECTUALITY OF RESEARCH
ON PHRASEMES: KINDS OF UNITS,
THEIR SPECIFICITY AND FUNCTIONS**

María Auxiliadora Barrios Rodríguez

Universidad Complutense de Madrid

Spain

<https://orcid.org/0000-0001-6830-4797>

Lexical Functions and Pragmatic Functions: a Proposal for the Formalization of the Pragmatemes within the Meaning-Text Theory¹

Abstract. Pragmatemes have been defined within the Meaning Text-Theory as phrasemes pragmatically restricted, such as *for rent*, *drive slow(ly)*, *do not enter*, *beware of the dog*. This concept is close to the previous concepts of *pragmatic formulae*, *pragmatic markers*, *speech formula*, *linguistic cliché* and *gambits*, which cover expressions such as *I regret that*. We claim that *pragmatemes* are related to these concepts and also to a broad range of other expressions, such as: a) speech acts characterized by the influence of the extra-linguistic features in their meaning, as *¡soy humano!* (*only human!*), expression that does not mean that someone is human but that it is understandable he has made something wrong; and b) speech acts characterized by cultural aspects, as in the Spanish question *¿quién es el último?* (who is the last person?) in a queue (waiting in a commerce, for instance), due to the Spanish habit of forming messy groups instead of long queues. So far within the Meaning-Text Theory pragmatemes have been formalized by Lexical Functions, a formal tool useful for the categorization of the lexical relations. However, pragmatemes express not a lexical relationship between words but a pragmatic relationship by means of words between participants in a communicative situation. Consequently, they demand some extra-linguistic features. In this paper we summarize the different proposals regarding the concept of pragmateme, we define it, we claim that Lexical Functions are not adequate for them and we present what we call the Pragmatic Functions, a new tool inspired in Lexical Functions, useful for the formalization of pragmatemes, particularly in regard to Second Language Acquisition dictionaries and Natural Language Processing applications. Pragmatic Functions could also serve at the same time as a new taxonomy of illocutionary verbs.

Key words: *pragmateme*, *pragmatic formulae*, *lexicology*, *lexicography*, *Meaning-Text Theory*

¹ This paper has been founded by the grant FFI2017-83293 of the Spanish Ministry of Science and Economy for the project "Diretes: Diccionario Reticular Español. Diccionario analógico y relacional con acceso en red desde el sentido y desde la forma". I would like to express my gratitude to the anonymous reviewers for their helpful observations, to Deborah Paton for all her suggestions on the English revision of the manuscript and to Joanna Szeszunowicz for her outstanding help in editing. Any remaining errors are the sole responsibility of the author.

1. Introduction

Pragmatemes have been defined within the Meaning Text-Theory (MTT) as phrasemes pragmatically restricted; that means expressions such as *for rent*, *drive slow(ly)*, *do not enter*, *beware of the dog* (Mel'čuk: 2015a: 29). The author claims: "the situation that the speaker wants to describe phraseologically binds the phrase A + B (...), the situation prescribes what to say and may be how to say it". The scholar admits that there may also be a broad concept of pragmateme: "a formuleme is a pragmateme if it is pragmatically constrained" (Mel'čuk: 2015a: 29).

The concept of pragmateme is relatively new and there is neither sufficient research nor agreement among scholars. Kauffer (2017) reviews the concept and Fréchon, Frassi and Polguère (2012) propose a broad pragmateme concept to include lexemes and locutions with pragmatic value. Barrios (2017) also claims for a spread use of the concept of pragmateme to create materials for Spanish Second Language Learners.

She considers that *hazme un favor* (*do me a favor*) is a collocation, whilst an expression such as *haz el favor de* (*do me a favor and ...*) is a pragmateme. The first Spanish expression is similar to the English collocation *to do a favor*, whilst the second Spanish expression can easily be rude and usually demands a higher authority of the speaker and a feeling of anger (such as the mother demanding her son or daughter to make the bed, when this is a daily petition without any success). Both (authority and anger) can be labelled as extralinguistic features and should be considered when describing the meaning of this expression in any dictionary (Barrios, 2017). But, how could such subjective features be described in an objective and formal way? Plus, on the other hand, could these types of productive structures also have been considered pragmatemes? What does it mean "pragmatically constrained"?

There is no consensus on the concept of pragmateme within the MTT but nonetheless, as we will summarize in the next section, the concept has been growing in significance during the last few years, even though outside the MTT framework. García Page (2007) claims that this term corresponds to the previous term *pragmatic formulas*, and that they are characterized because they are semantically, syntactically and phonically autonomous. Some other aspects have been studied by several MTT outsider scholars, such as pragmatemes as a signal of processing relevance and cognitive coherence (Komlósi, 2009), and as a signal of different psychotype of the speakers (Gorbunova, 2017). They have even been analysed in one language of South Africa by Nikuze (2014), who proposed a type of lexicographic representation of pragmatemes for general Kinyarwanda dictionaries.

The concept of pragmateme is close to a specific type of *marker* defined by Fraser (1996). He claims that there is a non-propositional part of sentence meaning that can be analysed into what he called *pragmatic markers*: markers that “taken to be separate and distinct from the propositional content of the sentence, are the linguistically encoded clues which signal the speaker’s potential communicative intentions” (Fraser: 1996, 326). The author recognized four types of Pragmatic Markers: basic marker, which is an expression of belief (such as *I regret that*); commentary marker, which comments on the basic message (such as *frankly*); parallel marker, which adds something optional (such as *in God’s name*); and discourse marker, which signals the relationship between the basic message and the discourse (such as, *incidentally*). As may be seen, pragmatic markers are attached to the discourse structure whiles pragmatemes are attached to the extra-linguistic context.

Fraser’s proposal is not so far from the *pragmatic formulas* (Smichdt, 1993) and from one particular type of pragmatic formula, called *gambit* (Keller, 1979). A gambit is a “formulaic expression whose primary role is strategic rather than propositional in nature” (such as *the main point is, or may I interrupt for a moment?*); “it makes it easier for the hearer to process the discourse by providing them with opportunities for top-down processing”; if there is an inappropriate use, the speaker can be viewed as impolite (Dufon, 1993: 27–28). As in the preceding proposal, we claim that the concept of pragmateme does not equate necessarily with the concepts of pragmatic formulae and gambits.

Finally, this concept is close to the concept of *speech formula* or *linguistic cliché* (Cowie, 2011). Cowie claims that a speech formula is not a *routine formulae* (such as *how do you do?*) and recognises two types of speech formula: independent clause type (such as *you know what I mean*) and dependent clause type (such as *if anything*). He claims that some of them contains the pronoun *you*. Some other recent works reflect the interest in all these kinds of expressions for learners of any language, such as Alessandro and Zamora (2011) paper on pragmatic speech acts from the Italian-Spanish contrastive perspective.

In this paper we reflect on the concept of pragmateme and its formalization in order to be included in dictionaries useful for Natural Language Processing (NLP) and Second Language Acquisition (SLA). We claim that pragmatemes can be formally described, and we present a new concept, the Pragmatic Functions, a formal tool that could assist towards this goal.

The paper is organized in seven sections. After this introduction, in section 2 we focus on the state of the art regarding the concept of pragmateme and in section 3 we present the one we are working with. Section 4 de-

finishes the concept of Lexical Function and summarizes the problems that arise when formalizing pragmatemes with Lexical Functions, which has hitherto been the usual practice to formalize them within the Meaning-Text Theory. Section 5 presents our proposal to formalize pragmatemes, the concept of Pragmatic Function, and section 6 shows a preliminary typology of pragmatemes by this formalization. Finally section 7 presents the conclusions.

2. The concept of pragmateme within the Meaning-Text Theory

The concept of pragmateme was defined for the first time within the Meaning Text-Theory as a set phrase composed of two lexemes A and B, such that it signified 'A + B' is not unrestrictedly constructed on the basis of the given Conceptual Representation out of the senses 'A' and 'B' (Mel'čuk, 1995: 179–189); the so called "conceptual representation" is a cognitive level attached to the extra-linguistic situation. In later works, the author claims that the pragmatemes are classified as a subclass of clichés (Mel'čuk, 2015b: 55)². The main point of his work is precisely related to the conceptual structure: whilst an idiom such as *kick the bucket* is used following a non-standard semantic rule (because it does not mean 'kick the bucket'), a pragmateme (we could think, for instance, on *wet paint*), demands a non-standard conceptual rule (Mel'čuk, 2015a: 60): in the following lines, we will try to explain what a conceptual rule is although very few works were performed on them within the Meaning-Text Theory. We should add that most MTT scholars have been working on the concept of pragmateme as defined by Mel'čuk (1995), which means that they do not work on a broad pragmateme concept, as proposed by Fréchon, Frassi and Polguère (2012) and Barrios (2017).

Regarding idioms, there is a non-standard semantic rule that orders "to block its free manipulation by syntactic and morphological rules"; then, when a speaker uses an idiom and says that someone *kicked the bucket*, he means that someone died and that he "is flipped" because of that

² The Mel'čuk hierarchy of phrasemes is presented step by a step in one of his later works: "A phrase E of L is a lexical phraseme if and only if it is not free, that is, if and only if it is constrained (...) A lexical phraseme is a semantic-lexemic phraseme if and only if its meaning and its lexemic implementation are both constrained with respect to its referent- that is, to its conceptual representation (...) A semantic-lexemic phraseme is a cliché if and only if it is compositional (...) A cliché is a fomuleme if and only if it has a specific abstract referent (...) A fomuleme is a pragmateme if and only if it is pragmatically constrained" (Mel'čuk, 2015b: 59; 74; 65; 69; 74; and 83 respectively).

(Mel'čuk, 2015a: 60). For Mel'čuk, this non-standard semantic rule explains the relationship between the meaning the speaker wants to express and the phrase he uses. We should highlight that *kick the bucket*, as with many idioms, includes the attitude of the speaker. As it is outside our specific field of research here, in this paper we will omit the role of the speaker's attitude of the idioms but we study its presence on the pragmatemes.

Returning to the issue of pragmatemes, as said before, besides a non-standard semantic rule, a pragmateme demands a non-standard conceptual rule (Mel'čuk, 2015a: 60). As it is not so easy to understand what is meant by non-standard conceptual rule within the MTT, we will try to approximate to its meaning by an example: when a speaker wants to express, for instance, the idea of *being careful* because there is fresh paint on something (such as handrails, fences or doors), at the semantic level he must choose the meaning 'wet' and 'paint' in English (*wet paint*), or 'touch' (*do not touch*), but neither the meaning 'be careful' nor 'fresh'; actually the expression *fresh paint* is used more frequently to refer to an entire room or house which was recently painted³. Consequently, we could say that there is a non-standard selection of the meaning ('wet paint') and its expression (*wet paint*) from the concept ('I want to warn you because there is something freshly painted and you could get dirty'). The non-standard conceptual rule can change from one language to another: in fact, for the same concept, at the semantic level a Spanish speaker will choose 'recently' and 'paint' to express the same idea (*recién pintado*, lit. *recently painted*) or will even add 'be careful' (*¡Cuidado, recién pintado!*, lit. *Be careful, recently painted!*)

To sum up, the complexity of the extra-linguistic situation attached to a pragmateme implies so many features that there are several meanings that could potentially be chosen to express any idea. Each language selects some of these meanings; consequently, we could claim that the same situation could be attached to different conceptual rules in different languages. Actually, the *wet paint* example shows different selections of meanings and words for each language, more than a rule linking concepts and words (consider, in any case, that the explanation for *wet paint* is ours, trying to illustrate the Mel'čuk's proposal).

Conceptual rules are being deeply analyzed by the Moscow' School in their ETAP4 applications (<http://cl.iitp.ru/etap4>). However, as far as we know, pragmatemes are not yet the subject of their research. As we have no experience working on conceptual rules, we will not work on them

³ We have confirmed our views on this subject with Google images (last revised January 15, 2019).

in this paper. However, we consider that the conceptual rules are attached to the extra-linguistic situation, and we will focus on this last point.

Few scholars have worked with the concept of pragmatemes within the MTT. Among them Blanco (2013, 2014) recognizes that pragmatemes are frozen statements conditioned not only semantically (as collocations and idioms) but also pragmatically, by the situation of communication. Blanco's research focused on the lexicographic study of the pragmatemes. He characterizes the pragmatemes by its *lexical anchorage*, a model for the pragmateme lexicographic processing previously proposed by Mel'čuk (2008). For instance, the lexical anchorage for the pragmateme *no parking* is *vehicle* (Blanco, 2014: 16). As Mel'čuk, Blanco proposes that the pragmatemes should not be stored on a dictionary as lemmas (2010); that means that *no parking* should not be a lemma but an addition of the lemma *vehicle*. This is, on the other hand, the usual way to work with the idioms, not only in the MTT dictionaries but also in the outsiders MTT general Spanish dictionaries.

As a result, the lexicographic entry of a dictionary for Blanco should be counted on some basic fields: lemma and equivalents in other languages (*prohibido aparcar, no parking*), variants (*it is forbidden to park*), lexical anchorage (*vehicle*), speech act (*order*), semantic structure (X says that Y cannot park there), synonyms and antonyms, and some other feature such as Lexical Functions (on the section 4 we present this concept) (2013, 2014).

Some other works have been published within the MTT. Most of them focus on small data and do not propose a new and clear definition of pragmateme (see, among others, Iriarte Sanroman (2000) for the Portuguese and Barrios (2007, 2008) for the routines); some other papers demand a revision of the concept (such as Barrios, 2017). Because of its relevance, we will comment on the content of a few more MTT proposals in the following section.

3. The broad concept of pragmateme

From our point of view, there are two MTT works that are particularly interesting for our subject: Mel'čuk (2015a) and Polguère (2016). The first one claims for the use of pragmateme in a narrow sense (as proposed by Mel'čuk, 1995). Both of them classify pragmatemes as a type of cliché, and clichés as a type of phraseological expressions. Polguère indicates that clichés (*have a nice day; sorry, I'm late*) are compositional but, on the other hand, they are idiomatic because "their content is prefabricated as much as its form" (2016: 13). He notes that they are always used as a speech act, but the speaker

does not build the speech act, he does actually mention it; in other words, the speaker uses the cliché as a quote (2016: 5). Polguère adds that there are several features attached to the concept of a cliché, such as variants related to the lexical and/or syntactic combinatory (*un instant / un petit instant; just a minute!, just a second!*); pragmatic features derived from the speaker, the listener and the medium (a medium could be a letter, an advertisement, etc.) and problems in their translation (2016: 5–6).

As shown in the above paragraph, Polguère studies clichés more than pragmatemes, and recognizes that any cliché is a Lexical Entity (2016: 5–6). Considering that Lexical Entity are mainly lexical units and *vocables* (Gader, Olliger et Polguère, 2014), we assume that it is a type of linguistic entity different from Lexical Units. Lexical Entity, as we understand it, is a higher category, and includes not only Lexical Units but also expressions, such as idioms, pragmatemes and so on.

In this regard, the traditional general dictionary's lemma is no longer necessarily a simple word or a compound noun. There is no problem of space, nor problem of alphabetic order in an e-dictionary. Moreover, we consider that *dictionary* denotes a smaller product than *e-dictionary*: we understand that the new e-dictionaries are big databases containing more than one type of dictionary (Barrios, 2019a; 2019b). This is the reason why we include the Lexical Entities as entries of our e-dictionary. In fact, we add each cliché and each pragmateme to our database as an individual entry of the dictionary, not as a part of some other lexical entry (such as Blanco proposes for the pragmatema *no parking*, which is a part of the lemma *vehicle*, see section 2). Furthermore, in our database we not only have two different entries (*no parking* and *vehicle*), but also an explicit relationship between both of them by means of the Pragmatic Functions, as we will see in the section 5.

We are working with a broad concept of pragmateme, which means that some of the examples of clichés from Polguère (2016) are labelled as pragmatemes in our database. We recognize that the concept of pragmateme, as well as the concept of cliché and formuleme, should be analysed for several scholars in order to get a consensus within the MTT approach, at least from the terminological perspective.

As this task exceeds the possibility of our individual paper, presently we will settle for partially adopting the typology proposed by Mel'čuk (2015a). He recognizes nick-names (*Eternal City*) and termemes (*State Department*) as a type of cliché. We claim that both can have a function as referential expressions, which means that they are used when we want to talk about an entity of the real world. Quite different are the other two types of Mel'čuk's clichés, never used to point at any entity but to evaluate something or to cre-

ate a social frame for the interaction between speakers: sentencemes (*better late than never*) and formulemes (*what time is it?*) He distinguishes the pragmatemes as a subtype of formuleme, characterized by being pragmatically constrained (*I'll pass the phone to...*) (Mel'čuk, 2015a: 83).

Comparing the last two examples (formulemes and pragmatemes), we could consider that only the second one requires a concrete extralinguistic situation: in this case, answering a phone call. This is apparently the essential feature for the distinction between formulemes and pragmatemes proposed by Mel'čuk (2015a). However, we would like to underline that, from his explanations, the limits between both of them still remains unclear. In fact, it is necessary to have a concrete extra-linguistic situation for most of the formulemes he proposes. For instance, the expression *what time is it?*, an example of formuleme, demands a situation in which there is one person without any clock, mobile, tablet, pc or any kind of device showing the time, and at least a second person with some of these devices.

Furthermore, as in English, there are two Spanish expressions related to this question, the first one is literally *what time is it?* (*¿qué hora es?*), but the second one is slightly different, *¿tiene hora?* (*do you have the time?*) There is a condition for the speaker to formulate both questions (he should not have any artefact to know the time) but there is only an extra condition for the first one: the listener should have an artefact and the speaker should know he has it before asking him; if not, the speaker should ask about the time by means of the second question. Then, the extra-linguistic situation imposes the selection of the adequate expression: it is quite impolite in Spanish to ask someone in the street *what time is it?* (we prefer *do you have the time?*) Consequently, the attitude of the speaker is also present: if he asks *what time is it?* to a stranger on the street, he will show a rude attitude and may provoke a not so positive answer. However, in daily life, as the second condition is fulfilled when we are with the family, colleagues or friends, we ask just simply *what time is it?*

In summary, the limits between formulemes and pragmatemes are not clear: Mel'čuk proposed that formulemes are clichés with abstract referents, and pragmatemes are formulemes restricted by extra-linguistic situation (2015b). From our point of view formulemes and pragmatemes share most of the features, but for the pragmatemes the extralinguistic situation is more specific and the dependency of the extralinguistic situation is stronger.

We define pragmateme as following:

A **pragmateme** is an **expression** (word, multiword, set of words or sentence) which has the following properties: a) it is **fixed** but it could have **variants** from the syntactic and lexicological point of view; b) it func-

tions as a **speech act**; c) it can be attached to a **pragmatic function** (such as to thank, to order, to greet, to congratulate, to evaluate, etc.); d) it involves the potential speaker and listener through **grammatical features** (such as person or time morphemes, deictics, and so on); e) it can be expressed by means of **oral or written text**; f) it could be attached to some particular **extra-linguistic situations** (such as an interaction in a restaurant, in the doctor's room, or an advertisement in a poster, direction for use, recipe or handmade signboard among others); g) it could express an **attitude** of the speaker and consequently may provoke an attitude from the listener.

Some of these characteristics have been proposed by several scholars, particularly by Polguère (2016), as shown before. However, we don't claim that compositionality is one of them, because as we will see it can be unnecessary for the concept of pragmateme we are working with. On the other hand, we have added three new factors: the grammatical features, the pragmatic function and the attitude of the speaker and listener. We will explain these features at the same time we present some examples in section 6. However, prior to that we need to present the concept of Lexical Function and explain why we think it is not adequate for the pragmatemes.

4. Lexical Functions and the problems that arise when formalizing pragmatemes

A Lexical function (LF) is a function that associates a given lexical expression *L* (such as *sound*), which is called the argument or keyword, with a set of lexical expressions, which are called values (such as *loud*, *strong*, *heavy*, *deafening*): a LF expresses a specific meaning associated (for this example, 'intense') (Mel'čuk, 1996). Then, he calls this function *Magn*, a Lexical Function that means 'intense' and is associated with several collocations as (1) and (2) shows, and (3) details:

- (1) $\text{Magn}(\text{sound}) = \text{heavy, loud, strong, deafening}$
- (2) $\text{Magn}(\text{rain}) = \text{heavy, intense, torrential}$
- (3)
 - a. 'intense' (specific meaning associated with *Magn*)
 - b. *sound, rain* = arguments or keywords of both sets of collocations
 - c. *loud, strong, heavy, deafening, intense, heavy, torrential* = values of *Magn*

LFs are usually a productive sense; for instance, *Magn* is useful when formalizing hundreds of collocations expressing the meaning 'intense'. Values

(adjectives in these examples) change from one keyword (*sound*) to another (*rain*). As (3) shown, there are only few values shared by different keywords (in (1–2) examples there is only one, *heavy*).

As their name indicates, LFs express a relationship between words. That is the reason why it is quite complicated to formalize any pragmatic relationship by LFs: as pragmatemes involve extralinguistic features, the LFs are inadequate for them. There is one way, used so far, to solve this problem: the called non-standard Lexical Functions, a type of LF created *ad hoc* for some specific and not so productive lexical relationships (Polguère, 2007).

In (4) we show the non-standard LF proposed by Mel'čuk (2008) for the lemma *peindre*_{1.1} (*to paint*); here, the non-standard LF are the paraphrases [*this object was*] *recently painted*, which simply points out the extra-linguistic situation:

(4) [Cet objet a été] peint récemment : Peinture fraîche [sur un signe, pour prévenir qu'on peut se tacher]

[This object was] recently painted: Fresh paint [on a sign, to avoid someone touching it]

As the example (4) proves, the complex set of extra-linguistic features is present by means of a set of words that, without any previous explicit structure or template, shows some characteristics of the circumstances (something was recently painted), the message (lit. *fresh paint*) and the medium for the message (on a sign).

From our point of view, it is hardly understandable that a Lexical Function (which is a general meaning) can be assimilated to a single and particular paraphrase explaining a concrete situation of life: there is no meaning in a situation but an interactive and complex set of linguistic and extra-linguistic features linking the speaker and the listener.

That is the reason why we do not use Lexical Functions for pragmatemes but a type of formalism that try to make explicit what we call the Pragmatic Function.

5. A new proposal: the concept of Pragmatic Function

We understand a Pragmatic Function (PF) as a function that expresses a speech act (such as to thank, to order, to greet, to congratulate, to evaluate, to warn, etc.) and associates a given extra-linguistic situation (such as an encounter or something freshly painted) which is called the argument, with a set of expressions (pragmatemes, formulemes) which are called values

(such as *how do you do, how are you going; wet paint, do not touch*); (5–6) shows the formalization of these examples:

- (5) To Greet (greeting encounter) = *how do you do?; how are you doing?*
 (6) To Warn (something freshly painted) = *wet paint; do not touch; fresh paint.*

A Pragmatic Function involves a set of complex features; consequently, most of the apparently equivalent pragmatemes are different from one another in some sense. There is a fine granularity that could be expressed adding new formalism to the proposals of (5–6), as (7–8) shown:

- (7) To Greet_formally (greeting encounter) = *how do you do?*
 To Greet_colloquially (greeting encounter) = *how are you going?*
 (8) To Warn (something freshly painted) = *wet paint; do not touch.*
 To Warn (a house/room freshly painted) = *fresh paint.*

As (7–8) proves, the granularity of the information can be spread either by means of the Pragmatic Function, adding some features, such as *formally* or *colloquially* as in (7); or by changing the argument (such as *something freshly painted*) for a more specific one, a sort of hyponym (such as *house/room freshly painted*) as in (8).

However, not all the extra-linguistic features that can be included in this situation could be expressed in this way. For instance, in section 1 we indicated that on the point of Fraser's proposal (1996) we were more interested in the expression of the speaker's potential communicative intentions, which corresponds to the feature 'attitude' in our definition of pragmateme (see section 4). As far as we know, there is no proposal for a frame for the description of all the features that are playing a role in the use of pragmatemes or formulemes. As in our e-dictionary the space is not an issue, we are solving this problem in our database by different fields on our table for formulemes.

Consider the expression *he did nothing*. It could be used as any other group of words in a compositional way, for instance in the sentence *this student should have finished his task but he did nothing*. In this case we only have to apply the grammatical rules in order to understand its meaning. However, if we think about the isolated expression *he did nothing!*, we probably discover it could be attached to different meanings depending on the extra-linguistic situation, as Table 1 shows.

As this set of situations shows, the extra-linguistic situation determines the meaning of this expression: it could be paraphrased as 'he really disappointed me, I had expected more from him', for the first situation; 'he is guilty' for second one; and 'he is innocent' for the last one.

Table 1. Possible different communicative intentions of a speaker when he says *he did nothing!*

Pragmatic Function	Extra-linguistic Situation	Typical Extra-linguistic Situation	Sentence/ Formuleme	Possible example
To vent (his/her frustration)	Conversation	A conversation with a friend about someone else	<i>He did nothing!</i>	Someone has a friend who could help him to solve a serious situation, but his friend did nothing
To accuse	Trial	Public prosecutor about the accused	<i>He did nothing!</i>	Someone is accused of neglect in a case of grave emergency
To defend	Trial	Lawyer about the accused	<i>He did nothing!</i>	Someone is falsely accused of committing a crime and the lawyer talk passionately to the jury about him

Source: own research.

We are interested in collecting any kind of expression that depends on the extra-linguistic situation, although it is not yet clear for us if it could be classified as a pragmateme, as a formulème or as a cliché. In section 6 we will show with more detail some features we analysed when building our preliminary typology of pragmatemes.

6. A preliminary typology of pragmatemes

In our data base, we use a table called “Formulemes” for all the expressions that are under the conditions delineated in section 3, which includes the pragmatemes. Table 2 shows some of our data (we write the English version of the Spanish expressions).

At the present time, we have collected more than eight hundred expressions corresponding to pragmatemes or formulèmes. On the one hand, we include in our corpus speech acts lightly characterized by the influence of the extra-linguistic features in their meaning, closer to the concept of formulèmes, as *¡soy humano!* (*only human!*), expression that does not mean that someone is human but that someone wants to apologise and that in this situation it is understandable that he has made something wrong. And, on the other hand, we also include speech acts with a stronger influence of the extra-linguistic situation, closer to the concept of pragmateme, some of them

Table 2. Some pragmatemes and formulemes of our database and their classification

Pragmatic Function	Extra-linguistic Situation	Typical Scenario	Pragmateme	Register
To greet	Greeting encounter		<i>How do you do?</i>	Formal
To greet	Leave-taking greeting		<i>Good bye!</i>	Formal
To greet	Greeting encounter		<i>How are you doing?</i>	Colloquial
To greet	Leave-taking greeting		<i>Bye, bye!</i>	Colloquial
To accept	After an order or indication		<i>All right!</i>	
To accept	After an invitation		<i>Nice!</i>	Colloquial
To warn	Something freshly painted	Handrails, fences, doors, etc. recently painted.	<i>Wet paint! Do not touch!</i>	
To order and to express disappoint	Someone made something wrong	Someone else angry and with authority gives him an order	<i>Do me a favour and...</i>	
To apologise for something	Someone made something wrong	Someone wants someone else to feel not so bad because of that	<i>Only human!</i>	
To ask an to get in line	Someone arrive to a queue	A queue in a mini market in Spain	<i>¿Quién es el ultimo? (Who is the last person?)</i>	

Source: own research.

characterized by cultural aspects, as the Spanish question *¿quién es el ultimo?* (who is the last person?), usual in a queue (waiting in a commerce, for instance), due to the Spanish habit of forming messy groups instead of long queues.

Our methodology was based on dictionaries, observation and introspection. We are building a preliminary typology of formulemes and pragmatemes at the same time that we are adding the data to our database. At this stage we have classified almost one hundred of them.

Lack of space does not allow us to show here some other fields of our table for formulemes (which includes pragmatemes), but we have some more columns: among others the lemma, which corresponds to the lexical anchorage (such as *greeting* for the four first files); the attitude (where we can add both, the speaker's attitude and the listener's attitude it could provoke); and the Second Language Acquisition level (the recommended level for any

student to learn this pragmateme). Not all the boxes have to be filled, just the boxes necessary for each expression.

The main point, as table 2 shows, is that the Pragmatic Functions are expressed in a natural language in the fifth column of our table, which means that the Pragmatic Functions proposed in (7) as *To Greet_formally*, is expressed here in a more simple way, *To greet + formally*; and similarly with the PF *To Greet_colloquially*, expressed as *To greet + colloquially* (see files one to four). From our point of view, the first formalism we proposed in 5–8 (which remains to the formalism of Lexical Functions) only makes sense if it is beneficial for a NLP application; if not, the second way (the one on the table) is easier for humans. In any case, it is just a formal question.

Some problems that we hope to solve in the next few months, still remain. The most relevant for us is related to the set of list of all the Pragmatic Functions we want to propose; we would like to see if our data is consistent with the Searle distinction: “we must carefully distinguish a taxonomy of illocutionary acts from one of illocutionary verbs” (1975: 368). Until now, all Pragmatic Functions are expressed by means of verbs that could be labelled as ‘illocutionary verbs’, but not all of them correspond to illocutionary acts.

7. Conclusions

The revision of the state of the art related to *pragmatic formulae*, *pragmatic markers*, *speech formula*, *linguistic cliché* and *gambits*, proves that all these concepts can be included by the concept of *formulème*, and some of them to the concept of *pragmateme* in its spread sense (Mel’čuk 2015a). However, there is no consensus on the concept of *pragmateme* proposed by Mel’čuk (1995) and, what is more important, the concept has not yet been applied to e-dictionaries accessible to anyone, which means that we cannot analyse big data related to *pragmatemes* in any existent dictionary. As we are working on a new Spanish e-dictionary, we are using our own data to arrive at some conclusions. There are different approaches to the concept of *pragmateme* within the Meaning-Text Theory, particularly by Blanco and Polguère proposals, but we understand for *pragmateme* something with some different features.

We have defended that non-standard Lexical Functions are not adequate to formalize the *pragmatemes*, even if so far it has been the normal way within the Meaning-Text Theory. Instead of that, we have proposed what we call a Pragmatic Function.

As an e-dictionary does not present problems of lack of space, we are collecting the pragmetemes and formulemes not only by Pragmatic Function but also reflecting some other words (lemmas) related to them, which corresponds to the lexical anchorage; and the Second Language Acquisition level demanded for its learning.

As we have not finished our project, we have not yet worked with the Pragmatic Function related to declarations, nor with pragmatemes such as *you're fired* or *I resign*. So we need to finish our task before submitting a definitive proposal for the taxonomy of pragmatemes and formulemes, and before proposing that our list of Pragmatic Functions could be definitively understood as a corpus of illocutionary verbs. We may say that we cannot know if our analysis of the complete set of data will confirm the conclusions we present here, not only in regard to the concept of pragmateme but also to the potential use of Pragmatic Functions as a set of Illocutionary Verbs. We hope to finish our project in one more year.

References

- Alessandro, Arianna; Zamora, Pablo. 2011. Un análisis lingüístico y traductológico de algunos enunciados pragmáticos en una perspectiva contrastiva español-italiano. In: Pamies, A.; Luque Nadal, L.; Pazos Bretana, J. M. (eds.) *Multilingual Phraseography: Second Language Learning and Translation Applications. Phraseologie und Paremiologie*. Band 28. Breten: Schneider Verlag Hohengehren. 139–146.
- Barrios Rodríguez, María Auxiliadora (2007): A tool for learning routines. In Christina Gitsaky (ed.). *Language and Languages. Global and local tensions*. Cambridge Scholars Publishing: Newcastle/United Kingdom. 90–113.
- Barrios Rodríguez, María Auxiliadora (2008): Propuesta de descomposición semántica de fórmulas rutinarias del español en el marco de la Teoría Sentido-Texto. In Mellado Blanco, C. (ed.) *Colocaciones y fraseología en los diccionarios*. Peter Lang Internationaler Verlag der Wissenschaften: Frankfurt. 211–231.
- Barrios Rodríguez, María Auxiliadora. 2017. Hacia un concepto amplio de pragmatema y sus aplicaciones en ELE: el caso de ¡qué + sust./adj.! In Almeida Cabrejas, B., Canales, A.B., García Sánchez, J.J. and Jiménez López, M.D. (eds.) *Investigaciones actuales en Lingüística (II). Semántica, Morfología y Lexicología*. Alcalá de Henares: Servicio de Publicaciones de la Universidad de Alcalá. 19–35.
- Barrios Rodríguez, María Auxiliadora. 2019a. Del concepto de planta de un diccionario al de base de datos léxica: cambios asociados a la lexicografía digital. En Calvo Rigual, Cesáreo y Ferran Robles i Sabater (eds.) *La investigación en lexicografía hoy: diccionarios bilingües, lingüística y uso del diccionario. Anejos. Quaderns de filologia: estudis lingüístics*. Valencia: Universitat de Valencia, vol 2, 605–620.

- Barrios Rodríguez, María Auxiliadora. 2019b. ¿Aún queda alguien para quien no exista un diccionario? *Diretes*, un diccionario electrónico apto para máquinas. En Cazorla Vivas, C., García Aranda, M. Á., Nuño Álvarez, M. P. (eds): *Lo que hablan las palabras. Estudios de Lexicología y Lexicografía en honor de Manuel Alvar Ezquerro*. Lugo: Axac.
- Blanco, Xavier. 2013. Les pragmatèmes : définition, typologie et traitement lexicographique. *Verbum*. 4 : 17–25. doi: 10.15388/Verb.2013.4.4977.
- Blanco, Xavier. 2014. Microestructura lexicográfica para unidades frásticas: los pragmatemas. *Revista Káñina*. XXXVIII: 13–18.
- Blum-Kulka, Shoshana (eds.) *Interlanguage Pragmatics*. Oxford: Oxford University Press. 21–42.
- Cowie, Anthony Paul. 2011. Speech formulae in English: problems of analysis and dictionary treatment. In: Geart van der Meer et Alice G.B. ter Meulen (eds.), *Making sense: from lexeme to discourse, in honor of Werner Abraham at the occasion of his retirement*. Groningen: University of Groningen, GAGL 44. 1–12.
- Dufon, Margaret A. 1993. The acquisition of gambits by classroom foreign language learners of Indonesian. In: Alves, Mark (ed.) *Papers from the Third Annual Meeting of the Southeast Asian Linguistics Society*. Tempe: Arizona State University. 27–42. <http://www.dssg.uchicago.edu/8c7i/02-emmie-lind/9781881044123-papers-from-the-third-annual-meeting-of-the-sout.pdf>. Retrieved: 01.01.2019.
- Fraser, Bruce. 1996. Pragmatic Markers. *Pragmatics*, 6(1): 167–190.
- Fréchon, Frassi; Polguère, Alain. 2012. Les pragmatèmes ont-ils un charme indéfinissable ? In Pierluigi Ligas et Paolo Frassi (ed.) *Lexiques. Identites. Cultures*. QuiEdit. pp. 81–104.
- Gader, Nabil ; Ollinger, Sandrine; Polguère, Alain. 2014. One Lexicon, Two Structures: So What Gives? *Seventh Global Wordnet Conference (GWC2014)*, Jan 2014, Tartu, Estonia. 163–171. [ffhal-00937187](https://hal.archives-ouvertes.fr/hal-00937187/document). <https://hal.archives-ouvertes.fr/hal-00937187/document>. Retrieved: 01.01.2019.
- García Page, Mario. 2007. Los pragmatemas: algunas consideraciones. In: Cuartero Otal, J., Emsel, M. (eds.), *Festschrift für Gerd Wotjak zum 65. Geburtstag. Band 1: Vernetzungen Bedeutung in Wort, Satz und Text*. Frankfurt: Peter Lang. 161–173.
- Gorbunova, Darya Alexandrovna. 2017. Spontaneous oral speech from the psycholinguistic side: pragmatemes in the speech informants with different psychotype. *Communication Studies*. 4(14). 73–82.
- Iriarte Sanromán, Álvaro. 2000. *A unidade lexicográfica: palavras, colocações, frasesmas, pragmatemas*. Braga: Universidade do Minho. <https://repositorium.sdum.uminho.pt/handle/1822/4573>. Retrieved: 01.01.2019.
- Kauffer, Maurice. 2017. De la notion de pragmatème en phraséologie. In : Cosimo De Giovanni (ed.) *Fraseologia e Paremiologia Passato, presente, futuro*. Milan: Francoangeli. 70–81.
- Keller, Eric. 1979. Gambits: Conversational Strategy Signals. *Journal of Pragmatics*- 3 (3–4): 219–238.

- Komlósi, László I. 2009. The choice of the Context Yielding the Joys of Cortex : Relevance and Pragmatic Enrichment for the Lexicon. *Topics in Linguistics. Interface Between Pragmatics and Other Linguistic Disciplines*, 3. Bratislava : Universita Konstantína Filozofa v Nitr. 39–47.
- Mel'čuk, Igor. 1995. Phrasemes in language and phraseology in linguistics. In: Everaert, Martin *et al.* (eds.). *Idioms: structural and psychological perspectives*. Nueva Jersey: Erlbaum Associates. 167–232.
- Mel'čuk, Igor. 1996. Lexical functions and lexical inheritance for emotion lexemes in German. In Wanner, L. (ed.), *Lexical functions in lexicography and natural language processing*. Amsterdam/Philadelphia. John Benjamin. 209–278.
- Mel'čuk, Igor. 2008. Phraséologie dans la langue et dans le dictionnaire. *Repères & Applications*, VI. Barcelona: Universidad Autónoma de Barcelona. 187–200.
- Mel'čuk, Igor. 2015a. Clichés, and Understudied Subclass of Phrasemes. *Yearbook of Phraseology*. 6(1): 55–86.
- Mel'čuk, Igor. 2015b. *Semantics. From meaning to text III*. Amsterdam/Philadelphia, John Benjamins.
- Nikuze, Emmanuel. 2014. *Towards a theoretically motivated model for the lexicographic representation of Kinyarwanda phrasemes in general Kinyarwanda dictionaries*. Stellenbosh: Stellenbosh University.
- Polguère, Alain. 2007. Lexical function standardness., In: Wanner L. (ed.) *Selected Topics in Meaning Text Theory, In Honour of Igor Mel'čuk*. Amsterdam: John Benjamins. 43–92.
- Polguère, Alain. 2016. Il y a un traître par minou : le statut lexical des clichés linguistiques, Corela [En ligne], HS-19 / 2016, mis en ligne le 08 juin 2016, consulté le 01 sept.-2019. URL: <http://corela.revues.org/4486>; DOI: 10.4000/corela.4486.
- Smichdt, Richard. 1993. Consciousness, Learning and Interlanguage Pragmatics. In: Kasper, Gabriele and

Lexical Functions and Pragmatic Functions: a proposal for the formalization of the pragmatemes within the Meaning-Text Theory

Resumen

En este artículo presentamos las conclusiones a las que hemos podido llegar tras analizar más de cien pragmatemas de los ochocientos que hemos recogido en nuestro diccionario. Hemos revisado el concepto de pragmatema, nacido en el marco de la Teoría Sentido-Texto (TST), y hemos visto que incluye, aunque es más extenso, conceptos previos como los de *pragmatic formulae*, *pragmatic markers*, *speech formula*, *linguistic cliché* y *gambits*.

Tras un repaso a los conceptos de pragmatema con los que se ha trabajado en la TST, hemos propuesto qué entendemos nosotros por pragmatema: un acto de habla asociado a una situación extra-lingüística particular, que está fosilizado de algún

modo aunque presente variantes, que se vincula a rasgos gramaticales (como los morfemas de persona o tiempo, o los déicticos), que se puede expresar oralmente o por escrito, que expresa una actitud en el hablante y puede provocar una actitud en el oyente.

Hasta el momento, en la TST se ha utilizado una herramienta llamada Función Léxica no Estándar para expresar de un modo formal el significado de los pragmatemas. A nuestro juicio es un error: una función léxica, como su propio nombre indica es léxica; para poder poner en relación una expresión con la situación extralingüística que le corresponda se necesita un recurso que no solo relaciones significados con unidades léxicas.

Hemos propuesto la existencia de lo que llamamos Función Pragmática (FP): una función que pone en relación un pragmatema con la situación adecuada de uso. Dado que un diccionario electrónico no tiene problemas de espacio, la FP permite recoger pequeños datos extralingüísticos, la actitud del hablante u oyente, el grado de formalidad del acto comunicativo, una entrada del diccionario a la que se pudiera asociar, etc. Hasta el momento las FPs que hemos formulado han podido ser expresadas como verbos ilocutivos.

Tatsiana Maiko

Università degli Studi di Milano

Italy

<https://orcid.org/0000-0002-1740-1615>

What Can You Give in Italian that You Can't Give in Russian?

A Contrastive Study of Constructions with the Light Verbs *dare* in Italian and *davat'/dat'* in Russian

Abstract. Light verb constructions (LVCs), which consist of a semantically reduced verb together with a noun (as the direct object or embedded in a prepositional phrase) conveying core lexical meaning to the combination, are widely used in both Russian and Italian. Contrastive research on LVCs in the two languages is practically non-existent. This study focuses on constructions formed by the Italian light verb *dare* 'to give' and the Russian light verb *davat'/dat'* 'to give', with a noun in the function of the direct object. The Construction Grammar model (Goldberg 1995, 2006) and the notion of a family of constructions were adopted to account for relatedness of LVCs within each language. Productive systematic metaphors that license extensions from the basic sense of the verbs were identified and convergences and divergences between the constructions of the two languages were established.

Key words: *light verb constructions, contrastive study, Construction grammar, Russian, Italian*

1. Introduction

Combinations of a semantically reduced verb together with a noun that conveys core lexical meaning to the combination, either in the direct object position (1, 2) or embedded in a prepositional phrase (3, 4), constitute a conspicuous class of idiosyncratic, yet semi-compositional and semi-productive constructions both in Italian and in Russian:

- (1) Italian: *prendere una decisione* 'to take a decision', *fare un'intervista* 'to do an interview';

- (2) Russian: *stavit' vopros* 'to put a question', *prinimat' učastie* 'to take part';
- (3) Italian: *mettere a confronto* 'to put in contrast', *prendere in considerazione* 'to take into consideration'.
- (4) Russian: *brat' pod ochranu* 'to put under protection' (lit. to take under protection), *stavit' pod somnenie* 'to put into doubt' (lit. to put under doubt).

Combinations such as these have been labelled *light verb constructions* (Jespersen 1946, Stein 1991), *delexical verb constructions* (Sinclair, Renouf 1988), *stretched verb constructions* (Allerton 2002) or *support verb constructions* (Mel'čuk 1996, Langer 2005), among other terms used. The criteria defining a light verb construction (LVC) vary depending on the chosen theoretical framework: for the purposes of this study a broad definition was adopted, embracing all combinations of a light verb and an eventive noun, including those which do not have a synonymous verb related to the noun.

Some scholars claim that light verbs have lost their semantic content and their role is limited to providing grammatical information to the predicate (Jespersen 1942: 117, Iordanskaja, Mel'čuk 2007: 239, Nordlund 2007: 84). Other studies based on material from several languages (Wierzbicka 1982, Alba-Salas 2002, Apresjan 2008, 2009; Samvelian et al. 2014) have demonstrated that light verbs display semantic preferences and their choice is not completely arbitrary, but is motivated by the meaning of the verb. Semantic classes of nouns that combine with a set verb can be identified, even though these generalizations are "probabilistic rather than rule-based" (Kopotev et al. 2016: 137).

Yet the choice of light verbs often differs cross-linguistically, and different light verbs may be used to denote the same situation in Italian and Russian:

- (5) *fare una diagnosi* 'to make a diagnosis' vs *stavit' diagnoz* 'to make a diagnosis' (lit. to put a diagnosis);
- (6) *fare un'intervista* 'to do an interview' vs *brat' interv'ju* 'to do an interview' (lit. to take an interview).

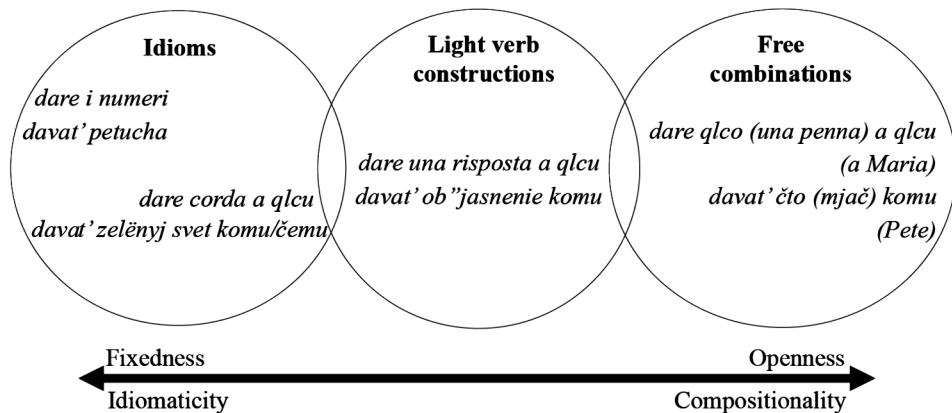
This is one of the reasons why acquisition of LVCs poses a real challenge for foreign language learners. There is a tendency for the interference of the learners' first language to cause errors in the production of these units (Altenberg, Granger 2001, Nesselhauf 2005, Gilquin 2007, Juknevičienė 2008).

Traditional foreign language teaching dealt with LVCs very randomly. Their acquisition relied mainly on learning various lists of units by heart without taking into account their frequency of use. Monolingual and bilingual Italian – Russian dictionaries also often treat LVCs unsystematically.

A contrastive study of LVCs in these two languages that detects differences and similarities and allows for some generalizations may be useful in the lexicographical and in the didactic fields. However, to date, only one contrastive study of Italian and Russian LVCs (Benigni, Cotta Ramusino 2011) has been conducted – on the Italian light verb *fare* ‘to do / to make’ and the Russian light verb *delat’* ‘to do / to make’.

This paper sets out to investigate LVCs with one of the most productive Italian light verbs *dare* ‘to give’ and its Russian counterpart *davat’/dat’* ‘to give’. These lie on a continuum between free and fixed instantiations of verbo-nominal combinations.

Figure 1. Continuum of verbo-nominal combinations with the Italian verb *dare* and the Russian verb *davat’/dat’*



Source: own research.

As depicted in Figure 1, this array can be represented as a gradient cline of idiomaticity (see Howarth (1998: 27), Nesselhauf (2005: 27–33)) with free combinations (7, 8) on the one side; and on the opposite idioms that more or less fixed and more or less semantically opaque¹, like lexically specified idioms (9, 10) and idioms with lexically open slots (11, 12):

- (7) *dare qlco (una penna) a qlcu (a Maria)* ‘to give sth. (a pen) to sb. (to Maria)’;
 (8) *davat’ čto (mjač) komu (Pete)* ‘to give sth. (a ball) to sb. (to Petja)’;
 (9) *dare i numeri* ‘to lose one’s marbles’ (lit. to give numbers);
 (10) *davat’ petucha* ‘to emit piercing sounds’ (lit. to give a rooster);

¹ On the motivation of idioms see Casadei (1996: 391).

(11) *dare corda a qlcu* ‘to listen to someone’ (lit. to give sb. a rope);

(12) *davat’ zelěnyj svet komu/čemu* ‘to give sb./sth. the green light’.

In the middle LVCs, where the meaning of individual words is retained to a certain extent, but substitutability of the components is restricted, are found:

(13) *dare una risposta a qlcu* ‘to give sb. an answer’;

(14) *davat’ ob’jasnenie komu* ‘to give sb. an explanation’.

Of course, these are not neatly separated classes and the boundary between them is often difficult to determine (Langer 2005: 188).

This paper is organised as follows: I begin with the corpora and the methodology employed (§ 2). The second section describes the primary meaning of the verbs (§ 3). Then several systematic metaphors that license extensions from the basic sense of the verbs are discussed (§ 4) and five groups of LVCs based on these metaphors in both languages are presented (§§ 4.1–4.5). Finally, the conclusions of this study are briefly discussed (§ 5).

2. Data and methodology

LVCs were extracted from the Italian Web 2016 (4,989,729,171 tokens) and Russian Web 2011 (14,553,856,113 tokens) corpora, accessed from the Sketch Engine corpus management system. The retrieval of combinations with the verb *davat’* followed by the accusative complement for Russian and with the verb *dare* and a direct object for Italian was carried out by using the Word Sketch search function. This identified 3,492,133 tokens sorted into 1559 types in Italian; and 2,946,415 tokens sorted into 1328 types in Russian. Manual processing of the first 550 types in both languages resulted in the elimination of a substantial amount of noise and free combinations, to obtain 352 LVCs in Italian and 308 LVCs in Russian.

Following the Construction Grammar approach, this analysis is based on the postulate that there is a continuum between lexicon and syntax and that highly idiomatic, idiosyncratic, unproductive patterns are not neatly separated from regular, productive ones (Fillmore, Kay 1996, Goldberg, Jackendoff 2004). All linguistic items are considered a construction “as long as some aspect of [their] form or function is not strictly predictable from [their] component parts”, and “even if they are fully predictable as long as they occur with sufficient frequency” (Goldberg 2006: 5).

Constructions that share some syntactic and semantic properties form a family. Establishing families of constructions helps “to capture generalizations across instances” (Goldberg 1995: 140) and “avoids assigning multiple and idiosyncratic senses for certain verbs” (Family 2014: 19). In a family of constructions there is a central construction that motivates other constructions through inheritance links. Postulating “inheritance allows us to capture the fact that two constructions may be in some ways the same and in other ways distinct” (Goldberg 1995: 72).

LVCs can be represented as usage-based conventionalized pairings which include a light verb and a nominal slot filled by a semantically restricted set of nouns, and which have a meaning that is entrenched in the language. They can be viewed as instances of an extension of the central sense of an argument structure in which they occur.

The Construction Grammar model and the notion of family of constructions were adopted in this study to account for relatedness between LVCs and prototypical non-fixed ditransitive structures with the verbs *dare* and *davat'* (for previous studies addressing LVCs as a family, or network, of constructions see Palancar (2003), Family (2009, 2014), Quochi (2016)). Productive systematic metaphors which motivate light verb uses were identified and similarities and differences between the constructions of the two languages were established.

3. Primary meaning of the verbs

When the verbs in question are used in their primary meaning, they denote a dynamic telic punctual action that brings about the transfer of a physical object from a volitional agent to a willing recipient (Goldberg 1995: 38), typically using the hands and, together with the object, they form the following pattern:

Form:

[X_{subj} *davat'* / *dare* Z_{obj1} Y_{obj2}]

Sagent: [+human]

Opatient: [+concrete]

Orecipient: [+animate]²

Meaning:

'X CAUSES Y TO RECEIVE Z'

² In Italian the theme is codified as the direct object and the recipient – as the indirect object introduced by the preposition *a*. In Russian the theme bears accusative case marking and the recipient argument bears dative case.

The most prototypical instances imply actual successful transfer (Goldberg 1995: 32) which involves a change in possession of the object together with a change of its location:

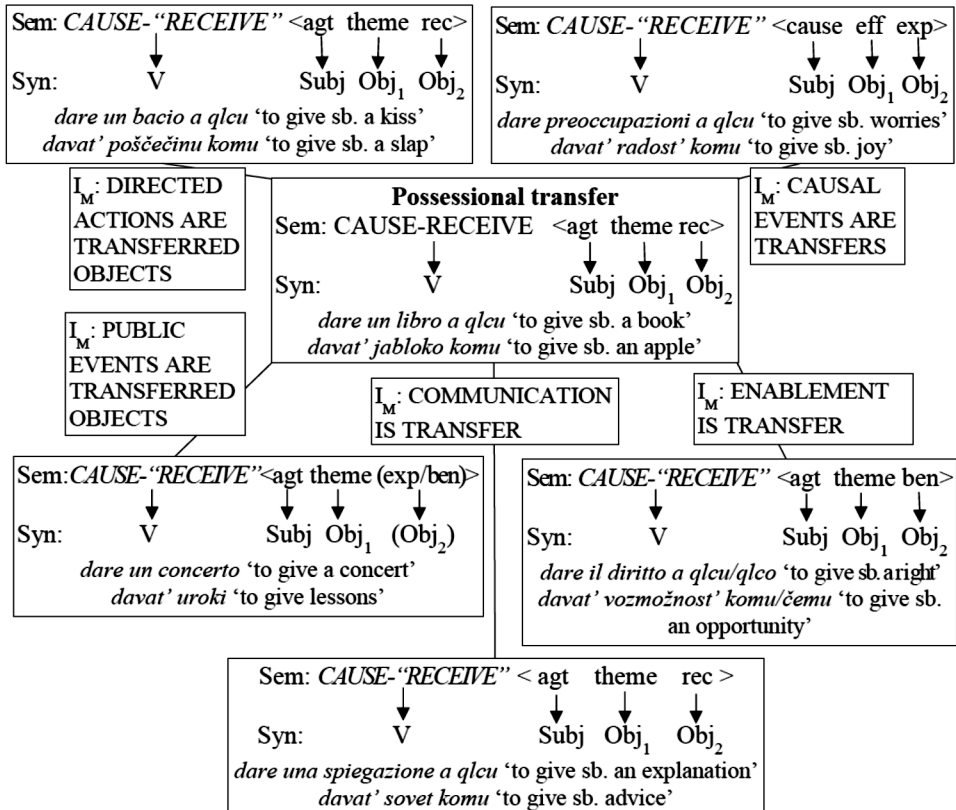
(15) *Lei ha dato un libro all'insegnante* 'She gave a book to the teacher'.

(16) *Mal'čik dal jabloko mame* 'The boy gave an apple to Mom'.

4. Systematic metaphors

Partially compositional ditransitive structures with the light verbs *dare* and *davat'* can be represented as extensions that inherit from the basic sense of the verbs via a metaphorical link. After retrieval, the relevant LVCs were classified according to different metaphors that seem to motivate them.

Figure 2. Naïve representation of the family of *dare*-LVCs and *davat'*-LVCs



Several systematic metaphors were identified in both languages, including: CAUSAL EVENTS ARE TRANSFERS, COMMUNICATION IS TRANSFER, ENABLEMENT IS TRANSFER, DIRECTED ACTIONS ARE TRANSFERRED OBJECTS, and PUBLIC EVENTS ARE TRANSFERRED OBJECTS. This does not imply that any noun that can be attributed to the semantic class, for example, of public events, is used with the verbs *dare* and *davat'*, but only that among the whole variety of possible collocates combined with these verbs, several semantically homogeneous groups motivated by a certain metaphor can be distinguished. Although, as mentioned above, light verbs display semantic preferences which can be represented as semantic classes of collocates, these generalizations are probabilistic and not rule-based.

Figure 2 shows a naïve representation of the family of constructions based on the verbs under study. Each extension is illustrated by a prototypical example in both languages.

4.1. Communication is transfer

One of the most conspicuous groups both in Italian and in Russian is formed by LVCs which designate communication. Communication is metaphorically represented as transfer (Reddy 1979, Goldberg 1995: 148), where the message is an object given by a mostly human Agent (or a text which metonymically stands for a human being) and the Addressee of the information is interpreted as the Recipient.

This group of LVCs can be schematically represented as follows:

Sem: CAUSE-“RECEIVE” ⟨ agt theme rec ⟩
 Syn: *dare/davat'* Subj Obj₁ Obj₂
 Subj: [+human] / [+text]
 Obj₁: [+communicative act]
 Obj₂: [+human]

Examples motivated by this metaphor and attested in the CORIS corpus of written Italian (17), and the Russian National Corpus³ (18) include:

(17) *Non sono in grado di darti una spiegazione precisa perché non conosco bene il linguaggio specialistico* 'I can't give you an exact explanation because I don't know the specialized language well'.

³ All the examples hereunder were extracted from the CORIS Corpus (for Italian) and from the Russian National Corpus (for Russian).

(18) *Ne možeš' li ty mne dat' sovet kak vydajuščijsja političeskij dejatel' našego vremena?* 'Can you give me advice as an outstanding political figure of our time?'

Table 1 lists, in descending order, the nouns that most frequently fill the nominal slot of the constructions of this group in the two languages.

Table 1. Nouns most frequently used in the direct object slot of LVCs denoting communicative act

Italian Web 2016			Russian Web 2011		
Lemma	Absolute frequency	Ipm	Lemma	Absolute frequency	Ipm
risposta 'answer'	97821	19,60	otvet 'answer'	92885	6,38
notizia 'news'	36715	7,36	sovet 'advice'	65659	4,51
consiglio 'advice'	35916	7,20	ob"javlenie 'announcement'	55064	3,78
informazione 'information'	32072	6,43	garantija 'guarantee'	41054	2,82
indicazione 'indication'	26203	5,25	informacija 'information'	36740	2,52
comunicazione 'announcement'	18832	3,77	rekomendacija 'recommendation'	34729	2,39
benvenuto 'welcome'	18488	3,71	pokazanija 'testimony', 'evidence'	29418	2,02
spiegazione 'explanation'	16389	3,28	ukazanie 'indication'	19132	1,31
giudizio 'judgement'	12642	2,53	soglasie 'consent'	17758	1,22
parere 'opinion'	9999	2,00	komanda 'command'	15007	1,03

Source: own research.

We can observe a wide margin of overlap between the light verb constructions in both languages. This semantic group can be further divided into subgroups, a few of which are:

- an order: *dare un ordine* – *davat prikaz* 'to give an order', *dare un compito* – *davat' zadanie* 'to give an assignment', *dare un comando* – *davat' komandu* 'to give a command';
- advice: *dare un consiglio* – *davat' sovet* 'to give advice', *dare una raccomandazione* – *davat' rekomendaciju* 'to give a recommendation', *dare suggerimento* – *davat' podskazku* 'to give a suggestion';

- a reaction to a verbal act: *dare una risposta – davat' otvet* 'to give an answer', *dare una smentita – davat' oproverženie* 'to give a denial', *dare una conferma – davat' podtverždenie* 'to give a confirmation';
- promise: *dare un'assicurazione – davat' zaverenie* 'to give assurance', *dare una garanzia – davat' garantiju* 'to give a guarantee';
- permission: *dare un permesso – davat' razrešenie* 'to give permission', *dare il consenso – davat' soglasie* 'to give consent';
- information: *dare un'informazione – davat' informaciju* 'to give information', *dare un annuncio – davat' ob''javlenie* 'to make an announcement' (lit. to give an announcement), *dare una comunicazione – davat' soobščenie* 'to make an announcement' (lit. to give an announcement).

There are also points of divergence between the LVCs denoting communication in the two languages. One quite productive subgroup in Italian that does not exist in Russian designates greetings:

- (19) *dare il benvenuto* 'to give a welcome', *dare la buonanotte* 'to say goodnight' (lit. to give goodnight), *dare un saluto*⁴ 'to give a greeting'.

To express the corresponding concept in Russian there is a simple verb (*privetstvovat'* 'to welcome', *privetstvovat' / zdorovat'sja* 'to greet') or a combination with a full verb (*poželat' spokojnoj noči* 'to say goodnight' (lit. to wish goodnight)).

4.2. Causal events are transfers

Another highly productive metaphor in both languages provides a link between the basic sense of the verbs *dare* and *davat'* and that of causation. Causal events are interpreted as transfers (Goldberg 1995:144, Lakoff, Johnson 1999: 195–199) of an effect (a physical or emotional state or a property) from the Cause, which is usually non-volitional and non-human, to a mostly human Experiencer.

The LVCs of this group form the following pattern:

Sem: CAUSE-“RECEIVE” ⟨ cause eff exp ⟩

Syn: *dare / davat'* Subj Obj₁ Obj₂

Subj: [±concrete], [±animate], [-volitional]

Obj₁: [+physical state] / [+emotional state] / [+property]

Obj₂: [±animate]

⁴ To express the same idea in Italian a construction with a light verb *fare* 'to do, to make' (*fare un saluto* 'to say hello', 'to give a greeting' lit. to make a greeting) or a simple verb *salutare* 'to greet' are more frequently used.

This metaphor licenses the occurrence of the light verbs *dare* and *davat'* in the following sentences:

- (20) *Guido, in quel gruppo di amici, era uno dei pochi a cui un'escursione in montagna dava una soddisfazione incredibile* 'Guido, in that group of friends, was one of the few to whom an excursion in the mountains gave incredible satisfaction';
- (21) *Oni ljubjat svoj gorod – i èto daët im sily dlja uspešnoj tvorčeskoj raboty* 'They love their city – and this gives them the strength for the successful work'.

The most frequent lexical fillers of the noun slot of this group of LVCs in both languages can be seen in Table 2.

Table 2. Nouns most frequently used in the direct object slot of LVCs designating causal events

Italian Web 2016			Russian Web 2011		
Lemma	Absolute frequency	Ipm	Lemma	Absolute frequency	Ipm
vita 'life'	195777	39,24	predstavlenije 'idea', 'insight'	52774	3,63
origine 'origin'	34216	6,86	žizn' 'life'	15087	1,04
sensò 'sense', 'meaning'	33994	6,81	sila 'strength'	13315	0,91
inizio 'start'	30237	6,06	znanie 'knowledge'	12703	0,87
forma 'form'	29709	5,95	načalo 'start'	12687	0,87
fastidio 'bother'	29230	5,86	oščuščenije 'feeling', 'sensation'	5833	0,40
idea 'idea'	28923	5,80	uverennost' 'confidence'	4756	0,33
importanza 'importance'	21753	4,36	otdych 'rest'	4542	0,31
forza 'strength'	21554	4,32	ponimanie 'understanding'	2887	0,20
valore 'value'	20725	4,15	čuvstvo 'feeling'	2434	0,17

Source: own research.

The Russian verb *davat'* combines only with nouns that have positive connotations, while the Italian verb *dare* is also used with nouns which designate negative emotional or physical states:

- (22) *dare fastidio a qlcu* 'to give sb. bother', *dare preoccupazioni a qlcu* 'to give sb. worries', *dare la nausea a qlcu* 'to give sb. nausea'.

4.3. Enablement is transfer

The metaphor ENABLEMENT IS TRANSFER licenses the conceptualization as a transferred physical object, of the assistance and advantage provided by the subject referent to the dative argument. The Beneficiary metaphorically receives favorable circumstances for the realization of the goal.

The schematic representation of this extension is the following:

Sem: CAUSE-“RECEIVE” ⟨ agt theme ben ⟩

Syn: *dare/davat'* Subj Obj₁ Obj₂

Subj: [±concrete], [±animate]

Obj₁: [+abstract]

Obj₂: [±concrete], [±animate]

The light verbs' uses motivated by this metaphor are exemplified in sentences (23–24):

- (23) *Esso dà il diritto alle università di brevettare scoperte nate dalla ricerca finanziata dal Governo e di concederle a società commerciali* 'It gives universities the right to patent discoveries arising from government-funded research and to grant them to commercial companies'.

- (24) *Ja polagaju, što nužno uveličivat' zarabotnuju platu, davat' vozmožnost' graždanam polučit' rabotu* 'I believe that it is necessary to increase wages, give citizens the opportunity to get a job'.

As can be seen from the list of nouns that most frequently fill the direct object slot (Table 3), the constructions in this group also present many cases of convergence between the two languages.

Table 3. Nouns most frequently used in the direct object slot of the LVCs designating enabling

Italian Web 2016			Russian Web 2011		
Lemma	Absolute frequency	Ipm	Lemma	Absolute frequency	Ipm
possibilità 'possibility'	82415	16,52	возможность 'possibility'	649093	44,60
contributo 'contribution'	59951	12,01	pravo 'right'	78535	5,40
spazio 'space'	48632	9,75	osnovanie 'basis'	31521	2,17
opportunità 'opportunity'	18580	3,72	šans 'chance'	21109	1,45
modo 'opportunity'	17905	3,59	preimuščestvo 'advantage'	19660	1,35
impulso 'impulse'	17525	3,51	povod 'reason', 'opportunity'	18722	1,29
diritto 'right'	17210	3,45	volja 'freedom'	15275	1,05
tempo 'time'	17209	3,45	tolčok 'push'	13581	0,93
aiuto 'help'	11735	2,35	svoboda 'freedom'	12412	0,85
accesso 'access'	9450	1,89	impul's 'impulse'	9041	0,62

Source: own research.

4.4. Directed actions are transferred objects

By virtue of the next metaphor, physical, perceptual and cognitive actions intentionally directed at a Recipient are interpreted as objects given (Goldberg 1995: 149, Turner, Fauconnier 1999: 411).

The first group of constructions designate an act where physical force is applied to a mostly animate non-willing Recipient (Maleficiary) causing some damage and can be schematically represented as follows:

Sem: CAUSE-“RECEIVE” ⟨ agt theme mal ⟩

Syn: dare/davat' Subj Obj₁ Obj₂

Subj: [+human]

Obj₁: [+action], [+contact]

Obj₂: [±animate], [-volitional]

The instances of the LVCs which form this group in both languages are exemplified in the following sentences:

- (25) *Uno di loro gli ha dato un calcio e lo ha fatto saltare in aria* 'One of them kicked it and blew it up'.
- (26) *Kogda ja vpervyje poceloval Svetlanu, ona dala mne poščečinu* 'When I first kissed Svetlana, she gave me a slap'.

As shown in Table 4, not only are there more nouns filling the nominal slot of the construction in Italian, but they are also more frequent.

Table 4. Nouns most frequently used in the direct object slot of the LVCs designating striking

Italian Web 2016			Russian Web 2011		
Lemma	Absolute frequency	Ipm	Lemma	Absolute frequency	Ipm
colpo 'blow'	8725	1,749	poščečina 'slap'	966	0,066
spinta 'push'	8699	1,743	pinok 'kick'	173	0,012
calcio 'kick'	4244	0,851	podzatylnik 'blow (on the back of the head)'	166	0,011
scossa 'shake'	3916	0,785	opleucha 'slap'	111	0,008
schiaffo 'slap'	1808	0,362			
pugno 'blow' (lit. fist)	1652	0,331			
pacca 'slap'	1475	0,296			
spallata 'shove with the shoulder'	1182	0,237			
sferzata 'whipping'	959	0,192			
botta 'blow'	932	0,187			

Source: own research.

The majority of nouns found within this group are deverbal, as in (27):

- (27) *dare un colpo* 'to give a blow', *dare una spinta* 'to give a push', *dare un calcio* 'to give a kick'.

This semantic group also includes denominal nouns with the suffix '-ata', which indicate the part of the body (28) or the instrument with which you strike a blow (29):

(28) *dare una spallata* 'to give a shove with the shoulder', *dare una testata* 'to give a head butt';

(29) *dare una martellata* 'to give a blow with a hammer', *dare una coltellata* 'to knife'.

In the construction *dare un pugno* 'to give a blow' (lit. to give a fist), the name of the instrument metonymically indicates the action for which it is used. To translate these constructions in Russian it is necessary to resort to the full verb *udarit'* 'to hit' followed by the instrument:

(30) *dare una testata* 'to give a head butt' > *udarit' golovoi* 'to hit with the head';

(31) *dare una martellata* 'to give a blow with a hammer' > *udarit' molotkom* 'to hit with a hammer';

(32) *dare un pugno* 'to give a blow' > *udarit' kulakom* 'to hit with the fist'.

The lexicalization of this concept in Russian may be entrusted in some cases to a construction with a noun derived from a noun which indicates the part of the body which receives the blow, as is the case with *davat' poščěčĭnu* 'to give a slap' (derived from *ščeka* 'cheek'), and *davat' podzatyĭ'nik* (derived from *zatylok* 'the back of the head').

Another type of LVCs motivated by the metaphor DIRECTED ACTIONS ARE TRANSFERRED OBJECTS designates physical contact between the Agent and the Recipient of the action as represented in the scheme:

Sem: CAUSE-"RECEIVE" < agt theme rec >

Syn: *dare/davat'* Subj Obj₁ Obj₂

Subj: [+human]

Obj₁: [+action], [+contact]

Obj₂: [+animate], [+volitional]

This group is rather productive in Italian:

(33) *dare un bacio* 'to give a kiss', *dare un abbraccio* 'to give a hug', *dare una carezza* 'to give a caress', *dare un pizzico (un pizzicotto)* 'to pinch'.

In contrast, Russian only has simple verbs to express these actions: *celovat'* 'to kiss', *obnimat'* 'to hug', *laskat'* 'to caress', *ščipat'* 'to pinch'.

Relevant examples attested in the CORIS corpus include:

- (34) *Un giorno portai alla maestra una mela e lei mi diede un bacio* 'One day I brought the teacher an apple and she gave me a kiss'.
 (35) *Diede un pizzicotto sul naso a Sophie e ottenne una risatina* 'She pinched Sophie's nose and got a chuckle'.

Another divergence between the two languages is found in the Italian construction with a deverbal noun derived from the past participle, which denotes a non-durative singular physical (36) or cognitive action (37):

- (36) *dare una ripulita a qlco* 'to give sth. a clean', *dare una sistemata a qlco* 'to arrange sth.';
 (37) *dare una letta a qlco* 'to have a read', *dare una controllata a qlco/qlcu* 'to check sth./sb.'.

Sem: CAUSE-"RECEIVE" ⟨ agt theme rec ⟩

Syn: *dare/davat'* Subj Obj₁ Obj₂

Subj: [+human]

Obj₁: [+physical act] / [+cognitive act], [±contact]

Obj₂: [±animate]

To express the semantic value of brevity and uniqueness of the action in Russian we must resort to a main verb eventually accompanied by a modifying adverb, like *bystro* 'quickly'.

The peculiarity of this group of constructions compared to the previous two is that the Recipient is almost always a physical object, for instance:

- (38) *Ha acceso il forno e ha dato una spolverata alle tute che nell'armadio non hanno fatto la muffa* 'He has lighted the oven and given a dusting to the suits that didn't mold in the closet'.
 (39) *È bene abituarsi comunque a dare una controllata generale a qualunque oggetto (...)* 'However, it is good to get used to checking any object'.

4.5. Public events are transferred objects

The last metaphor to be discussed here motivates some instantiations of LVCs in both languages but has more occurrences in Russian than in Italian (see Table 5).

Table 5. Nouns denoting public events most frequently used in the direct object slot

Italian Web 2016			Russian Web 2011		
Lemma	Absolute frequency	Ipm	Lemma	Absolute frequency	Ipm
lezioni (pl.) ⁵ 'lessons'	4166	0,835	concert 'concert'	12106	0,832
festa 'party'	1387	0,278	urok 'lesson' (mostly pl.)	10696	0,735
concerto 'concert'	1340	0,269	master-klass 'workshop'	1999	0,137
spettacolo ⁶ 'performance', 'show'	202	0,040	spektakl' 'performance', 'show'	997	0,069
ricevimento 'reception, party'	95	0,019	press-konferencija 'press conference'	861	0,059
pranzo 'dinner'	79	0,016	obed 'dinner'	767	0,053
			zanjatie 'lesson' (mostly pl.)	687	0,047
			predstavlenije 'performance', 'show'	673	0,046
			bal 'ball'	634	0,044
			lekcija 'lecture'	550	0,038

Source: own research.

The theme argument of the LVCs of this group denotes a public event which is metaphorically interpreted as an object given:

(40) *dare una festa* 'to give a party', *dare ricevimento* 'to give a reception';

(41) *davat' press-konferenciju* 'to give a press conference', *davat' priëm* 'to give a reception', *davat' obed* 'to give a dinner'.

⁵ The singular form of the Italian noun *lezione* 'lesson' is used in the idiom *dare una lezione a qlcu* (lit. to give sb. a lesson), which expresses an act of punishment. The Russian idiom *davat' urok komu* (lit. to give sb. a lesson) means 'to teach sb. a lesson', 'to correct sb.'

⁶ The noun *spettacolo* 'show, performance' forms with the verb *dare* a collocation *dare uno spettacolo (teatrale)* 'to put on a show', but is also used idiomatically: *dare (uno) spettacolo* 'to make an exhibition of oneself'.

Performing educational services is also metaphorically interpreted as transferring an object to a recipient:

(42) *dare lezioni (di musica)* 'to give (music) lessons';

(43) *davat' uroki/zanjatja* 'to give lessons', *davat' master-klass* 'to give a workshop', *davat' lekciju* 'to give a lecture'.

Some of the LVCs are motivated by the metaphor SENSORY PERCEPTIONS ARE TRANSFERRED OBJECTS (Goldberg 1995: 148, Turner, Fauconnier 1999: 411, Paszenda 2017: 260):

(44) *dare un concerto* 'to give a concert', *dare uno spettacolo (teatrale)* 'to give a performance';

(45) *davat' koncert* 'to give a concert', *davat' spektakl' / predstavlenie* 'to give a performance'.

Schematically the LVCs forming this group can be represented in the following way:

Sem: CAUSE-“RECEIVE” ⟨ agt theme (ben/exp) ⟩

Syn: *dare/davat'* Subj Obj₁ Obj₂

Subj: [+human]

Obj₁: [+public event]

(Obj₂: [+human])

The Recipient argument is often not lexically profiled as its referent's identity is either irrelevant or unknown.

Some examples of this metaphorical extension found in the corpora are:

(46) *Graham aveva accettato di dare il primo concerto a Sydney il 15 febbraio* 'Graham had agreed to give the first concert in Sydney on February 15th'.

(47) *Stasera daremo una festa per tutti i nostri amici più cari* 'Tonight we are giving a party for all our dearest friends'.

(48) *I eščë ja davala chastnyje uroki russkogo jazyka i literatury dlja postupajuščich v universitet* 'I also gave private lessons in Russian language and literature for applicants to university'.

(49) *V subbotu v Teatre na Taganke budet dan tysjačnyj spektakl' po romanu Bulgakova "Master i Margarita"* 'On Saturday, in the Taganka Theater, will be performed the thousandth play based on Bulgakov's novel The Master and Margarita'.

5. Conclusions

Non-fully compositional ditransitive structures with the light verbs *dare* and *davat'*/*dat'* have the same syntactic structure and express the idea of transfer as prototypical non-fixed ditransitive constructions. The only difference is that the transfer in the light verb constructions is metaphorical and the transferred entity is abstract. LVCs can be presented as extensions that inherit from the basic sense of the verbs via a metaphorical link and that form a family of related constructions.

Systematic metaphors that license extensions from the primary meaning of the verbs *dare* and *davat'* have been identified and LVCs were grouped according to these extension mechanisms. The most productive metaphors in both languages are COMMUNICATION IS TRANSFER, CAUSAL EVENTS ARE TRANSFERS, and ENABLEMENT IS TRANSFER. The domains largely overlap across the two languages, but some are richer in one language or the other.

The convergences and divergences between the constructions of the two languages resulting from this study may become the subject of further research on the phenomenon and may be taken into consideration while teaching Russian and Italian as a foreign language.

Literature

- Alba-Salas, Josep. 2002. *Light Verb Constructions in Romance: A Syntactic Analysis*. PhD. diss. Cornell University.
- Allerton, David J. 2002. *Stretched Verb Constructions in English*. London: Routledge.
- Altenberg, Bengt; Granger, Sylviane. 2001. The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics*. 22 (2): 173–194.
- Altenberg, Bengt. 2001. Contrasting delexical English make and Swedish göra. In: Aijmer, Karin (ed.). *A Wealth of English. Studies in Honour of Göran Kjellmer Göteborg*. Gothenburg: The University of Gothenburg. 195–219.
- Apresjan, Jurij D. 2008. O semantičeskoj motivirovannosti leksičeskich funkcij-kollokatov. *Voprosy jazykoznanija*. 5: 3–33.
- Apresjan, Jurij D. 2009. *Issledovanija po semantike i leksikografii*. Vol. I. Moskva: Jazyki slavjanskich kul'tur.
- Benigni, Valentina; Cotta Ramusino, Paola. 2011. Italjanske konstrukcii s glagolom podderžki fare v sopostavlenii s russkim jazykom. In: Kibrik, Aleksandr E. et al. (eds). *Komp'juternaja lingvistika i intellektual'nye tehnologii: Po materialam ežegodnoj Meždunarodnoj konferencii "Dialog"*. Bekasovo. 68–83.

- Casadei, Federica. 1996. *Metafore ed espressioni idiomatiche: uno studio semantico sull'italiano*. Roma: Bulzoni.
- Corpus of written Italian – CORIS / CODIS. <http://corpora.dslo.unibo.it>
- Family, Neiloufar. 2009. Mapping Semantic Spaces: A constructionist account of the “light verb” eat in Persian. In: Vanhove, Martine (ed.). *From Polysemy to Semantic Change: Towards a Typology of Lexical Semantic Associations*. Amsterdam: John Benjamins. 139–161.
- Family, Neiloufar. 2014. *Semantic Spaces of Persian Light Verbs: A Constructionist Account*. Leiden: Brill.
- Fillmore, Charles J.; Kay, Paul. 1996. *Construction Grammar Coursebook*. Unpublished manuscript. UC Berkeley.
- Gilquin, Gaëtanelle. 2007. To err is not all. What corpus and elicitation can reveal about the use of collocations by learners. *Zeitschrift für Anglistik und Amerikanistik*. 55 (3): 273–291.
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: Chicago University Press.
- Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Goldberg, Adele E; Jackendoff, Ray. 2004. The English Resultative as a Family of Constructions. *Language*. 80: 532–568.
- Howarth, Peter. 1998. Phraseology and second language proficiency. *Applied Linguistics*. 19 (1): 24–44.
- Iordanskaja, Lidija N.; Melčuk, Igor' A. 2007. *Smysl i sočetaemost' v slovare*. Moskva: Jazyki slavjanskich kul'tur.
- Italian Web 2016 (itTenTen16)*. <https://www.sketchengine.eu>
- Jespersen, Otto. 1942. *A Modern English Grammar on Historical Principles*. London: George Allen and Unwin Ltd.
- Juknevičienė, Rita. 2008. Collocations with high-frequency verbs in learner English: Lithuanian learners vs native speakers. *Kalbotyra*. 59 (3): 119–127.
- Kopotev, Mikhail et al. 2016. Constructional generalization over Russian collocations. In: Sanromán Vilas, Begoña (ed.). *Mémoires de la Société néophilologique de Helsinki, Collocations Cross-Linguistically*. 121–140.
- Lakoff, George; Johnson, Mark. 1999. *Philosophy in the Flesh: The Embodied Mind And Its Challenge To Western Thought*. Harper Collins Publishers.
- Langer, Stefan. 2005. A formal specification of support verb constructions. In: Langer, Stefan; Schnorbusch, Daniel (eds). *Semantik im Lexikon*. Tübingen: Narr. 179–202.
- Melčuk, Igor. 1998. Collocations and Lexical Functions. In: Cowie, Anthony P. (ed.). *Phraseology. Theory, Analysis and Applications*. Oxford: Oxford University Press. 23–53.
- Nesselhauf, Nadja. 2005. *Collocations in a Learner Corpus*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

- Nordlund, Marie. 2007. On the different uses of the Swedish verb *ta* 'take' and its translations into English. In: Magnusson, Ulf et al. (eds). *Further Insights into Semantics and Lexicography*. Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej. 73–85.
- Palancar, Enrique L. 2003. La polisemia dei verbi *dar*, *pegar* e *meter* in spagnolo. In: Gaeta, Livio; Luraghi, Silvia (eds). *Introduzione alla Linguistica Cognitiva*. Roma: Carocci, 197–211.
- Paszenda, Joanna. 2017. Motivation behind the extended senses of the Polish ditransitive construction. In: Ruiz de Mendoza Ibáñez, Francisco J. et al. (eds). *Constructing Families of Constructions: Analytical perspectives and theoretical challenges*. Amsterdam/Philadelphia: John Benjamins Publishing Company. 241–275.
- Pinker, Steven A. 1989. *Learnability and Cognition: The Acquisition of Argument Structure*. Cambridge, Mass.: MIT Press.
- Reddy, Michael. 1979. The conduit metaphor: A case of frame conflict in our language about language. In: Ortony, Andrew (ed.). *Metaphor and Thought*. Cambridge: Cambridge University Press. 284–324.
- Russian Web 2011 (ruTenTen11)*. <https://www.sketchengine.eu>
- Sinclair, John; Renouf, Antoinette. 1988. A lexical syllabus for language learning. In: Carter, Ron; McCarthy, Michael (eds). *Vocabulary in Language Teaching*. London: Longman. 140–160.
- Stein, Gabriele. 1991. The phrasal verb type 'to have a look' in Modern English. *International Review of Applied Linguistics*. 29: 1–29.
- The Russian National Corpus*. www.ruscorpora.ru
- Turner, Mark; Fauconnier, Gilles. 1999. A Mechanism of Creativity. *Poetics Today*. 20 (3): 397–418.
- Quochi, Valeria. 2016. Development and representation of Italian light-fare constructions. In: Yoon, Jiyoung; Gries, Stefan Th. *Corpus-based Approaches to Construction Grammar*. Amsterdam/Philadelphia: John Benjamins Publishing Company. 39–64.
- Wierzbicka, Anna. 1982. Why can you Have a Drink when you can't *Have an Eat? *Language*. 58 (4): 753–799.

**Was kann man in Italienisch geben, das man in Russisch nicht geben kann?
Die kontrastive Studie von den Funktionsverben *dare* in Italienisch
und *davat'*/*dat'* in Russisch**

Zusammenfassung

Funktionsverbgefüge, Konstruktionen aus einem Verb mit einem abgeschwächten semantischen Inhalt und einem der Konstruktion sinngebenden Substantiv in der Rolle des Akkusativobjektes (*prendere una decisione* 'eine Entscheidung treffen')

(wörtlich 'eine Entscheidung nehmen'), *stavit' vopros* 'eine Frage stellen') oder in einer Präpositionalphrase (*mettere a confronto* 'vergleichen' (wörtlich 'in den Vergleich stellen'), *brat' pod ochranu* 'unter Bewachung stellen' (wörtlich 'unter Bewachung nehmen')), bilden eine auffällige Klasse von idiosynkratischen, aber halbkompositionellen und halbproduktiven Konstruktionen in Italienisch sowie in Russisch. Diese Studie konzentriert sich auf Konstruktionen aus dem Funktionsverb *dare* 'geben' in Italienisch und dem Funktionsverb *davat' / dat'* 'geben' in Russisch und einem Nomen in der Rolle des Akkusativobjekts. Das Model der Konstruktionsgrammatik (Goldberg 1995, 2006), das einen einheitlichen Rahmen zur Untersuchung der Konstruktionen mit direkter sowie figurativer Bedeutung anbietet, und das Familienkonzept wurden angewendet, um die Verwandtschaft der Konstruktionen zu erforschen.

Halbkompositionelle ditransitive Konstruktionen mit dem Funktionsverb *dare* und *davat' / dat'* haben dieselbe syntaktische Struktur und drücken die Idee des Transfers aus wie prototypische nicht feste ditransitive Konstruktionen. Der einzige Unterschied ist, dass der Transfer in den Funktionsverbgefügen metaphorisch ist und die transferierte Einheit abstrakt ist.

Systematische Metaphern, die die Erweiterung der Grundbedeutung der Verben erlauben, wurden identifiziert und Funktionsverbgefüge wurden nach Erweiterungsmechanismen gruppiert. Die produktivsten Metaphern in beiden Sprachen sind KOMMUNIKATION IST TRANSFER, KAUSALE EREIGNISSE SIND TRASFER und ERMÖGLICHUNG IST TRANSFER.

Die in der Studie erschlossenen Übereinstimmungen und Unterschiede zwischen den Konstruktionen in den zwei Sprachen können zum Gegenstand weiterer Forschung werden und beim Unterrichten des Russischen und des Italienischen als Fremdsprache berücksichtigt werden.

Antonio Pamies

University of Granada

Spain

<https://orcid.org/0000-0001-8193-9359>

Phrasal Verbs, Idiomaticity and the Fixedness Continuum

Abstract. Leonard Talmy assigns the *phrasal verb* to a typological group called *satellite-framed languages*, whose prototype would be the Germanic languages, however, this construction exists also in Romance and Slavic Languages (in lesser quantity). Besides, these structures have much more in common with phrasemes than with syntactic combinations, and the function played by these verbs is quite similar to that of a proportional number of Romance and Slavonic prefixed verbs. This work analyzes both mechanisms within a *continuum* of verbal constructions with different degrees of fixedness and idiomaticity, going from regular syntactic combinations to completely lexicalized verbs, with an intermediate zone occupied by several idiomatic structures: light verb collocations, verbal idioms, separable phrasal verbs, inseparable phrasal verbs and compound verbs.

Key words: *phraseology, phrasal verb, compound verb, syntagmatic verb, preverb, postverb, prefix*

1. Introduction

Assuming that the most basic element of an event is the expression of displacement in space-time, Talmy (1985) divides the world languages into types which depend on whether they encode the “trajectory” into the verb itself (*verb-framed languages*) or into a “satellite”, that is, a directive element external to the verb but associated with it (*satellite-framed languages*). If we take examples such as sp. *entrar/salir*; fr. *entrer/sortir*, and compare them with rs. *входитъ/выходитъ*, pol. *wejść/wyjsć*, grm. *hereinkommen/herauskommen* and eng. *come in/come out*, it could seem reasonable to assign Romance languages to the first type, whereas Slavonic and Germanic languages would belong to the second (Kopecka 2004: 114).

A prototypical example illustrating this dichotomy would be the syntactic model *Paul ran out of the school*, where Romance languages start from a specific verb expressing an abstract trajectory, complementing it with a second verb describing the concrete “manner” of moving. By contrast, Germanic and Slavonic use only one verb, describing a manner of the movement, adding a locative particle to indicate its direction (or *path*).

Table 1. Trajectories expressed by verbs and “satellites”

TRAJECTORY EXPRESSED BY A VERB	TRAJECTORY EXPRESSED BY A “SATELLITE”
sp. Pablo <i>salió</i> de la escuela <i>corriendo</i> fr. Paul est <i>sorti</i> de l'école <i>en courant</i> pt. o Paulo <i>saiu</i> da escola <i>correndo</i> it. Paolo è <i>uscito</i> <i>correndo</i> dalla scuola cat. en Pau ha <i>sortit</i> de l'escola <i>corrent</i> rmn. Pavel a <i>iesit</i> <i>fugind</i> din școală	eng. Paul <i>ran</i> <i>out</i> of the school grm. Paul <i>rannte</i> <i>aus</i> der Schule rs. Павел <i>выбежал</i> из школы ukr. Павло <i>вубіг</i> із школи cz. Pavel <i>vyběhl</i> ze školy pol. Paweł <i>wybiegł</i> ze szkoły

Source: own research.

However, in fact, phrasal verbs may be found in several language families. They were identified long ago in the Spanish language by J. D. Luque Durán (1972: 808), who called them *verbos con extensión preposicional*. Raffaele Simone also observed them in Italian, and called them *verbi sintagmatici*, which is a literal translation of the term *phrasal verbs* (1996: 156–157), being followed later by Iacobini (2009) Artusi (2016), and also by Calvo Rigual, who compared these Italian verbs with their Spanish and Catalan counterparts (2008: 59–61).

Several Russian constructions, traditionally considered as idioms or as verbs with restricted arguments, belong to the same lexical association between a verb and an adverbial or prepositional particle, forming synthetically a verbal expression with a different meaning: e.g. eng. *piss off*; it. *mandare giù* (*send down: “swallow”); fr. *sauter dessus* (*jump over: “attack”); sp. *venirse abajo* (*come+self+down: “collapse”); pt. *estar por dentro* (*be by inside “be well informed”) (cf. Pamies 2018b).

An English sentence, such as *the apple falls down from the tree*, may be expressed in Italian with exactly the same structure: *la mela cade giù dall'albero* (Russo 2014: 97). Conversely, English may also use *verb-framed* constructions, such as *the frog escapes from the jar* whereas Italian uses a phrasal verb, therefore, a *satellite-framed* construction: *la rana salta via dal barattolo* (*the frog jumps out of the jar) (Anastasio 2014: 110–111).

Although many Russian linguists believe that their language *has no phrasal verbs at all* (e.g. Mudraya et al. 2008: 296), we can find counter-examples like rs. *идти налево* (*go to+left: “be unfaithful” [sexually]), which fits exactly in the definition of the phrasal verb: multi-word units composed by a verb followed by an adverbial particle, with a spatial or locative meaning, whose sense is global and cannot be derived from the individual meanings of the parts (cf. Biber, Conrad & Leech 2002). Even if they are called *idioms* (2008: 297), the fact is that Mudraya et al. quote examples that could be analyzed as canonical phrasal verbs: *сводитъ вместе* (“bring together”), *оставлять позади* (“leave behind”); *двигатся вперед* (“move forward”).

Phrasal verbs have been considered as a syntactic phenomenon, a kind of *verbal periphrasis* (cf. Bolinger 1971; Quirk et al., 1972; Fraser 1976; Courtney 1983), though they fulfill the definition of phraseological units: several lexemes, formal fixedness, semantic idiomaticity. E.g., *to play something down* has little to do with a game or a sport, since its unique and indivisible meaning is “minimize the importance of something” (Riguel 2014: 120). The same can be said about the German “separable” and “inseparable verbs”. However, as well as these Russian *idioms*, they all belong to phraseology, since they are multi-lexemic, fixed, and idiomatic (cf. Kunin 1996: 308–309). On the other hand, phrasal verbs share many properties with certain verbs which are mistakenly considered as “prefixed”.

2. Postverbs and preverbs

A preverb is an element that is attached to a verb but retains a relative autonomy, that allows it to make semantic modifications giving rise to new verbs (Buridant 1995: 292). Formally, it is opposed to the postverb, which, however, fulfills the same function in the inverse position. Therefore, the locative English adverb *down* is a “preverb” in *download* and a “postverb” in *kneel down*, but the mechanism is basically the same: joining a locative adverb to a verb in order to form a new unit, that acquires a single function in the sentence, with an indivisible meaning. Depending on the language, it is even possible that, with the same verb and the same sense, a preverb may be convertible into a postverb, depending on its conjugation. E.g., grm. *mitkommen* → *kommst mit mir* (“accompany me”); *austrinken* (*out drink “drink completely”) → *trink das Glas aus* (Pamies & Pazos 2018); fr. *bien vouloir* (*well want “accept”) → *je veux bien* (*I want well); *bien tomber* (*well fall “to come/happen in the right moment”) → *ça tombe bien* (*it falls

well). The commutability between these preverbs and postverbs confirms their functional synonymy, which can also be observed in English, e.g. between *break up/upbreak*; *hold up/uphold*: *look over/overlook*.

If we do not pay attention to the graphic separation and the order of the components (*ran out* vs. *выбежал*), the Slavic system is not so different from the Germanic one: in both cases the particle identifies the abstract direction of the movement, and the verb indicates only its concrete manner (*walk, run, jump, ride, drive, swim, navigate, fly*, etc.). However, the fact that an analytical procedure such as the addition of lexemes or morphemes can fulfill the same function as a synthetic procedure does not rule out the possibility that the same language may use both mechanisms (cf. Kopecka 2004; Michot et al. 2015; Pamies 2018b). On the other hand, nothing prevents these mechanisms from being applied to verbs that are alien to physical displacement. Movement is one of the most productive source domains of figurative language, and there is no guarantee either that metaphorical displacements have the same morpho-syntactic regime as the literal ones.

3. Aspectuality

Particles may also contain information about the *internal tense* of an action, either as a morpheme of aspect (throughout the whole verbal system) or as an inherent feature of its lexical meaning (*Aktionsart*). Slavic languages systematically distinguish the imperfective from the perfective aspect, by adding a prefix to the imperfective, in order to obtain its perfective correlate, never the opposite¹ (e.g., rs. *играть* “play” ≠ *сыграть* “having played”). Due to the high percentage of homonymy between preverbs and prefixes in these languages, Slavists often use these terms as synonyms, but they divide them into two subclasses: the “lexical” one (which entails a new verbal meaning) and the “grammatical” one, also called “de-lexicalized” or even “empty”, which conveys exclusively the perfective aspect (Fougeron 1995: 256). In practice, this distinction is very similar to the aforementioned borderline between preverb and prefix, because the former would be lexemes and the latter would be mere grammatical morphemes.

¹ A verb with a new meaning can only be derived from a perfective one, and another imperfective verb can emerge from it by adding the appropriate suffix (Fougeron 1995; Cygal-Krupa 1995; Žaucer 2002). We leave aside here the few exceptions, such as verbs expressing a state (e.g., rs. *жить*; pol. *żyć* “live”) or borrowed French verbs (e.g., rs. *резервировать* “to book”; pol. *replikować* “retort”; sln. *reorganizati* “reorganize”).

A complementary test to avoid confusing them is translating them into a non Slavic language: if the particle is an aspectual prefix: the translation does not change the verb but its conjugation (morpheme):

rs. *читать* ≠ *прочитать* → eng. *read* ≠ *have read* → sp. *leer* ≠ *haber leído*;
rs. *писать* ≠ *написать* → eng. *write* ≠ *have written* → sp. *escribir* ≠ *haber escrito*.

By contrast, if the particle is a preverb, the translation does not change the conjugation but the verb (lexeme):

rs. *ходить* ≠ *входить* → eng. *walk* ≠ *enter* → sp. *andar* ≠ *entrar*;
rs. *играть* ≠ *выиграть* → eng. *play* ≠ *win* → sp. *jugar* ≠ *ganar*.

However, besides this dichotomy, other aspectual values can also be marked by particles, whose relationship with prepositions is obvious (Fougeron 1995: 267). For example, in Russian, the polysemic preposition *по* ("by"/"on"/"along"/"because of"/"according to") acts as an **inchoative** prefix in *пойти* (*by+go: "begin to go"), *побегать* (*by+run: "begin to run"); while the preposition *из* ("from") is a **terminative** prefix in *измылить* (*from+soap: "spend the soap to the end") (Fougeron 1995: 258–259). In Polish, the preposition *z* ("from") is an inchoative prefix at *zsinieć* (*from+pale+ACTION: "turn pale") and *po* ("behind") has this function in *pokochania* (*behind+love: "fall in love"), while the preposition *do* ("until") is a terminative prefix in *dopić* (*until+drink: "drink to the end"), *dośpiewać* (*until+sing: "sing to the end" [a melody]), *dolecieć* (*until+fly: "arrive (flying)"), *dorysować* (*until+draw: "finish a drawing") (Cygalski-Krupa 1995: 270–274). English can use postverbs for aspectual purposes (e.g., *clean* ≠ *clean up*; *drink* ≠ *drink out*), as well as German: *er trank das Bier* ≠ *er trank das Bier aus*). But, even in these languages, aspectuality is neither limited to this aspectual couple nor to these markers. Other aspects, such as the inchoative, may have their own (figurative) locative marker.

By metaphorically projecting space over time, the telic Aktionsart can express the passage from a starting point to an ending point (e.g., sp. *entrar en la casa* *enter into the house), whereas the atelic one represents the trajectory (e.g., *caminar hasta la casa* *walking until the house), and, in both cases, the mark of this dichotomy would be the preposition. (cf. Kopecka 2004: 112). In Romance languages, the inchoative aspect of an event may be marked by the addition of locative preverbs (sp. *enloquecer*; it. *impazzire*; cat. *embogir*; pt. *endoidecer* (*in+crazy+VSUF: "to get crazy"). The same may happen in German: *lieben* ("love") ≠ *verlieben* ("fall in love"). This system may also express the **terminative** aspect in Germanic languages: e.g., grm. *austrinken* (*out+drink) "drink to the end", a formulation almost identical to that of their Polish semantic equivalence, *dopić* (*until+drink). There are also pairs

of Romance preverbs that symmetrically oppose between them inchoative and terminative actions, as in fr. *emménager* (“to start living in a new house”) vs. *déménager* (“to stop living in one house, to move to another one”), where the figurative spatial movement refers to the initial and final phase of the same process. This is also possible for Slavic and Germanic preverbs: rs. *включить* (*in+connect) vs. *выключить* (*out+connect), grm. *anschalten* (*near+switch) ≠ *ausschalten* (*out+switch), whose Aktionsart refers to the initial and final phases of a process (in this case, to start or stop functioning), by means of a grammatical metaphor, mapping two antonymic spatial positions on the temporal domain. In the inverse order, the same system is ruling English postverbs, such as *switch on* ≠ *switch off*.

Both in Romance and Slavic languages, the preverb marker of inchoativity is not always the same, and no preverb expresses only this aspect² (cf. Cygal-Krupa 1995: 278). The reflexive pronominal form has an inchoative value in sp. *irse* (*go+oneself: “to leave”) but not in *pasearse* (“have a walk”). In Russian, the preverb *в* (“in”) is inchoative in *влюбиться* (*in+love+oneself: “fall in love”) but not in *всадить* (*in+sit: “to stick”). In French, spatial preverbs also express inchoativity in *s’envoler* (*oneself+in+fly) and something similar can be found in German (*wegfliegen* *far+fly), and Russian (*улететь* *in+fly), all of them meaning “fly off”, whereas English and Italian use here a postverb (*fly off*; *volare via* *fly away). In this particular case, Spanish does not have a syntagmatic verb and resorts to inchoative periphrastic means (*salir volando* *exit flying), or lexicalized collocations (*tomar vuelo* *take flight / *alzar el vuelo* *raise the flight). Inchoative preverb constructions are therefore closer to composition than to derivation, since their overall sense is unpredictable, either by a rule or by the meaning of its components (Pamies 2018b). Here, the distribution of languages no longer corresponds to Talmy’s typology mentioned in table 1, since languages of the same family would not necessarily share the same column. Furthermore, this distribution may change from one verb to another.

4. Composition vs. derivation

Martinet (1960) included among the *synthèmes* any kind of amalgam between *monemes*, thus, not only composition but also derivation. However,

² In addition, there are some verbs that, by nature, are only interested in one phase of a process, either initial or final, but not in the intermediate trajectory. E.g., esp. *to saddle/to unsaddle* (a horse); fr. *seller/déseller* (un cheval).

their degree of lexicalization is far from being homogeneous. In fact, we may even distinguish different degrees among what morphology calls *prefixes*, since they can also proceed from prepositions and/or adverbs, some of them may have kept part of their original autonomy, and, therefore, behave as preverbs. Verbal prefixes produce analogical series: speakers may assume that eng. *retwit* is to *twit* what *reappear* is to *appear*. Preverbs do not behave this way; the semantic modifications they bring about are neither systematic nor predictable, but lexically specific: e.g., in the English preverbs *down* and *under*, whose semantic role is not the same in *download*, *downturn*, *downshift*, *understand*, *undergo*, *underwrite* or *undertake*. The Spanish preverb *entre* ("between") does not produce the same meaning effect in *entrecortar* ("cut intermittently"), *entrebir* ("half-open"), *entrever* ("to catch a glimpse"), *entretener* ("distract"). The same can be said of postverbs: e.g. eng. *around*, does not fulfill the same role in *stick around* than in *hang around* or *mess around*.

According to this criterion, the preverb (*under*<1>) would be the synchronic homonym of a prefix (*under*<2>), whose value is regular and predictable (e.g. *underrate*, *underlie*, *undersell*, *understate*, *underestimate*, *underfeed*, *undercut*). This item is (more or less) equivalent to the Latin prefix *sub-* that we find in sp. *subdividir*, *subcontratar*, *subestimar*, *subyugar*, *subrogar*, *subordinar*). The German particle *ver-* ("wrong") is perfectly regular³ in *verachten* (*wrong+estimate: "despise"), *vertreiben* (*wrong+drive: "expel"), *verraten* (*wrong+advise: "betray"); *verlaufen* (*wrong+walk: "to go wrong", which can be applied to many other movement verbs, as in *verfahren*, *verfliegen*, *verschwimmen*... *verspazieren*, all of them meaning "to miss the right way" (riding/driving, flying, swimming or taking a walk), thus, it is a prefix. However, this particle is quite idiomatic in *verstanden* (*wrong+stand: "comprehend"), *vergeben* (*wrong+give: "forgive"), *verführen* (*wrong+lead: "seduce") or *verkaufen* (*wrong+buy: "sell"), where *ver-* is a preverb. The same could be said about *über*<1> and *über*<2> ("over"): this preposition is a regular prefix in *überbieten* (*over+provide: "surpass"); *überdauern* (*over+last: "survive"); *übereilen* (*over+rush: "to be too hasty"), *überessen* (*over+eat: "eat too much"), etc., but it is an idiomatic preverb in *übersetzen* (*over+sit: "translate") or *überlaufen* (*over+walk: "overflow").

In Romance languages, we find also this homonymy. The spatial particle *sobre* ("on/over") behaves as a regular prefix in predictable derived verbs, such as *sobrecargar* ("overload"), *sobrevalorar* ("overrate"), *sobreestimar* ("overestimate"), *sobreescribir* ("overwrite"), *sobrealimentar* ("overfeed"). But it behaves as preverb in idiomatic compounds, such as *sobrecoger*

³ Cf. Bayley 1997.

(*on+take: “excite”/“impress”), *sobrellevar* (*on+carry: “endure”) *sobreentender* (*on+comprehend “take for granted”). Unlike prefixes, preverbs and postverbs are idiomatic components, whereas prefixes are grammatical instruments.

A complementary criterion is that, unlike a prefix, a preverb exists as a word by itself (Pamies 2018b). Verbs like *reprobar* (“condemn”) *prometer* (“promise”) or *disparar* (“shot”) are actually prefixed derivatives, because *re-*, *pro-* and *an-* do not exist alone. This is even more evident when the verb itself does not exist alone too; e.g., sp. *producir*, *traducir*, *inducir*, *deducir*, which cannot be compounds since their “parts” are not words. Since the verb **ducir* does not exist anymore in modern Spanish, the particle becomes a mere prefix.

By contrast, verbs like eng. *overcome*, *underlie*, sp. *entretener* (between+have: “distract”) or *sobrecoger* (on+take: “impress”) are compounds, because they have two lexemes, fixedness and idiomaticity (Pamies 2007, 2017). The distinction between preverb and prefix coincides with the extreme border between words and phrasemes (Pamies 2018b).

Although all of them are of prepositional or adverbial origin, only the preverb is still a “satellite”, whereas the prefix has become totally amalgamated to the verb: speakers no longer divide semantically verbs such as *suppose*, *compose*, nor *suppress*, *repress*, *compress*, *express*, because – from a communicative point of view – distinguishing their “parts” would be a nonsense⁴. Therefore, what the morphological tradition calls inaccurately *prefix*, would be an overlapping mixture, covering three degrees of lexicalization: (1) preverb of a compound verb (*upgrade*, *download*), (2) actual prefix of a derived verb (*retwit*, *unlock*), (3) etymological sediment of a current simple verb, morphologically irrelevant (*repeat*, *produce*).

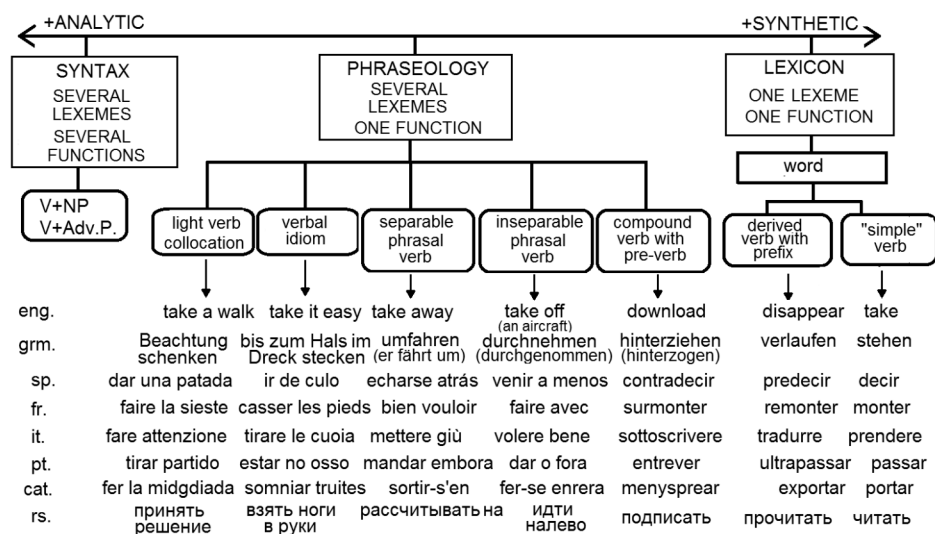
In Slavic languages, the boundary between preverb and prefix may be even more blurred: one could deduce that, since rs. *ব্যি* / pol. *wy* (“outwards”) do not exist as words, they would be prefixes in rs. *ব্যি* *ব্যি* *ব্যি* / pol. *wy* *wy* *wy*: “exit”, whereas rs. *в* and pol. *w* (“in”) would be preverbs (rs. *в* *в* *в* / pol. *w* *w* *w*).

⁴ The (apparent) exceptions are generally due to the semantic darkening or disappearance of the verbal base itself. The Latin verbs *ducere* “lead”, *fugire* “flee”, *petere* “beg”, *premere* “push” became unrecognizable in Spanish, so that their derived verbs ceased to be analyzable (e.g. *reducir* “reduce”, *conducir* “lead/drive”, *refugiar* “shelter”, *repetir* “repeat”, *competir* “compete”, *reprimir* “repress”, *comprimir* “compress”. These examples are perceived by current speakers as “simple” words, whereas *releer* (“rereading”), *recalcular* (“recalculating”), *reformular* (“reforming”), *conformar* (“make up”), are easily recognizable as derivatives, since their verbal element is still a Spanish word (*reescanear* “rescan”, *redimensionar* “resize”, *recauchutar* “retread”, *codirigir* “co-direct”, *cofinanciar* “to finance jointly”, *configurar* “to shape/set up”, *reconfigurar* “reset”).

wchodzić: “enter”), since these particles also exist as prepositions. In order to avoid such an exaggeration, a **secondary criterion** must be applied: that of systematic productivity⁵. There is a stable and general rule, which makes it possible that a displacement verb with the prefix *вы/вы* implies “outwards” and, with the one of *в/в*, implies “inwards”, it is not an unpredictable casuistry but a regular productive phenomenon, therefore, both *вы/вы* and *в/в* can be considered as prefixes **when they follow a rule**: *выводить* (*out+lead: “take out”), *выезжать* (*out+ride/drive: “leave”), *носить* (*out+carry: “take away”), *вылетать* (*out+fly: “fly out”) etc. According to this point of view, there would be two homonymic particles “в” in Russian, the prefix of *входить* (literal and rule governed) and the preverb of *влюбиться* (idiomatic and unpredictable).

According to their degree of fixedness and idiomaticity, we may set out a cline of verbal constructions, placed along a *continuum* going from regular syntactic combinations to completely lexicalized verbs, with several intermediary constructions, based on different phraseological mechanisms.

Figure 1. Verbal phraseologisms within the syntax-lexicon continuum



Source: adapted from Pamies & Pazos 2018 and Pamies 2018b.

⁵ *Productivity* is understood here as the ability of coining new units following regular rules, contrary to *creativity*, which refers to the possibility of creating units without a rule (Dal 2004: 3).

5. Literal and figurative displacement

The locative particles are not always “directive” since their relations with space can be figurative: e.g., in phrasal verbs such as *give up* (“surrender”), *burn out* (“exhaust”), *knock out* (“render unconscious”) (cf. García Vega 2011). Such “displacements” are metaphorical in many ways; for example, by virtue of an archi-metaphor shared by many languages, the inchoativity of certain mental states (e.g. LOVE, DRUNKENNESS, MADNESS, FEAR) is conceptualized as an INWARDS MOVEMENT (Iñesta & Pamies 2003). The typology that Talmy applied to literal displacement should lead us to expect that Romance languages would also represent this imaginary “entry” with a simple directional verb (as it happens in sp. *entrar en trance* [*enter into trance] or *entrar en coma* [*enter into coma]), whereas Germanic and Slavic languages would always do it with a “concrete” verb linked to a directional particle (as it happens in pol. *zakochać się* *behind+love+oneself: “falling in love”) or rs. *нануться* (*on+drink+oneself: “get drunk”). But we do not find always this distribution either, since Romance languages also use preverbs (this, locative prepositions) to designate the beginning of certain mental states: sp. *enamorarse* (*in+love+oneself: “to fall in love”), *asustarse* (*at+fear+oneself “to get frightened), *entristecerse* (*in+sad+oneself: “to become sad”). If movements are figurative, nothing prevents two language from using inverse trajectories to express the same idea. E.g. “to get crazy” is *entering* madness in Spanish, and *escaping* out from good sense in Russian: sp. *enloquecerse* (*in+crazy+V.suf.) = rs. *сходить с ума* (*exit from spirit), therefore, in order to represent the beginning of the new mental state, the directionality of the path is not always inwards.

Comparing between languages the expressions meaning “get crazy” by means of figurative motion metaphors, we see that the metaphoric model of each unit is obviously unpredictable, either from one action to another, or from one linguistic family to another. For inchoative SLEEPING, we find a spatial preverb in French (*s’endormir* *oneself+in+sleep: “to fall asleep”) but not in Spanish, where there is a collocation (*quedarse dormido*: *remain slept). The opposite happens with LOVE and MADNESS, where Spanish uses a locative preverb (*enamorarse*: *in+love+oneself), whereas French uses a collocation (*tomber amoureux*: *fall lover), similar to eng. *to fall in love*. If enlarging the number of languages (table 1), we can see that, when the displacement is only figurative, the linguistic distribution is completely different from what corresponded to Talmy’s dichotomy for the “real” movement (tables 2 to 6).

Table 2. Love

LOVE	
INCHOATIVITY AS A METAPHORICAL TRAJECT BY MEANS OF A SYNTHETIC VERB, COLLOCATION OR IDIOM	INCHOATIVITY AS A METAPHORICAL TRAJECT BY MEANS OF A SATELLITE (PREVERB OR POSTVERB)
<p>eng. <i>to fall in love</i> fr. <i>tomber amoureux</i> (*to fall lover)</p>	<p>grm. <i>sich verlieben</i>; nl. <i>verliefd worden</i> sp. <i>enamorarse</i>; pt. <i>se apaixonar</i>; it. <i>innamorarsi</i>; cat. <i>enamorar-se</i>; rmn. <i>se îndrăgostească</i> pol. <i>pokachania / zakochać się</i> cz. <i>se zamilovat</i>; slk. <i>sa zamilovat'</i> cro. <i>se zaljubiti</i>; sb. <i>се залуби</i>; sln. <i>se zaljubiti</i> bul. <i>се влюби</i>; mcd. <i>се влюби</i> rs. <i>влюбитъся</i>; ukr. <i>закохуватися</i> blrs. <i>улюбляјуца</i></p>

Source: own research.

Table 3. Drunkenness

DRUNKENNESS	
INCHOATIVITY AS A METAPHORICAL TRAJECT BY MEANS OF A SYNTHETIC VERB, COLLOCATION OR IDIOM	INCHOATIVITY AS A METAPHORICAL TRAJECT BY MEANS OF A SATELLITE (PREVERB OR POSTVERB)
<p>eng. <i>get drunk</i> <i>-get hammered</i> <i>-get primed to the muzzle</i> grm. <i>sich zu betrinken</i> (*oneself to drink) nl. <i>zich bedrinken</i> (*oneself to drink) sp. <i>pillar una borrachera</i> (*to catch a drunkenness) <i>-pillar una castaña</i> (*to catch a chestnut); <i>-ponerse hasta el culo</i> (*to put oneself until the arse) <i>-ponerse ciego</i> (*to put oneself blind) <i>-ponerse hasta las orejas</i> (*to put oneself until the ears); it. <i>prendersi una scimmia</i> (*to catch a she-monkey) fr. <i>se bourrer</i> (*to stuff oneself) <i>-se bourrer la gueule</i> (*to stuff one's snout) <i>-charger la mule</i> (*to load the mule) svn. <i>se piti</i> (*to drink oneself) cro. <i>se piti</i> (*to drink oneself)</p>	<p>eng. <i>to inebriate</i> it. <i>inebriare</i> sp. <i>emborracharse</i> fr. <i>s'ennivorer</i> pt. <i>se embebedar / se inebriar</i> cat. <i>entrompar-se</i> rmn. <i>să se îmbete</i> pol. <i>upić się</i> cz. <i>se opít</i> svk. <i>sa opit'</i> rs. <i>нанитъся</i> ukr. <i>нанитися</i> blrs. <i>нанівајуца</i> bul. <i>се нание</i> mcd. <i>се опие</i> sb. <i>се наније</i></p>

Source: own research.

Table 3. Sleep

SLEEP	
INCHOATIVITY AS A METAPHORICAL TRAJECT BY MEANS OF A SYNTHETIC VERB, COLLOCATION OR IDIOM	INCHOATIVITY AS A METAPHORICAL TRAJECT BY MEANS OF A SATELLITE (PREVERB OR POSTVERB)
eng. <i>fall asleep/pass into sleep</i> nl. <i>in slaap te vallen</i> sp. <i>ponerse a dormir</i> (*to put oneself at sleeping) <i>-quedarse dormido</i> (*to remain slept) <i>-quedarse frito</i> (*to remain fried) fr. <i>se mettre à dormir</i> (*to put oneself at sleeping)	grm. <i>einschlafen</i> ; nl. <i>inlapsen</i> lat. <i>obdormiscere</i> fr. <i>s'endormir</i> ; it. <i>addormentarsi</i> pt. <i>adormecer</i> ; cat. <i>adormir-se</i> rmn. <i>să adoarmă</i> pol. <i>zasnąć</i> ; blrs. <i>заснути</i> rs. <i>уснути/заснути</i> ; ukr. <i>заснути</i> cz. <i>usnout</i> ; slk. <i>zaspaf</i> sln. <i>zaspati</i> ; cro. <i>zaspati</i> ; sb. <i>засну</i> bul. <i>засну</i> ; mcd. <i>заснива</i>

Source: own research.

Table 4. Madness

MADNESS	
INCHOATIVITY AS A METAPHORICAL TRAJECT BY MEANS OF A SYNTHETIC VERB, COLLOCATION OR IDIOM	INCHOATIVITY AS A METAPHORICAL TRAJECT BY MEANS OF A SATELLITE (PREVERB OR POSTVERB)
eng. <i>to get crazy</i> <i>-to go nuts</i> grm. <i>verrückt werden</i> ; (*mad become) <i>-verrückt geworden sein</i> (*mad become be) sp. <i>volverse majara</i> (*to become nutty) <i>-perder la cabeza</i> (*to lose one's head) fr. <i>perdre la tête</i> (*to lose one's head) <i>-perdre l'esprit</i> (*to lose one's spirit) <i>-péter les plombs</i> (*to blow the fuses) rs. <i>сходитъ с ума</i> (*to exit from spirit) <i>-помешаться умом</i> (*to agitate spirit); <i>-помешаться в рассудке</i> (*to agitate reason) <i>-тронутый умом</i> (*to be touched in spirit) blrs. <i>сысци з розуму</i> (*to exit from spirit) ukr. <i>зійти з розуму</i> (*to exit from spirit) sln. <i>postati nori</i> (*to become crazy)	sp. <i>enloquecer</i> pt. <i>endoidecer</i> it. <i>impazzire</i> cat. <i>embogir</i> rmn. <i>se înnebunească</i> pol. <i>zwarłować</i> cz. <i>se zbláznit</i> slk. <i>sa zbláznit'</i> cro. <i>poludjeti</i> sb. <i>полуди</i> bul. <i>да се побърка</i>

Source: own research.

Besides, the same language can use both systems to mark the inchoativity of FEAR conceptualized as a DISPLACEMENT: either *verb-framed* (sp. *entrar miedo [a alguien]* *fear enters [into sb.]) or *satellite-framed* (sp. *asustarse* *at+fear+self). Comparing compound verbs with idioms, we verify again that the abstract directionality of the metaphor can be inverted. For example, between eng. *to be invaded by fear* and sln. *biti iz sebe od strahu* (*to be out of one's mind by fear)⁶. The same "directional paradox" also affects preverbs, mixing the INWARDS path (sp. *intimidarse* / it. *intimorirsi*) and the OUTWARDS path (pol. *wystraszysz się* / rs. *ucnyzamься* / eng. *freak out*) though expressing the same idea.

Table 5. Fear

FEAR	
INCHOATIVITY AS A METAPHORICAL TRAJECT BY MEANS OF A SYNTHETIC VERB, COLLOCATION OR IDIOM	INCHOATIVITY AS A METAPHORICAL TRAJECT BY MEANS OF A SATELLITE (PREVERB OR POSTVERB)
eng. <i>to be invaded by fear</i> <i>-to be full of fear</i> sp. <i>entrarle miedo</i> (*fear enters [into somebody]) <i>-estar lleno de miedo</i> (*to be full of fear) fr. <i>être envahi par la peur</i> (*to be invaded by fear) it. <i>riempirsi di spavento</i> (*to fill+oneself with fear) pt. <i>ser invadido pelo medo</i> (*to be invaded by fear) rmn. <i>a se umple de spaimă</i> (*to fill oneself of fear) pol. <i>strach go obleciał</i> (*fear flew around him) rs. <i>напал страх [на кого-либо]</i> (*fear fell [on somebody]) <i>-захвачен страхом</i> (*conquered/occupied by fear) cz. <i>dostat strach</i> (*fear enters [into sb]) sln. <i>biti iz sebe od strahu</i> (*to be out+of oneself by fear)	eng. <i>freak out</i> / <i>flip out</i> grm. <i>sich erschrecken</i> sp. <i>asustarse</i> <i>-atemorizarse</i> <i>-intimidarse</i> fr. <i>s'apeurer</i> pt. <i>se assustar</i> it. <i>intimorirsi</i> cat. <i>atemorir-se</i> <i>-acovardir-se</i> rmn. <i>a se înspăimânta</i> <i>-a se intimida</i> pol. <i>wystraszysz się</i> cz. <i>vystrašit se</i> sln. <i>prestraši se</i> rs. <i>ucnyzamься</i>

Source: own research.

⁶ The Slovenian particle *pre* is not a pre-verb but a prefix, since it does not exist by itself as a word, and its function is just to mark perfectivity. However, it represents metaphorically an "end-to-end" traject inside a limited space, as it can be deduced from the meaning of literal movement verbs: *preplavati* "to swim from one point to another", "to bike from one point to another", *preteči* "to run from one point to another".

Phraseology is characterized by representing “imaginary” displacements to express other kinds of event, and, at the same time, by representing “real” displacements as if they were something else, by virtue of the bidirectionality between source domain and target domain (cf. Pamies 2014a; Pamies; Craig & Ghalayini 2014). Idioms are semantically indivisible per definition (Čermák 1998; 2007), so, when expressing “real” movements, they neutralize the *verb-sallelite* opposition, by simultaneously expressing together the abstract directionality (“outwards”, “inwards”, etc.) and the concrete description of movement (“running”, “flying”, “riding”, “swimming”, etc.):

Table 6. “RUNNING OUT”

“RUNNING OUT” EXPRESSED BY IDIOMS
<p>eng. <i>to go hell for leather; take to one’s heels; to show a clear pair of heels; make tracks; to cut and run; to take the midnight express;</i></p> <p>grm. <i>die Beine in die Hand nehmen</i> (*to carry one’s legs in one’s hands); <i>die Hufe schwingen</i> (“swing one’s hoofs”); <i>Fersengeld geben</i> (*give money to the heels); <i>den Adler machen</i> (*to make the eagle); <i>die Flattern machen</i> (*to make the flapping); <i>das Weite suchen</i> (*to look for space); <i>den Sittich machen</i> (*to make the parrot); <i>einen langen Schuh machen</i> (*to make a long shoe); <i>sich aus dem Staub machen</i> (*to make oneself from the dust);</p> <p>sp. <i>poner pies en polvorosa</i> (*put [one’s] feet in dusty); <i>tomar las [calzas] de Villadiego</i> (*to take Villadiego’s [hose]); <i>darse el bote</i> (*to give+oneself the jump); <i>darse a la fuga</i> (*to give+oneself to the fleeing);</p> <p>fr. <i>prendre ses jambes à son cou</i> (*to take one’s legs around one’s neck); <i>prendre la poudre d’escampette</i> (*take the clear+up powder); <i>foutre le camp</i> (*fuck the camp); <i>se sauver à toutes jambes</i> (*to save oneself at all legs); <i>se faire la malle</i> (*to make [one’s] trunk); <i>se faire la belle</i> (*to make oneself the beautiful [one]); <i>plier bagage</i> (*to fold [one’s] luggage); <i>tirer ses grègues</i> (*to pull+up one’s hose); <i>prendre la clé des champs</i> (*take the key of+the fields); <i>se déguiser en courant d’air</i> (*to disguise oneself as an airstream);</p> <p>it. <i>darsela a gambe</i> (*to give it oneself to legs); <i>tagliare la corda</i> (*to cut the rope); <i>alzare i tacchi</i> (*to lift the heels); <i>mostrare le calcagne</i> (*to show the heels); <i>mettersi le gambe in spalla</i> (*to put one’s legs on one’s shoulder); <i>mettersi le ali ai piedi</i> (*to put+oneself wings to one’s feet); <i>filare a rotta di collo</i> (*to fray at breakneck); <i>prendere la direttissima</i> (*to take the very+straight); <i>darsi alla fuga</i> (*to give+oneself to the fleeing);</p> <p>rs. <i>не чуя ног под собой</i> (*not to feel paws under oneself); <i>взять ноги в руки</i> (*to carry [one’s] feet in hands); <i>миаться со всех ног</i> (*to gallop with all legs); <i>смотреть удачи</i> (*to put+away the rods); <i>одна нога здесь другая там</i> (*[be with] one leg here, the other+one there);</p> <p>pol. <i>wziąć nogi za pas</i> (*to take the legs to one’s belt); <i>dać nogę</i> (*to give leg); <i>dać drapaką</i> (*to give broom); <i>zmyć się</i> (*to wash oneself); <i>pokazać pięty</i> (*to show heels);</p> <p>cz. <i>vzít nohy na ramena</i> (*to carry [one’s] legs on shoulders); <i>prásknout do bot</i> (*to burst even [one’s] boots); <i>vzít roha</i> (*to carry a horn).</p>

Source: own research.

Naturally, “simple” words with metaphorical meanings can also merge both kinds of information in a synthetic way. E.g., to mean “running away”, we find esp. *esfumarse* (*become smoke: “disappear”); fr. *détaler* (*pack one’s goods: “escape”); *déguerpir* (*abandon one’s possessions: “flee”), although, etymologically, these words were also analyzable as preceded by a locative preverb.

Of course, if, instead of verbal expressions, we observe adverbial idioms, the types will coincide with Talmy’s classification, because adverbial idioms express only the concrete form of the movement, not including the path, that still needs a satellite in Germanic and Slavonic languages, whereas it is embedded in the verbal meaning in Romance languages.

Table 7. Verbs + adverbial idioms

[DIRECTIVE VERB] + ADVERBIAL IDIOM	[CONCRETE VERB & SATELLITE] + ADVERBIAL IDIOM
sp. <i>Pablo salió de la escuela a toda hostia</i> fr. <i>Paul est sorti de l'école à toutes jambes</i> (*Paul is exited from the+school at all legs); pt. <i>Paulo saiu da escola pé na tábua</i> (*Paul exited from school foot on áiãðä); it. <i>Paolo è uscito dalla scuola a gambe levate</i> (*Paul is exited from+the school at raised legs); cat. <i>en Pau ha sortit de l'escola a corre-cuïta</i> (*Paul has exited from the+school at run-boiled).	eng. <i>Paul ran out of the school at break-neck speed</i> grm. <i>Paul rannte aus der Schule wie aus der Pistole geschossen</i> (*Paul ran out+of the school as shot by a pistol) rs. <i>Павел выбежал из школы сломя голову</i> (*Paul out+ran from school breaking head); pol. <i>Paweł wybiegł ze szkoły jak błyskawica</i> (*Paul out+ran from school like lightning).

Source: own research.

6. Provisional conclusions

The typological opposition between the *verb-framed* and *satellite-framed* languages is more quantitative than qualitative, since the same models are in fact available in Slavic, Romance and Germanic languages. Only their proportions and frequencies differ in each family. Besides, the selection between both kinds of mechanisms completely changes when the movement is metaphorical. The porosity of the boundaries between these categories is related with the impossibility of exclusion between the syntactic and lexical

domains, because of the in-between of phraseology, which was traditionally ignored. As John Sinclair (2008: 407) observed:

Phrases have never had a proper status in linguistic theory, and, as a consequence, are *anomalous* in descriptions. The reason for their omission from theory is that received theories require grammar and lexis to be separated from the outset, and they are then described without reference to each other.

As far as verbal predicates are concerned, phraseology occupies a large and central space in the lexico-grammatical continuum, including different areas, some of them, traditionally attributed to syntactic structures (light verb collocations and phrasal verbs) or to lexicon (compound verbs). Given that these constructions fulfill all the defining features of the phraseological unit (multi-lexemic sequences with different degrees of fixedness and lexicalization), some metalinguistic boundaries must be displaced, if not blurred.

Literature

- Anastasio, S. 2014 "Diversité dans l'expression de l'espace dynamique en L1 : étude contrastive entre italien, français et anglais". In : *COLDDOC 2014 Diversité des langues : Les universaux linguistiques à l'épreuve des faits de langue*. Paris : CNRS / Univ. Paris-Ouest / Univ. Paris-Descartes : 106–116 (<http://coldoc2014.free.fr>).
- Artusi, A. 2016. "Estudio traductológico de los verbos sintagmáticos del italiano al español. El caso de *Lessico familiare*". In: Calvo Rigual, C. & Spinolo, N. (eds.) *La traducción de la oralidad / Translating orality*. Alicante: Universidades de Valencia, Alicante y Castellón: 77–102.
- Bader, F. 1997. "Prépositions dans les langues indo-européennes anciennes". *Faits de langue*, 9 (Mars): 49–60 [rééd. http://www.persee.fr/doc/flang.1244-5460_1997_num_5_9_1140].
- Bayley, Ch.G. 1997. *The etymology of Old High German weak verbs* (I). Doctoral thesis. Newcastle: University.
- Biber D., Conrad, S. Leech, G. 2002. *Longman Grammar of Spoken and Written English*. Harlow: Pearson education Limited.
- Bolinger, D. 1971. *The Phrasal Verb in English*, Cambridge (MA): Harvard University Press.
- Buridant, C. 1995. "Les préverbes en ancien français". In: Rousseau, A. (ed.) *Les préverbes dans les langues d'Europe. Introduction à l'étude de la préverbation*. Lille: Presses du Septentrion: 287–323.
- Calvo Rigual, C. 2008. "I verbi sintagmatici italiani, con appunti contrastivi con lo spagnolo e il catalano". In: González Royo, C. & Mogorrón, P. (eds.) *Estudios y análisis de fraseología contrastiva*. Alicante: Universidad: 47–66.

- Čermák, F. 1998. "La identificación de las expresiones idiomáticas". In: Luque Durán, J.d.D. & Pamies Bertrán, A. (eds.): *Léxico y fraseología*. Granada: Método: 1–18.
- Čermák, F. 2007. *Fraseologie a idiomatika česká a obecná: Czech and General Phraseology*. Praha: Karolinum.
- Courtney, R. 1983. *Longman Dictionary of Phrasal Verbs*. Essex: Longman Group Limited.
- Cygal-Krupa, Z. 1995. "Les préverbes en polonais". In: Rousseau, A. (ed.) *Les préverbes dans les langues d'Europe. Introduction à l'étude de la préverbation*. Lille: Presses du Septentrion: 269–283.
- Dal, G. 2003. "Productivité morphologique : définitions et notions connexes". *Langue française*, 140 : pp. 3–23 (http://www.persee.fr/doc/lfr_0023-8368_2003_num_140_1_1063).
- Fougeron, I. 1995. "Préfixe et aspect en russe contemporain". In: Rousseau, A. (ed.), *Les préverbes dans les langues d'Europe. Introduction à l'étude de la préverbation*. Lille: Presses du Septentrion: 224–267.
- Fraser, B. 1976. *The Verb-Particle Combination in English*, New York: Academic Press.
- García Vega M. 2011. "Transitive phrasal verbs with the particle out: A lexicon-grammar analysis". *Southern Journal of Linguistics*, 35/1: 75–110 (hal-01369825).
- Iacobini, C. 2009. "Phrasal verbs: between syntax and lexicon". *Italian Journal of Linguistics*, XXI: 97–117.
- Iñesta, E.M. & Pamies, A.: 2002. *Fraseología y metáfora*. Granada: Método.
- Kopecka A. 2004. *Étude typologique de l'expression de l'espace: localisation et déplacement en français et en polonais*. Thèse doctorale. Université Lumière-Lyon2.
- Kunin, A.V. 1996. *Курс современной английской фразеологии*. Дубна: Феникс+.
- Luque Durán, J.d.D. 1972. *Las estructuras preposicionales en español y en inglés*. Tesis doctoral (inérita). Universidad de Granada.
- Martinet, A. 1960. *Éléments de linguistique générale*, Paris: Colin, [rééd. 1967].
- Masini, F. 2006. "Diacronia dei verbi sintagmatici in italiano". *Archivio Glottologico Italiano*, XCI/1: 67–105.
- Mejri, S. 2012. "Délimitation des unités phraséologiques". In: Ortiz Álvarez, M.L. & Huelva Unternbaümen, E. (eds.), *Uma [re]visão da teoria e da pesquisa fraseológicas*. Campinas: Pontes: 139–156.
- Mel'čuk, I. 2003. "Collocations : définition, rôle et utilité". In: Grossmann, F. & Tutin, A. (eds.), *Les collocations: analyse et traitement*. Travaux et recherches en linguistique appliquée. Amsterdam: de Werelt: 23–31.
- Michot, M-E.; Goldschmitt, S. & Pierrard, M. 2015. "Il saute dehors / Il grimpe dessus : particules adverbiales de trajectoire en français L1 et L2". *Pratiques: Linguistique, littérature, didactique*, 167–168 [<http://pratiques.revues.org/2773>] (DOI: 10.4000/pratiques.2773).

- Mudraya, O.; Piao, S.S.L., Löfberg, L.; Rayson, P. & Archer, D. 2005. "English-Russian-Finnish Cross-Language Comparison of Phrasal Verb Translation Equivalents". In: Cosme, C., Gouverneur, C., Meunier, F. & Paquot, M. (eds.) *La phraséologie dans tous ses états*. Louvain-la-Neuve: Université Catholique: 277–283.
- Mudraya, O., Piao, S.S.L., Rayson, P., Sharoff, S., Babych, B. & Löfberg, L. 2008. "Automatic extraction of translation equivalents of phrasal verbs and light verbs in English and Russian". In: Granger, S. & Meunier, F. (eds.), *Phraseology: an interdisciplinary perspective*. Amsterdam: John Benjamins: 293–312.
- Pamies, A. 2007. "De la idiomaticidad y sus paradojas". In: Conde, G. (ed.): *Nouveaux apports à l'étude des expressions figées*. Cortil-Wodon: InterCommunications: 173–204.
- Pamies, A. 2014a. "Provérbios fitonímicos e plantas proverbiais". In: S. Silva (ed.) *Fraseologia & Cia: entabulando diálogos reflexivos*, vol II. Campinas (Brasil): Pontes: 79–104.
- Pamies, A. 2014b. "A metáfora gramatical e as fronteiras (internas e externas) da fraseologia", *Revista de Letras*, 33/1 (Fortaleza: UFC), pp. 51–77.
- Pamies, A. 2016. "Metafora grammaticale e metafora lessicale: implicazioni teoriche per la fraseologia". In: Dal Maso, E. & Navarro, C. (ed.) *Gutta cavat lapidem. Indagini fraseologiche e paremiologiche*. Mantova: Universitas Studiorum: 87–120.
- Pamies, A. 2017. "Grammatical metaphor and functional idiomaticity". *Yearbook of Phraseology*, 8(1): 69–104.
- Pamies, A. 2018a. "Les concepts d'unité et de construction en phraséologie". In: Soutet, O., Mejri, S. & Sfar, I. (eds.): *La phraséologie: Théories et applications*. Paris: Champion: 59–79.
- Pamies 2018b. "El verbo sintagmático en las lenguas románicas". In: Briz, A. et al. (eds.), *Estudios lingüísticos en homenaje a Emilio Ridruejo*. Valencia: Universidad. Vol. II: 1057–1070.
- Pamies, A. & Pazos, J.M. 2018. "Acerca del "phrasal verb" en español y alemán". *XIII Congreso Internacional de Lingüística General*. Universidad de Vigo: 13–15 junio 2018 [texto en prensa].
- Quirk, R. & S. Greenbaum 2000. *A University Grammar of English*. Singapore: Longman.
- Riguel, É. 2014 "Les phrasal verbs dans les productions écrites des apprenants de l'anglais langue seconde". In: *COLDOC 2014 Diversité des langues : Les universaux linguistiques à l'épreuve des faits de langue*. Paris : CNRS / Univ. Paris-Ouest / Univ. Paris-Descartes : 117–130 (<http://coldoc2014.free.fr>).
- Rousseau, A. (ed.) 1995. *Les préverbes dans les langues d'Europe. Introduction à l'étude de la préverbatton*. Lille: Presses du Septentrion.
- Russo, Rosa 2014 "Les verbes de mouvement italiens : stratégies de lexicalisation des natifs". In : *COLDOC 2014 Diversité des langues : Les universaux linguistiques à l'épreuve des faits de langue*. Paris : CNRS / Univ. Paris-Ouest / Univ. Paris-Descartes : 95–105 (<http://coldoc2014.free.fr>).
- Simone, R. 1996. "Esistono verbi sintagmatici in italiano?". *Cuadernos de Filología Italiana*, 3: 47–61.

- Talmy, L. 1985. "Lexicalization patterns: semantic structure in lexical forms". In: Shopen, T. (ed.): *Language typology and syntactic description, Volume III, Grammatical categories and the lexicon*, Cambridge, Cambridge University Press: 57–149.
- Talmy, L. 1991. "Path to realization: A typology of Event Conflation". In: Sutton, L.A., Johnson C. & Shields, R. (eds.), *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society*: 480–510.
- Žaucer, R. 2002. *The Role of Verbal Prefixes in Slavic: Evidence from Slovenian Locative Denominal Verbs*. Master Degree Thesis. University of Ottawa.

Les verbes composés, l'idiomaticité et continuum de figement

Résumé

Leonard Talmy assigne les *phrasal verbs* à un groupe typologique qu'il appelle *langues à satellite*, dont le prototype seraient les langues germaniques, cependant cette construction existe aussi dans les langues romanes et slaves (en moindre quantité). Par ailleurs, ces structures ont beaucoup plus de traits en commun avec les phrasèmes qu'avec les combinaisons syntaxiques, et la fonction jouée par ces verbes est assez similaire à celle d'un nombre proportionnel de verbes préfixés romans et slaves. Ce travail analyse les deux mécanismes dans un continuum de constructions prédicatives avec différents degrés de figement et d'idiomaticité, allant des combinaisons syntaxiques régulières aux verbes complètement lexicalisés, avec une zone intermédiaire occupée par plusieurs structures idiomatiques: collocations à verbe-support, locutions verbales, verbes syntagmatiques séparables, verbes syntagmatiques inséparables et verbes composés.

Iris Vogel

Universität Hamburg

Germany

<https://orcid.org/0000-0003-2066-3840>

Representing Gestural Idioms in the Lexicon¹

Abstract. Gestural idioms are verbal expressions based on conventionalized body movements and can therefore be regarded as the interface between nonverbal and verbal communication. This article examines the challenges of encoding gestural idioms in lexicography. An analysis of Japanese dictionaries yields patterns for the comprehensive and manageable description of gestural idioms not only in dictionaries but for language learning in general.

Key words: *gestures, lexicography, nonverbal communication, phraseology, Japanese*

1. Introduction

A number of somatic idioms derive their meaning from body language or, more specifically, from gestures. Through linguistic encoding, these nonverbal conventions are transformed into phraseological units, which reflect the metaphorical meaning of the gesture to a varying degree. Conventions evolve over time in a group; therefore the ability to interpret another person's body language is greatly dependent on the observer's cultural background, and cannot easily be transferred to a different culture. Typical examples would be shaking of the head or hands, which can be regarded as universally understood, at least in the cultures of Western Europe and North America, whereas the Japanese gestures of tilting one's head (*kubi o hineru*) or placing one's hands together (*te o awaseru*) are quite culture-specific.

In some cases, the linguistic encoding of nonverbal behavior survives the change in social conventions and can therefore outlive the actual gesture or

¹ The paper was originally scheduled for publication in 2018 and it reflects the state of scientific discourse at that time.

body movement from which it was originally derived. These idioms tell tales of times past, for instance *ashi o arau* (lit. washing one's feet)²; in Japanese, it evokes the picture of a traveler washing his feet before entering the house, because the etymology is well known.

In order to elucidate ways to define gestural idioms, I analyzed the representations of a number of Japanese gestural idioms and their counterparts in mono- and bilingual dictionaries (Japanese-Japanese, Japanese-English). Examples in monolingual and specialized dictionaries illustrate that the comprehensive representation of gestural idioms is possible and manageable. Bilingual dictionaries, however, tend to be rather vague in their description of this phenomenon. The comparison of different dictionaries reveals different ways and patterns of describing the facets or layers of meaning. The translations given for the monolingual dictionary entries are quasi literal translations of the Japanese definitions. They provide the basis for discussing the differences between the paraphrases used in the monolingual dictionaries. They are not intended or suitable in any way to represent the idioms in the bilingual dictionary context. The goal of this work is to derive best practice patterns for the representation of gestural idioms in dictionaries and language learning material.

2. Semantic levels of gestural idioms

Gestural expressions are a special case with regard to their semantics within the field of phraseology, as Burger points out (Burger 1976: 316). The specific nature of these *kinograms*, as Burger calls them, lies in the double-layered semantic structure, constituted by the factual behavior and the meaning of the behavior (communicative value of the behavior). The co-occurrence of the gesture and the phrase or simultaneous activation of both readings in a text is Burger's condition for a phraseological unit to qualify as a phraseological *kinogram* (e.g. *to shake one's head*). The fact that both layers of meaning, factual and idiomatic, are active at the same time is the constitutive characteristic of this group of idioms. The general rule, that a dominant idiomatic meaning in a certain context tends to shift the literal meaning into latency, does not apply here. The reason is that the symbolic meaning of the gesture

² In the time when people in Japan traveled on foot, they usually washed their feet before entering a house or inn, leaving the dust and dirt of the journey behind. The Japanese idiom *ashi o arau* (lit. washing one's feet; meaning: *get out, leave, cut ties with*) originates from this obsolete tradition and is now only used in its idiomatic interpretation of someone leaving the dirty deeds of the past behind.

is superimposed over the literal meaning, which leads to the characteristic semantic double-layeredness.

Burger distinguishes four semiotic levels (Burger 2003: 61–62):

1. the factual behavior (the “form” of the nonverbal behavior),
2. the conventional interpretation of this behavior (the “meaning” of the nonverbal behavior),
3. the linguistic encoding (the “form” of the utterance),
4. The double-layered meaning of the linguistic utterance:
 - a) The depiction of the meaning of the factual behavior (= 1),
 - b) The depiction of the meaning of the nonverbal behavior (= 2).

3. Contrastive Research on Gestural Idioms

Hashimoto (1993) and Otsuka (1994) found in their research on Japanese L2 teaching that Japanese learners have difficulty with deducing the meaning of unknown idioms if they involve cultural differences and the learners are not familiar with similar conventionalized behavior (Hashimoto 1993, Otsuka 1994). In addition, the existence of partial equivalents in the mother tongue leads to an increase in mistakes or misinterpretations (Otsuka 1994: 56–57). They classify idioms by grounding their nonverbal component in reality, distinguishing idioms based on real, semi-real and fictional body movements, their co-occurrence with the verbal manifestation, and their figurative, emblematic, rhetorical or metaphorical nature. The following analysis is based on these features in combination with the semiotic levels. The goal is to pinpoint potential sources for these erroneous interpretations and outline patterns for a comprehensive representation of gestural idioms in the dictionary.

4. Gestural idioms in the dictionary

To better understand what descriptive methods are used to encode information specific to gestural idioms for language learners, I analyzed entries in electronic dictionaries with regard to their representation of semiotic levels. Because of the abundance of dictionaries on the Japanese market, this analysis focuses on a selection of common mono- and bilingual dictionaries. Preference was given to online versions of the dictionaries, if available. Special attention is given to the following questions: How can the double-layered meaning be represented in a differentiated manner? What methods can be used to represent the correlation between the factual and the figura-

tive meaning in the description of the idiom? What kind of information is needed in a dictionary entry describing the meaning and use of a gestural idiom in order to bridge the cultural gap?

In order to obtain a clearer picture of the descriptive methods used in the dictionaries, the examples are arranged in four distinct groups, depending on their specific semantic features: Type 1 idioms are based on real body movements and both layers are active simultaneously. Type 2 comprises semi-real nonverbal behavior, where both layers still appear to be active. Types 3 and 4 are what phraseological research usually calls *pure idioms*, in which the factual meaning and the idiomatic meaning are distinct from one another. Type 3 are ‘real idioms’, which are based on real nonverbal behavior but the factual and the idiomatic layers are never active at the same time, whereas in Type 4 the gesture or body movement is totally imaginary and the meaning purely metaphorical. The examples chosen for this analysis and the grouping method are greatly indebted to the work of Hashimoto (1994) and Otsuka (1993), but idioms that do not denote a gesture in the narrower sense³ or phrases whose meaning is strongly based on the metaphorical interpretation of one constituent⁴ are not included in this analysis.

4.1. Expressions based on real movements

The analysis starts with gestural idioms based on real body movements. Since the comprehensive verbal descriptions of the gesture itself given in specialized dictionaries tend to be quite lengthy (Lynn 2014: 1504–1506) and go far beyond the scope of any general dictionary, illustrative material would be helpful for learners, especially for the group of idioms in which the nonverbal communicative behavior and the phrasal expression are directly linked (Hashimoto 1998: 154). Traditionally, however, lexicography is reluctant to use graphic material and relies strongly on verbal descriptions, because of the limited available space. None of the following dictionary examples provides an illustration for the factual layer. In the cases where double-layered semantic information is given, paraphrases rely on verbal descriptions of the factual layer. In this respect, digital editions of dictionaries are no different from their printed versions. The methods of description for the different

³ The Japanese idiom *tsume ni hi o tomosu* (literally: to light up your finger nail) as an expression of extreme stinginess is totally imaginary, but not a bodily action in the narrow sense, at least for a human.

⁴ This analysis also excludes metonymy: e.g. the meaning ‘intelligent’ of *atama ga yoku kireru* (lit: the head cuts well) is strongly based on the association of *head (atama)* with its function (*thinking, intelligence*).

types of gestural idioms were subjected to the following analysis. The focus lies on the definition sections of semantic commentary in the dictionaries. Example phrases are only taken into account if they make a substantial contribution to the encoding of semantic features by providing an interpretation different from the definition(s) or are the sole description contained in the dictionary. The reading of the lemma itself is presented in italics before the entry and a rough English translation of the Japanese definitions is given in square brackets. Japanese explanations within the English part of dictionary entries are marked with italic letters.

The group of expressions denoting nonverbal behavior in the narrower sense can be further divided into three subgroups, depending on whether: 1a) the simultaneous performance of the gesture is required in order to understand the meaning, 1b) the simultaneous performance of the gesture is optional and 1c) the expression is only used separately from the performance of the gesture. The first example of Type 1a is *atama o kaku* (lit. scratch one's head):

- (1) *atama o kaku*: 失敗を自ら恥じたり照れたりするさま。[Feel ashamed due to one's own failure.] (KJE)
- (2) *atama o kaku*: 思わず頭に手をやって軽くかく。恥ずかしく思ったり、てれたりしたときのしぐさ。[Inadvertently move one's hand to the head and scratch lightly. Gesture when feeling embarrassed or ashamed.] (DJR)
- (3) *atama o kaku*: 恥ずかしく思ったりはにかんだりして、思わず頭に手をやって軽くかく。失敗を恥じたり、はにかんだりする。閉口する。[Feel embarrassed and shy and inadvertently scratch one's head. Feel embarrassed and shy because of a failure. Be at a nonplus.] (NKDJT)

Most of the monolingual dictionaries encode both layers of the gestural idiom *atama o kaku*, with the exception of the Kōjien (KJE), where only the figurative meaning is represented (1). With regard to the simultaneous and obligatory nature of the gesture, the definitions state this clearly with expressions like *no toki no shigusa* (lit. gesture when [...]) (2) or a conjunctive sentence structure (~te, [...]) (3).

Polysemy is a phenomenon that also exists apart from the semiotic layers, as the example of *te o awaseru* (lit. place the hands together) shows:

- (4) *te o awaseru*: 1 手のひらを合わせて神仏を拜む。転じて、心から物事を頼む。心から感謝する。[Place the hands together and pray to the gods or Buddha. Figuratively, to wish something from the bottom of one's heart. To thank someone from the bottom of one's heart.] 2 手合せをする [Compete with someone.] (KJE)

- (5) *te o awaseru*: 1 両方の手のひらを合わせる。感謝の気持ち、懇願の気持ちなどを表す。[Place the palms of both hands together; expressing the feeling of gratitude or supplication.] 2 拝む。合掌する。[Fold one's hands in prayer.] 3 相手として勝負する。手合わせをする。[Compete with someone; enter a competition.] (DJR)
- (6) *te o awaseru*: 1 両方のてのひらを合わせる。拝む。また、心をこめて頼む。[Place both palms of one's hands together; pray. Also to wish something from the bottom of one's heart.] 2 相手となって勝負をする。手合わせをする。[Compete against an opponent; enter a competition.] (DJS)

The secondary reading of *te o awaseru* (to compete; to enter a competition) is purely figurative, not motivated by a specific gesture or bodily movement and therefore not included in this analysis. For the first reading, which is based on the verbal encoding of nonverbal behavior, all the dictionary entries take account of both semiotic levels in their definition. The most obvious difference between the dictionaries is that the Daijirin (DJR) (5) treats the placing of the hands together in prayer as a distinct reading, indicating that the motivation of the gesture is regarded as a distinctive feature. Another notable difference is that while all the dictionaries closely correlate the gesture of placing the hands together with the emotion, suggesting that the gesture is an integral part of the expression, KJE (4) uses the expression *tenjite* (lit. figuratively). This typical way of marking disjoint semantic layers calls into question the obligatory nature of the body movement. With *~toki no shigusa* (lit. gesture when ~), DJR (2) states the co-occurrence of the gesture more clearly than DJR with *~kimochi o arawasu* (lit. expresses the feeling of ~) in the later example (5). In the bilingual dictionaries the factual layer is most prominent within the definitions of Type 1a gestural idioms:

- (7) *atama o kaku*: scratch one's head (考えるジェスチャー) [(gesture of thinking)] (KWEDJT4)
- (8) *atama o kaku*: 褒められた彼は、照れくささを、あたまをかいてごまかした。He scratched his head to hide his embarrassment at this praise. (KWEDJT5 within examples under the lemma *atama*)
- (9) *te o awaseru*: 1 (pray) (posture during prayer) place one's hands together; (posture of solicitude) fold [clasp] one's hands. 2 (take on as opponent) play against sb. (KWEDJT5)

The fourth edition of Kenkyūsha's New Japanese-English Dictionary (KWEDJT4) translates *atama o kaku* with *scratch one's head* and supplies an interpretation of the semantic layer in Japanese explaining the conventionalized meaning of the English gesture and accordingly the English phrase (7).

It shows that within an English (or German⁵) context one typically associates this gesture with thinking or pondering over something; in Japanese culture, however, it is strongly associated with embarrassment. The choice of metalanguage and the fact that the common Japanese interpretation is not mentioned shows that this additional information is directed at a Japanese readership. The interpretation of embarrassment only surfaces in an example sentence in the fifth edition of Kenkyūsha's New Japanese-English Dictionary (KWEDJT5) (8).

The second subgroup (Type 1b) consists of gestural idioms based on real gestures, for which the performance of the corresponding motion is optional. There is no mention of the polysemic nature of *shita o dasu* (lit. stick out one's tongue) in Hashimoto (1993), but all of the monolingual dictionaries acknowledge it. For the secondary reading, three (12–14) of the four dictionaries refer to the fact that the gesture is not optional, therefore this facet of interpretation belongs to Type 1a rather than 1b.

- (10) *shita o dasu*: 1 陰でそしたり、ばかにしたりするさま。[Libel someone or make fun of someone behind their back.] 2 自分の失敗を恥じたり、てれかくしをしたりするさま。[Be embarrassed because of one's own failure and feel ashamed.] (KJE)
- (11) *shita o dasu*: 1 陰でばかにする。心の中であざわらう。[To secretly make fun of someone. To inwardly sneer at someone.] 2 自分の失敗を恥じたり、ごまかしたりするしぐさにいう。[Describes a gesture used to gloss over the shame caused by a failure.] (DJR)
- (12) *shita o dasu*: 1 陰で人をばかにしたり、あざけり笑ったりする。また、そういうときの動作。[Make fun of a person and laugh at them behind their back. Also the motion on such occasions.] 2 恥ずかしさなどをごまかす動作。[Manner of glossing over embarrassment.] (DJS)
- (13) *shita o dasu*: 1 陰でそしたり、嘲（あざけ）ったりするさまを表わすしぐさ。[Gesture expressing secret criticism or when ridiculing someone behind their back.] 2 自分の失敗を恥じたりてれたりするさまを表わす動作。[Manner of expressing one's shame and embarrassment at a failure.] (NKDJT)

⁵ In German the meaning of the gesture *sich am Kopf kratzen* (to scratch one's head) is not as conventionalized as in Japanese, for example, and is often accompanied by a semantic interpretation. This interweaving of the interpretation of the gesture within the idiom is quite a common phenomenon in German (Vogel 2016: 26).

With regard to their figurative (primary) meaning (the motion of sticking out the tongue), the definitions in the monolingual dictionaries are quite similar. There are differences, however, in the way they represent the interrelation of the semiotic levels: In KJE (10) there is no mention of the factual interpretation, whereas in *Nihon Kokugo Daijiten* (NKDJT) (13) the primary and secondary readings are phrased in a similar way, mentioning the motion as an integral part of the interpretation. The expression *kokoro no naka* (inwardly) in *Daijirin* (DJR) (11) could be interpreted as a hint suggesting that the co-occurrence of the motion is facultative, but only DJS (12) clearly states the optionality of the factual layer.

The bilingual KWEDJT5 also distinguishes the two readings of *shita o dasu* (14) but with regard to the Type 1b reading, the figurative use (optionality of the gesture) is conveyed by the example phrase, where *shita o dasu* is translated by ‘laughs at her behind her back’ or disrespect in general.

- (14) *shita o dasu* stick out one’s tongue in embarrassment [shame];
 〈陰であざけて〉 stick out one’s tongue ((at sb)); express contempt for sb. ⇒ あいつは先生の前で神妙にしているが、陰では～を出している。He is all docility in front of the teacher, but he laughs at her behind her back [really doesn’t respect her at all]. [KWEDJT5]

Finally, Type 1c also comprises gestural idioms based on real body movements, with the difference that the physical performance typically does not co-occur with the verbal use of the phrase. One example is *ago o dasu* (lit. stick one’s chin out), used metaphorically for being exhausted.

- (15) *ago o dasu*: ひどく疲れる。[Be very tired.] (KJE)
- (16) *ago o dasu*: ひどく疲れて、足が動かず、あごだけが前に出る。疲れ切ってしまうにもならない状態をたとえていう。[To feel exhausted to such an extent that the feet don’t move and only the chin sticks out. Used metaphorically for being exhausted to such a degree that nothing can be done.] (DJS)
- (17) *ago o dasu*: (長い間歩いて疲れると、腰がひけて顎が出る格好になるところから) 弱り果てる。疲れ切る。転じて、自分の手に負えないで困る。
 [(From the posture one assumes after being tired out from a long walk, when bending forward and sticking one’s chin out) Be exhausted. Be totally tired out. Figuratively, be troubled beyond one’s control.] (NKDJT)

Again, KJE does not mention the body motion, focusing only on the figurative meaning (15). DJS and NKDJT, however, acknowledge the fact that

the metaphorical interpretation is based on the posture somebody assumes when very tired after a physically demanding task like a long walk (16), (17). That this is purely a motivation, not an accompanying motion, is expressed by *tatoete iu* (used as an example/metaphorically for) (16) and *tenjite* (figuratively) (17) respectively.

The bilingual KWEDJT5 puts the sole focus on the communicative value (18):

(18) *ago o dasu*: get exhausted [tired out, worn out] (KWEDJT5)

Given that that there is no comparable figurative expression in the target language of the dictionary, it would be especially valuable for language learners to illuminate the background of the idiom in a similar way to some of the monolingual dictionaries (16), (17).

4.2. Type 2: Expressions based on semi-real body movements

Type 2 consists of idioms illustrating a posture, gesture or facial expression that represents a certain partly imaginary behavior or gesture. Examples of this type are *me o sankaku ni suru* (lit. to make one's eyes triangular) and *hana no shita o nagaku suru/nobasu*⁶:

(19) *me o sankaku ni suru*: 激怒するさまの形容。[Appearance when being infuriated.] (KJE)

(20) *me o sankaku ni suru*: 目を怒らす。怖い目つきをする。目に角(かど)を立てる。Give an angry look; glare at someone. *Me ni kado o tateru* (lit. to put up corners in the eyes).] (DJS)

(21) *me o sankaku ni suru*: 目に角(かど)を立てこわい目つきをする。目を怒らす。[*me ni kado o tate* (lit. to put up corners in the eyes) and look angry. Give an angry look.] (NKDJT)

The definitions of *me o sankaku ni suru* in the monolingual dictionaries DJS and NKDJT depend mostly on a synonymous idiom (*me ni kado o tateru*) and the causative construction *me o okorasu* (lit. to anger the eyes) (20), (21). DJS adds at least one non-figurative paraphrase *kowai metsuki o suru* (lit. give a frightening look) to the definitions, while KJE regards the appearance in general as essential, thereby expressing semi-factuality in a more generic fashion.

⁶ The idioms *hana no shita o nagaku suru* and *hana no shita o nobasu* are synonymous expressions and treated together in the dictionaries.

- (22) *me o sankaku ni suru*: give [shoot] an angry look «at...»; look daggers [menacingly] «at...» (KWEDJT5 under the lemma *sankaku*)
- (23) *me o sankaku ni suru*: have an angry look in one's eyes; look angrily «at...» (KWEDJT5 under the lemma *me*)

KWEDJT5 provides slightly different English equivalents in the entries for *sankaku* (22) and *me* (23) respectively. To *look daggers* in example (22) aligns quite neatly with the Japanese factual layer (*sankaku ni suru* meaning literally *make triangular*), and *have an angry look in one's eyes* in example (23) describes the expression of the eyes by means of its interpretation. The factual layer and the semi-realness of the facial expression do not need to be isolated here, because the expressions in the target language are quite similar.

The second example *hana no shita o nagaku suru* (lit. to lengthen the part below the nose (= between nose and mouth)) is much more culture-specific and therefore an interesting case for this analysis. Again KJE does not address the semi-real gesture (24) whereas DJS introduces the semi-real factual layer as a *kaotsuki* (facial expression) that refers to a certain look (26):

- (24) 女にあまい。女性に迷いやすい。鼻下長。[Have a soft spot for women. Be easily taken with feminine charms. Amorous sort of man.] (KJE)
- (25) 好色そうな顔つきをする。女性に甘いようすについていう。鼻の下を長くする。[Have a lustful look on one's face. Said about someone who has a soft spot for women. *Hana no shita o nagaku suru.*] (DJS)

Hashimoto regards these Type 2 expressions as a transitional area between reality and figurativeness (Hashimoto 1993: 144), but comparison of *me o sankaku ni suru* with *hanano shita o nagaku suru/hana no shita ga nagai* reveals a perceptible difference within this group. Examining the bilingual dictionaries, it appears that displaying anger by changing the appearance of the eyes is a more universal concept across cultural borders than expressing an amorous nature with the part between nose and upper lip. It is not surprising that the only gestural idiom given as equivalent in KWEDJT5 *have one's head turned* (26) is based on a different gesture, which is a movement rather than a facial expression.

- (26) [*hana*]no shita ga nagai: ~の下が長い be spoony [on] [over] a woman, be easily taken with [susceptible to] feminine charms, be an amorous sort ~の下を伸ばす[長くする have one's head turned [be struck silly] by the sight of a woman (KWEDJT5)⁷

⁷ Mentioned as a collocation within the entry of 鼻の下 (*hana-no shita*).

4.3. Type 3: Figurative expressions based on real body movements

Type 3 consists of the so-called pure idioms that are based on real gestures or motions, but where the performance of this gesture or motion is linked to the literal interpretation of the phrase. Only then is the factual layer active. The figurative interpretation of this gesture or motion is typically triggered by the verbal manifestation if the factual layer is inactive. To put it differently, the idiomatic use of the phrase depends on the absence of the gesture, which leads us to *agura o kaku* (lit. sit cross-legged) and *te o nobasu* (lit. reach out with the hand/arm):

- (27) *agura o kaku*: 1 両足を横にひろげ前に組んで、楽にすわる。[Sit comfort by spreading the legs to the sides and crossing them in the front.] 2 (既得の地位や権力をよりどころとして)自分では何もしないで、いい気な態度でいる。(Due to one's acquired status or authority) do nothing. Be complacent.] (KJE)
- (28) *agura o kaku*: 1 あぐらを組んで座る。[Sit cross-legged style.] 2 のんきにかまえて、何の努力もしないことのたとえ。[Metaphorical for a carefree attitude, without making any effort.] (DJR)
- (29) *agura o kaku*: 1 足を組んで楽な姿勢で座る。[Sit comfortably with the legs crossed.] 2 自分は何の努力もせず、あるものに頼ってゆうゆうと構えている。ずうずうしく構える。[Attitude of not making any effort oneself, but relying on what is there and taking it easy. Conduct oneself in an impudent manner.] (DJS)

Distinct readings represent the disjunctive nature of the semiotic layers in *agura o kaku* in all four⁸ monolingual dictionaries (27–29). Surprisingly, KWEDJT5 only mentions the phrase as a collocation of *agura* and does not give the figurative reading (30).

- (30) sit [squat (down)] cross-legged [tailor fashion, Indian fashion, Turkish fashion]; sit with one's legs crossed, sit in the lotus position (KWEDJT5)

The style of sitting down described by *agura o kaku* and *sit tailor fashion* is comparable between the languages, but the implications are slightly different. In Japanese *agura o kaku* nowadays has the implication of sitting in a relaxed fashion as opposed to *seiza* (lit. the proper way of sitting), for example during the tea ceremony. Sitting tailor fashion refers to the style previously used

⁸ The entry in NKDJT is very similar to the ones in KJE and DJS.

by tailors when working. Since it lacks the implication that there is a different, 'proper' way to sit, the mention of the second reading (complacency) should be encoded, as done in the monolingual dictionaries. The second example for disjoint semiotic layers is *te o nobasu* (lit. reach out with the hand):

- (31) *te o nobasu*: 取引先や仕事を広げる。手を広げる。[Expand the reach of one's dealings or business. *Te o hirogeru*.] (KJE)
- (32) *te o nobasu*: 今までしなかった事をやってみる。勢力をひろげる。手をひろげる。[To do something one has not done yet. Extend one's influence. *Te o hirogeru*.] (NKDJT)

There is no mention of the factual layer *reach out* in the monolingual dictionaries, possibly because of its compositional nature. In the bilingual KWEDJT5 dictionary, there is an apparent shift between editions: In the fourth edition of the dictionary both layers were represented together in one single entry, clearly marking the figurative use with *hiyuteki-ni* (figuratively)⁹. In the more recent fifth edition the editors moved the factual interpretation into the main entry of the first component *te*, while keeping a separate entry for the figurative meaning.

- (33) [*te*]o *nobasu*: ~を伸ばす stretch out (one's) hand; reach out ((to do, for sth)); put out one's hand; reach out ((to do, for sth)), put (out) one's [a] hand; reach after [across, over] ..., reach for ((a book)) (KWEDJT5 within the entry for the lemma *te*)
- (34) *te o nobasu*: 〈仕事などの範囲を広げる〉 become concerned in [with] ((a matter)) (KWEDJT5 as separate entry)

4.4. Gestural idioms based on imaginary body movements or functions

The section of gestural idioms for which there is no equivalent body motion or function known to us is etymologically the most interesting group. Literal meanings like *a hand reaches out from one's throat* or *to boil water with the belly button* are quite removed from any gesture in the world we know. Therefore this contra-factual literal meaning plays no part in the entries for *nodo kara te ga deru* in the monolingual dictionaries:

⁹ The entry for *te o nobasu* in KWEDJT4 mentions two distinct readings: 1. stretch (out) one's hand; reach out 《to》 [...]; 2. [比喩的に] concern oneself in [with] 《a matter》 (Vogel 2016: 18).

- (35) *nodo kara te ga deru*: 欲しくてたまらないたとえ。[Metaphorical for strongly yearning for something.] (KJE)
- (36) *nodo kara te ga deru*: 欲しいと思う気持ちが、抑えかねるほどであることのとえ。[Metaphorical for the feeling of wanting something so strongly that it is hard to suppress.] (DJR)
- (37) *nodo kara te ga deru*: ほしくてたまらないたとえにいう。[Metaphorical for wanting something to a degree that is hard to bear.] (NKDJT)

This truly strange image of a hand reaching out from someone's throat is clearly marked as a metaphor in all the monolingual dictionaries. The bilingual dictionary KWEDJT5 supplies the most common context of this phrase *nodo kara te ga deru* you na (as if ..) (38) and *nodo kara te ga deru* hodo hoshii (to an extent...), which provides clues for the contra-factual nature of the expression.

- (38) *nodo kara te ga deru*: ⇒ ~から手が出るような [*~kara te ga deru you na*] extremely tempting ((offer)); ((□)) mouthwatering / ~から手がでるほど欲しい [*~kara te ga deru hodo hoshii*] would desperately like; would absolutely love ((sth, to do)); covet; desire eagerly; feel like jumping ((at..)) (KWEDJT5)

The entries for *heso de cha o wakasu* (and its variations) are a good example of the absence of any encoding of the factual layer within the encoding of the semantic features of the idiom.

- (39) *heso de cha o wakasu*: おかしくてたまらないことにいう。[Said about something so ridiculous that it is hard to bear.] (KJE)
- (40) *heso de cha o wakasu*: おかしくてたまらないこと、また、ばかばかしくてしょうがないこと。多く、あざけていう場合に用いる。[Something unbelievably strange or absolutely absurd. Often used to ridicule someone/something.] (DJS)
- (41) *heso de cha o wakasu*: おかしくてたまらない。腹の皮がよじれる。また、ばかばかしくて話にならない。ちゃんちゃらおかしい。[Unbearably funny. To split one's sides. Also, too ridiculous to be worth considering.] (NKDJT)

KWEDJT5 does not even assign the idiom an entry but provides an example as a secondary layer of information in the semantic commentary (42).

- (42) *heso de cha o wakasu*: 彼は歌手志望だって?お〜で茶を沸かすよ。He wants to become a singer? What a joke! [Don't make me laugh!] (KWEDJT5)

5. Conclusions

Analyzing the gestural idioms with regard to the different semiotic layers (Burger 2003) and the facet of realness of the body movement (Otsuka 1994, Hashimoto 1993) provided a framework for showing how the aspects of embodiment and figurativeness can be encoded in definitions of dictionary entries. But it also revealed some shortcomings of the previous research for lexicography. Firstly, the classification into different types of gestural idioms needs to be done at the level of readings, since one lemma can belong to different groups, depending on the context. Polysemy can also be manifested at the level of the factual layer, which means that one verbal description can evoke different images, as seen in case of *te o awaseru* (lit. place the hands together).

Differences between the dictionaries show varying approaches: KJE and KWEDJT appear to target a mostly Japanese readership, as they focus strongly on the communicative interpretation and are not concerned with representing the factual layer, for instance by giving information on gestural elements accompanying the use of an idiom. This is very unfortunate, since gestures can support language learning in acquisition (Goldin-Meadow 2005: 253–254, Goldin-Meadow 2013).

In general, bilingual dictionaries add a different layer of complexity. Here the question of whether a figurative and a factual layer are encoded often depends on the inventory of the target language. If there is an idiom matching all aspects of the verbal and nonverbal communicative value, there seems to be no need to paraphrase the complex of verbal and nonverbal interpretation. Looking at the examples in detail, however, reveals that there is very little total equivalence between gestural idioms. Differences are manifested not only in the communicative value of the expression but also in the form of the gesture, as has been shown for *te o awaseru* (4–9). The complexity of an adequate verbal description of a gesture goes beyond the scope of a general dictionary, and equivalents in languages where the cultural background is different can be misleading. This might be another reason why KWEDJT5 mostly focuses on the figurative meaning, shying away from the challenge of adequately describing the factual layer. The complementary interpretations in different sections of the entry (like sub-lemma (33) versus separate entry (34)) and differing interpretations in the example sentences (14) are easily missed by the reader. The selective comparison between the previous fourth edition and the current fifth edition of KWEDJT suggests a shift towards more specifically targeting a Japanese readership.

Especially the monolingual dictionaries DJS, DJR and NKDJT contain many best practice examples for representing most aspects of the meaning of gestural idioms. For a more detailed analysis of the patterns expressing different semiotic aspects of the dictionaries, other factors such as the motivation of the metaphorical interpretation (emotion, religious or cultural practice, bodily function...) should be taken into consideration, since some examples have already shown that they are applicable in only very specific situations.

References

Dictionaries

- Daijisen*. 2012. Ikegami, Akihiko (ed.). 2nd ed. Tokyo: Shogakukan. (DJS)
- Daijirin*. 2006. Matsumura, Akira (ed.). 3rd ed. Tokyo: Sanseido. (DJR)
- Kenkyusha's new Japanese-English dictionary Kenkyūsha (Shin wa-ei dai-jiten)*. 1974. Masuda, Kō (ed.). 4th ed. Tokyo: Kenkyusha. (KWEDJT4)
- Kenkyusha's new Japanese-English dictionary Kenkyūsha (Shin wa-ei dai-jiten)* 2003. Watanabe, Toshirō; Skrzypczak, Edmund R. and Snowden, Paul (eds.). 5th ed. Tokyo: Kenkyusha. (KWEDJT5)
- Kojien*. 1998. Niimura, Izuru (ed.). 6th ed. Tokyo: Iwanami Shoten. (KJE)
- Nihon Kokugo Daijiten*. 2000. 2nd ed. 14 volumes. Tokyo: Shogakukan. (NKDJT)

Literature

- Burger, Harald. 1976. "Die Achseln zucken"- zur sprachlichen Kodierung nichtsprachlicher Kommunikation. *Wirkendes Wort*. 26: 311–334.
- Burger, Harald. 2003. *Phraseologie: eine Einführung am Beispiel des Deutschen*: Berlin: E. Schmidt.
- Burger, Harald. 2008. *Phraseologie: eine Einführung am Beispiel des Deutschen*: Berlin: E. Schmidt.
- Goldin-Meadow, Susan. 2005. The two faces of gesture: Language and thought. *Gesture*. 5: 241–257.
- Goldin-Meadow, Susan. 2013. How our gestures help us learn. In: Müller, Cienki, Fricke, Ladewig, McNeill and Bressemer (eds.). *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction*. Berlin, Boston: de Gruyter Mouton. 792–803. [HSK 38.1]
- Hashimoto, Hiroko. 1993. Teaching Japanese Body Language Phrases to JFL Students. *Journal of College of Japanese Studies*. 11: 141–159.

- Lynn, Ulrike. 2013. Gestures in dictionaries: Physical contact gestures. In: Müller, Cienki, Fricke, Ladewig, McNeill and Bressemer (eds.). *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction*. Berlin, Boston: de Gruyter Mouton. 1502–1511. [HSK 38.1]
- Otsuka, Yoko. 1994. Gestures and Body Part Phrases in Japanese. *Bulletin of Gifu College of Education*. 27: 51–65.
- Vogel, Iris. 2014. *Phraseographie Japanisch-Deutsch: Lexikographische Perspektiven bei der Entwicklung digitaler Nachschlagewerke*. Ph.D. diss., University of Tübingen. <http://hdl.handle.net/10900/52902>. Retrieved: 21.10.2016.
- Vogel, Iris. In press. Phraseologismen der Körpersprache im elektronischen Wörterbuch. In: Martina Ebi and Robert Horres (eds.): *Referate des 14. Deutschsprachigen Japanologentages*. Berlin: LIT Verlag. 17–40. <http://hdl.handle.net/11858/00-248C-0000-002D-E327-0>. Retrieved: 21.10.2016.

Leksykograficzny opis idiomów gestualnych

Streszczenie

Idiomy gestualne to jednostki frazeologiczne oparte na skonwencjonalizowanych ruchach ciała, które z tego powodu mogą być traktowane jako interfejs między komunikacją werbalną i niewerbalną. Niniejszy artykuł bada aspekty związane z leksykograficznym opisem tych idiomów. Analiza japońskich słowników pozwala na wskazanie modeli całościowego i zarazem praktycznego opisu idiomów gestualnych nie tylko w słownikach, ale i ogólnie na płaszczyźnie glottodydaktycznej.

PART II

**DYNAMICS OF PHRASEOLOGICAL
STOCK**

Tetiana Anokhina

KNLU

Ukraine

<https://orcid.org/0000-0002-8859-5568>

The Form Changes and Shifts of Meaning of the English MWUs

Abstract. The modern linguistics has a wide variety of approaches to describe the natural language meanings. In our research we rely on the corpus-based approach to offer a particular way of looking at Multi word units (MWUs) change, creation and rearrangement. The paper presents the insights into form changes and shifts of meaning of MWUs using the Corpus of Contemporary American English (COCA) friendly interface. The derived MWUs outlined in the paper are regarded as the secondary units where the actual meaning is changed, modified or erased. The operations of semantic deviation are described in the light of shift and rearrangement demonstrating mimesis when the initial meaning of MWUs can be partially modified or removed. The semantic shifts are viewed as the secondary and the post-secondary changes. In the process of semantic deviation the old meaning of the modified figurative expression is deleted and the brand new conceptual unit appears.

Key words: *MWUs, COCA, shift and rearrangement, neo, mimesis*

1. Introduction

In our research we use the procedure of analysis of MWUs (Multi words units) which is directed towards compiling the MWUs with understanding that these units may be new, unknown or partially known or unpredictably changed. The modified figurative expressions demonstrate mimesis, but the converted MWUs appear to be completely new conceptual units. What is the most important for our research and what we find interesting is compiling the registry of MWUs consisting of Phraseologic Units (PU) seen as Phraseologic N-grams: set expressions, free expressions, phraseologic units, neology units, shell words, other MWUs found in dictionaries and context. All the units denote the common idea that can be unknown, unique or lacunar thus needs extra clarification what can be done in compiling the registry of MWUs.

In this article we focus on the cognitive scenario “change in form and meaning” analyzing some types of MWUs in corpus data. Today, with the rapid development of the corpus-based studies and computational approaches along with the availability of huge electronic corpora has made possible to see the change of form and meaning in MWUs.

1.1. Insights into MWUs: multidisciplinary research

Reproducibility of the secondary and post-secondary rearrangement is the new linguistic area for research and analysis of the deviation of MWUs including modifications of form and meaning to be studied. The research problem is correlating with frame semantics, its cognitive scenarios that making the research area especially significant for cognitive semantics.

The linguists explore the idea of MWUs changes of form and meaning. The analysis has been done on the basis of corpus data. The interpreted data are used within the framework of cognitive semantics. The analysis has shown that MWUs are often used in modified form as corpus data illustrate and that MWUs change their form and meaning according to the cognitive scenario “change in form and meaning” from positive to negative, neutral to colloquial, bookish to neutral, etc. The core idea is often implicated from the previous form partially rearranged; the inner form being modified can be readable due to understood implications and the reconstructed previous empirical knowledge. MWUs are broadly interpreted as all phraseologic units (phraseologisms), non-fixed collocations (weakly idiomatic phrases), idiomatic phrases, all set phrases of a language including proverbs that can be found in corpora by linguistics including computational scholars (Colson 2017: 17). Other definitions illustrate the terminology growth related to MWUs, some of the most common terms being chunk, cliché, collocation, extended lexical unit, fixed expression, formulaic sequence, idiom, idiomatic expression, lexicalized phrase, multi-word unit, phraseme, phraseologism, phraseologic unit, phrasal lexical item, phrasal lexeme, prefabricated chunk, prefab (Hüning, M., & Schlücker, B. 2015: 450). Some of these terms are regarded synonyms and close terms according to definitions by different scholars, but for the most part meanings overlap only partially. In general all MWUs can be described as extended lexical units with different degrees of syntactic fixedness and semantic compositionality, fixed or free (idiom and collocation) (Hüning, M., & Schlücker, B. 2015: 450).

The recent studies explore corpus data, fixed expressions and phraseology (Colson 2017). As Masini states (2005: 145) MWUs are “lexical units larger than a word that can bear both idiomatic and compositional mean-

ings". Sprenger (2003: 4) finds that MWUs as lacunar fixed set-expressions can be defined as specific combinations of two or more words with an opaque meaning or a deficient syntactic structure. These MWUs can be often lacunar to nonnative speakers which can be overcome in the broad context of the corpus data.

MWUs are variably-based identified as phraseologic units, multiword lexical units, polylexical words, items characterized for fixedness, including psycholinguistic fixity (speakers conceiving it as a unit), structural fixity (with variations) and pragmatic fixity (Lopez 2015: 162).

The MWU being a part of the "*frozen phraseology*" correlate with cognitive endozone of lacunarity. The obvious characteristic of MWUs is indicating their part of speech. The lacunar taxons of MWUs corpus include several slots in the POS structure (parts of speech): fixed verbal idiom (e.g. *bite the bullet*), fixed frozen adverbial (e.g. *all at once*), fixed particle verbs (e.g. *stick out*), non-fixed complex nominal (e.g. *daycare center*). The modified proverb can be traced as changed MWUs: the archaic "*The suit does not make the man*" turned into modern "*Clothes do not make the man*" according to the cognitive scenario "change in form and meaning" due to changes in the socio-cultural context (Pintarić & Škifić 2006: 210).

1.2. MWUs as N-grams in COCA search

The idea of representing MWUs as bigrams (two-word units) as (1) *easy rider*, *New York*, *pay attention*, *sharp criticism* and three words units (trigram) as (2) *to be in a habit*, *take a bath*, *deliver a speech* and multi words units (MWUs) as (3) *add insult to injury*, *at the drop of a hat*, *back to the drawing board*, *barking up the wrong tree*, *beat about the bush* is popular in computational linguistics (Colson 2017: 19). Bigrams, trigrams and multigrams are easily taken from corpus environment as COCA or BNC and other available or self-made corpora and can be analyzed according to the frequency rate extraction from the corpus data to illustrate the rare lacunar usage or to indicate MWUs as popular collocations.

The search for MWUs in corpora is a modern tool in learning and teaching EFL. To know the word in context KWIC or searched for the English collocations of the base is searched (*habitude – habit*) and from its entry, the collocation expressing the terminative aspect of habit (i.e. *get out of a habit*) is retrieved (Heid 1994: 252).

The COCA Corpus gives possibility to access the authentic corpus data for analysis and comparison of the changed forms of MWUs. MWUs have changed as: the search entry *bird in the hand* gives the variation result (the

changed proverb): (1) “*A bird in the hand is worth two in the bush – Better a bird in the hand than AIDS in the ass*” (COCA 1991, ACAD, LatAmPop-Scult), the search entry *shaken not stirred* (the changed quotation) gives the variation result: (2) *classic shaken, not stirred martini – Diet Coke, shaken not stirred* (COCA 1990, FIC, Knopf A.: Jurassic Park), the search entry *eggs is eggs* gives the variation result (the changed metaphorical expression): (3) *as sure as eggs is eggs – And this, in his strained English: Eggs is eggs and pigs is pigs* (COCA 2015, FIC, IowaRev), the search entry gives the result of re-think (the rearranged verbal idiom): (4) *to kick the bucket – it’s easier to kick the bucket than to kick the habit, to shoot the breeze – telephoned at least once a month, usually to shoot the breeze* (COCA 2014, Mag, Science News); the search entry *make up* gives the variation result evident in the COCA context (polysemantic phrasal verbs) (*to make up* – (5) *The city of Los Angeles had to make up a funding gap in the project after the value of federal low-income housing tax* (COCA 2017, NEWS, Los Angeles Times).

Some MWUs demonstrate change in meaning depending on post prepositions as the light verb constructions *to have a look* gives variation results (*to have a look* – (1) *They called me to have a look at him, but I just smiled and waved* (COCA, 2016, FIC, The Antioch Review; *I’ll be there by about one-thirty to have a look around the area* (COCA, 2016, FIC, Bk: HardAsIce); the search entry *to wash car* gives the result of syntactic fixedness: *to wash car* – (2) *I wash car windows for all the jerks on their way to work* (COCA, 1998, FIC, scholastic).

The stereotyped similes may be changed and rearranged, for instance the search entry *as nice as* gives the result (1) *as nice as nice can be* (COCA, 2008, FIC, ContempFic), (2) *as nice as I can get* (COCA, 2007, SPOK, Fox_Susteren), (3) *as nice as I can put it* (COCA, 1993, SPOK, ABC_Special); the search entry *beg and [...]* with zero ending gives the result of the rearranged binomial expressions as (1) *beg and pray* (COCA, 2017, SPOK, NPR: Fresh Air), (2) *to beg and plead, shout and scream* (COCA, 2015 MAG, ChristToday), (3) *to beg and creak under bloated budgets* (COCA, 2008, MAG, AmSpect); t-Complex nominals vary in the context: *man about town* – (1) *a celebrity man about town* (COCA, 2015, SPOK, PBS: PBS Newshour), *man about town* – (2) *a true man about town* (COCA, 2015, NEWS, USA Today), *man about town* – (3) *sophisticated man about town* (COCA, 2009, MAG, AmericanSpectator), *man about town* – (4) *a single man about town* (COCA, 2007, MAG, Newsweek). MWUs in collocations can be fixed, with additional meanings added due to articles co-occurring in COCA (definite/indefinite/zero), for instance: *hard frost* – partially rearranged in before (1) *the first hard frost in October* (COCA, 1994, MAG, MotherEarth), *hard frost* – (2) *after hard frost* (COCA, 1995, MAG, MotherEarth), *hard frost* – (3) *the*

onions for storage any time after they lie down and before a hard frost (COCA, 1996, MAG, MotherEarth).

MWUs stay fixed in fossilized or frozen forms (*all of a sudden* – *All of a sudden, he's in a rush* (COCA, 2016, FIC, Bk: ConfessorsClub) and routine formulas unchanged as *Good morning* demonstrate high frequency rate (*all of a sudden* gives 7324 results and *Good morning* – 21803). It is possible to find non-fixed MWUs in the broad context of COCA, in the manner of making a query: for instance, *budge*.[v*] gives different verb forms as: *let alone budge, will barely budge at all, will not budge on these issues*, etc. (COCA 1997, 2011, SPOK, PBS Newshour).

1.3. The Theory of Semantic Change: deletions or coining a neo

MWUs transfer into other languages can be implemented by deletions cutting off the unknown or euphemized element causing the informative lacunae, by sense substitutions or formal form shifts forming the partial cognitive or formal lacunae, by innovations eliminating lacunae by creating new word unit, by commentary and other interpretations filling in lacunae by the accompanied explanation (Szerszunowicz 2013: 207).

When the initial meaning can be modified or partially removed, thus the secondary meaning arising out, so we deal with semantic shift. If the form of the word (nominative unit) or multi-word units (communicative units) change, the lexical and semantic changes occur. The changes took place according to the principles of the Theory of Semantic Change (Lopez 2015: 167). As Lopez interpreted the phraseologic process can be lacunar as phraseological meaning can be opaque along with source meaning unclear. Phraseologic units demonstrate the ideal “test bench” proving the validity of the Invited Inferencing Theory of Semantic Change.

As Lopes finds (2015: 174) phraseologic units carry a heavier semantic load, thus the conventionalized implicatures can be very comparing to other linguistic elements. The archaic expression like *down the pike* (194 COCA results) turned into other more evident etymology changing into *down the pipe* (69 COCA results), the set-expression *American Idiot* (COCA, NEWS, 2016, OCRegister) returns 84 results in COCA engine that were presupposedly borrowed from the idiomatic expression the *American Patriot* with 46 COCA results.

Sometimes it happens when the origin of the word/term is borrowed from Latin or Greek, or from other foreign languages, cf.: *tele-collection*. The neologism is exploring mimesis and looks like the old term, but it is completely brand new conceptual unit like *-ism* derivatives into abstract

nouns, cf.: *Trumpism* (55 units in COCA search). Old terms' traces are iterated by the new second-hand users disguised as pseudo-terms as in *Trumpism of the Day* where the neological unit *Trumpism* occurs.

MWU can be polyfunctional, thus the one meaning has to dominate contextually, Cf.: *JFK* (2207 units in Frequency rate) as (1) *Junior Fellowship Kidz*, or as *Junk Food Kitty* (796 entries for Junk Food), or (2) *Jihads For Kerry, John Fitzgerald Kennedy* as in (3) *the JFK assassination* (58 frequency rate) (COCA, 2017, NEWS, Chicago Sun-Times); (4) *John F. Kennedy = John F Kennedy International Airport* (erasure of the word airport) (COCA, 2017, SPOK, NPR: Planet Money).

Picture 1. JFK variation: *Junk Food Kitty*, *JFK assassination*, *John F Kennedy International Airport*



Source: the Internet.

The Phraseological MWUs are secondary units that are in comparison with free expressions have stable indirect meaning, can be described as highly figurative, metaphoric, being stick together in the preserved collocation, e.g.: *Early bird catches the worm*. Still, these secondary units can be modified, and changed into the post-secondary, tertiary, undergo fourth change and beyond, cf.: (1) *There's an early bird buy-one, get-one-free special*, (2) *to work the early bird on Friday*, (3) *with early bird discounts*, (4) *the Early Bird Detective Agency*, (5) *moderate early bird*, (6) *with early bird dinner specials*, (7) *early bird bingo*, etc.

According to Mieder (2009: 77) the phraseologic units and felicities are often used to demonstrate argumentation strategies for they are easy-to-understand expressions and demonstrate persuasive argument by able-to-get

way with conclusive proof supported by metaphors and popular phraseology, cf.: (1) “*Opportunity doesn’t come easy*”, (2) “*One man cannot make a movement*”, (3) “*A new politics for the new time*”. Some set-phrases, collocations and proverbs make politician’s speech popular (cf. Obama’s proverbial Rhetoric in “*Yes, we can*” (2009)), other fading away (2009: 161). As Mieder states (1993: 209) proverbs like: “*A woman’s tongue wags like a lamb’ tail*” or “*Spare the rod and spoil the child*” have disappeared or are on their way out while such proverbs as “*A woman without a man is like a fish without a bicycle*” or “*There is no free lunch*” are steadily gaining in currency. Some proverbs are rearranged: “*Use it or lose it*”, “*Pay as you go*” and “*lipstick on a pig*” (1993: 83).

The proverbs are seen as traditional signs of cultural values. The linguists have studied proverbs to observe expressions of national wisdom to change and reoccur (Mieder 1993: 205). The new proverbs as they emerge illustrate new values:

Nothing but money is sweeter than honey. Banks have no hearts,
and Money makes the mare go but not the nightmare.
Old shoes and old friends are best. Mud thrown is ground lost.
Friendship can’t stand on one leg alone.
You have to summer and winter together before you know each other.

(Mieder 1993: 228)

The secondary units or all rearranged units on the one hand have some similar traces, demonstrating mimesis; on the other hand they have some vivid formal replacements or semantic change or multiple semantic or other changes as in artifact rearrangements, cf.: *Diamonds Are A Girl’s Best Friend* introduced by Carol Channing, vivified in other tributes.

The term “law of Hobson-Jobson” is used in linguistics to define the process of phonological change when loan words can be adapted to the phonology rules of a new language. This term indicates that words are changing in appearance as the Spanish *cucaracha* becoming English “*cockroach*”, and English “*riding coat*” becoming French *redingote*. The secondary term that derived from *Hobson-Johnson* is *Hanklyn-Janklin* demonstrated mimesis to the earlier term denoting the contemporary glossary of Indian English terms and Indian-derived words in mainstream English by Nigel Hankin, named as a tribute to its 1886 forebear Hobson-Jobson (Purcell 2009).

Multi-word units correspond to one grammatical phrase constituted by several lexemes separated by a blank (Bolly 2009: 9). They differ from free combinations, whose constituents keep their syntactic and semantic independence, and from compounds, which are morphologically made up of

two elements which have independent status outside these word combinations.

Often MWUs alter their sense while entering other linguistic area. The borrowed set expressions appear in different areas like “two peas in a pod” in different areas: translations (retranslations), film industry (remakes), arts (repetitive copies, recopies, similar replicas), in internet (related videos), in marketing (mimetism representations as celebrity imitation in appearance, clothes or manners, twinsumers¹, etc.).

When MWUs are also terms from two languages differ considerably or when a term exists in one language only, there are three basic techniques for reproducing the term in another language: borrowing, creating new term and creating equivalent paraphrase (Arntz 1993: 15).

The use of a loan word, i.e. the direct coining of a term from another language, is indicated when the content of the term is especially typical for the area in which the source language is spoken and is therefore difficult to translate (e.g. drugstore in North America and ombudsman in Sweden). That language usage is by no means restricted to the case shown by word Engl. *Know-how*, Ukr. *nou-khau* and very many others which were taken over into Ukrainian unchanged. A loan translation (e.g. Engl. *contact lenses*, Ukr. *kontaktni linzy*) can facilitate the comprehension of a term which is unknown in the area in which the target language is spoken: however, this requires a motivated term in the source.

According to conceptual change theory the semantic change is like “rearranging nodes in the network” which needs addition or deletion some links and nodes, involving restructuring and replacing the whole conceptual networks (Pavel 1993: 22).

The other important issue of MWUs transfer and words’ migration from one language into other, from one sphere into another: e.g. the term *reproducibility* from biology migrates to linguistics, the term *nonlinear dynamics* used by linguistics in synergetic approaches.

MWUs change in headings changing their meaning, in film’s names, e.g.: “*Men prefer blondes*”. Creating sequels of books and numerous screen versions provide lacunar prototypical copies, relic or artifact parodies, such as “*Casino Royale*” (1967). New versions are not worthy without comparison to older ones. Translation and a new film are always intertextual. The intertextuality is a “right hand” of semantic lacunarity. Artifact, where you can find the gaps, is incomplete with respect to its real or virtual prototype. Memory

¹ TWINSUMER trend: consumers don’t connect to “just any other consumer” anymore; they are looking for the most relevant of their taste “twins”.

and Identity can be both lost in “Eternal Sunshine of the Spotless Mind” (King 2013). When something is lost in translation, it must be the identity of the original.

Among the most important factors of MWUs transfer and reproducibility are creativity and mimesis (Pavel1993: 30). Some characteristics of MWUs can be observed in scientific discourse, corpus data, and wide context search engines. The neos of MWUs show the lexical creativity as Pavel states: “concept-theme feedback loops display remarkable similarities” to the original MWUs in comparison to the post secondary MWUs (1993: 30).

Among the current best new words (hybrid compounds) in American English are there challenges to translation, e.g.: *masterdating*, *nonversation*, *cellfish*, *errorist*, *internest*, *chairdrobe*, *afterclap* (New Modern random words). These rearranged MWUs are highly frequent and rather colloquial.

2. Methodology and materials

Our approach relies on corpus-based studies with possibility of data extraction from different corpora. Due to the intuitive interface of the Corpus of the Contemporary American English we have chosen COCA. As our purpose to rely the corpus based approach to teaching EFL we prefer online method which gives the possibility to operate huge data without specialized tools.

2.1. Reproducibility in demo: artifact replicas, the units of deviation, the changed MWUs, neology

Often MWUs alter their sense while entering other linguistic area. Thus, the set expressions borrowed from other language and culture are changing their appearance. This is Hobson Jobson phenomenon. When artifact enters the new territory, the change in appearance often takes place, compare Vasily Perov’s “The Hunters on a halt” (1871). The replicas are slightly changed but the original is visible and identifiable, thus its reproducibility is evident.

The two illustrations (their titles are working as conceptual metaphors) bellow are not the only examples of reproducibility of the original “The Hunters on a halt”, but the following copies “*The Chukot Hunters*” and “*The Smoko time*” demonstrate how the intra-cultures react and rearrange the original into the secondary and the post-secondary replicas with embedded cognitive and conceptual changes shown by the different diachronic and national heritage.

Picture 2. The Original of “The Hunters on a halt”



Picture 2a. “The Smoko time”



Picture 2b. “The Chukot Hunters”



Source: the Internet.

The conceptual metaphors demonstrate reproducibility in MWUs as “*The Hunters on a halt*” (Picture 2) turns into “*The Smoko time*” (Picture 2a) and “*The Chukot Hunters*” (Picture 2b). Highly figurative MWUs are not phraseologic units but also have semantic and visual form change, with a reproducibility feature in common with other recopies: remakes, winged phrases, high profile cases, *cause celebre* and other reproduced copies of the original.

3. Results

3.1. Frequency data analysis and concordance search for MWUs: the current usage

The MWU is a good precedential turn of phrase, then the frequency of usage rises and the felicity turns into the stable collocation fixed in speech, then in writing, in fiction, magazine, academic discourse with different rate of occurrence, cf.: (1) *time flies so quickly* (COCA 1996, CBS SPOKEN), (2) *time flies like an arrow* (COCA 2009, FIC Analog), (3) *time flies by anyway* (COCA 1997, NEWS, SanFrancChron).

Syntactic “fixedness” relates to the degree of grammatical rigidity or frozenness of the unit, including constraints on syntactic order and transformational deficiencies of the MWU under study: “a sequence is considered syntactically fixed if it does not allow any of the combinatory or transformational possibilities that are typical of this kind of sequence” (Bolly 2009: 9).

Metaphorical and metonymic multi-word combinations are typical cases of semantic opacity. “Lexical restrictions” relate to the degree to which there are constraints on the “commutability” of the MWU’s constituents. That is, PUs show “preferred lexical realizations” on the paradigmatic axis: they display arbitrary lexico-grammatical restrictions in restraining possibilities of paradigmatic substitution of one constituent for another (quasi) synonymic item.

Then, after pushing “Find matching string” you are getting 100 collocations represented in the frequency of usage order in the next page by the KWIC concordance examples (short textual illustration) to each set expression like or *down the gauntlet*. The meaning derived from MWUs can be seen as modified phraseological unit, cf.: *The Republicans in the House have thrown down the gauntlet* (COCA, 2013, SPOK PBS).

Alongside this restricted interpretation of phraseology in “phraseologic approach”, there is another way of defining it, which she refers to

as the “frequency-based approach”. An expanded definition of phraseology would therefore include the study of all the “frequently occurring syntagmatic combinations” revealed by corpus linguistic analyses (Bolly 2009: 10). In order to follow the change of the phraseological unit we rely on search engine of COCA to find similar collocations with the structure “down the [*N]”, for instance: “*down the drain*”.

This corpus-based approach enabled the researcher to consider the advanced learner variety from a new perspective, in terms of overuse, underuse, and misuse. The combined method of investigation highlighted a striking difference in phraseologic behaviour between high-frequency MWUs in available dictionaries and in corpora data.

The constant change and modification of free and fixed expressions are given by dictionaries and corpora data also having data presented by styles (Magazine (MAG), Academic (ACAD), fiction (FIC), other items in the Corpus of Contemporary American English (COCA), Cf.: (1) *Blue Beard – a man who marries and kills his wives* (Longman, 129), (2) *Blue Beard: PHOTO (COLOR)*, 1991, pastel, 131/2 x 13 (COCA, AmerArtist 1992), (3) *portrait as Blue Beard* (COCA, 1992, AmerArtist), (4) *original story of Blue Beard* (COCA, 1992, AmerArtist), (5) *rewriting of Blue Beard, Carter story as a repeat performance of Blue Beard* (COCA, 1992, AmerArtist).

Due to the quantitative analysis of the corpus data by to the “Clusters” and “Collocates” tools of the *OxfordWordsmith Tools 4.0* concordancer our preliminary analysis reveals (a non-significant) difference in the frequency of use of some recurrent sequences and collocational combinations in the native and non-native corpora.

Some MWUs may be avoided (underuse) and some well-known MWUs may be overgeneralized (overused). The deviant production related to the word semiosis in the native area of the term. For instance, in America it is not popular to use old forms of words, so the paper-and-pencil game (now electronic also), where we put O or X in nine squares trying to win a row of three O's or X's was changed from *Noughts and Crosses* (OXO) (BrEngl) into *Tick-tack-toe* (AmEngl) derived from “*tick-tack*”, both with possible roots to the ancient Egypt game. The expression *Noughts and Crosses* is not available at COCA. The word *nought* is the old form of the word *zero*, rarely used in American English according to the Corpus of Contemporary American English (COCA), giving totally 72 results for Academic discourse (5), precisely: ACAD/Theological Studies (2), ACAD/PublicLaw (1), ACAD/Current Psychology (1), ACAD/ArtBulletin (1), FIC (51), MAG (9), NEWS (3), SPOKEN (1). This change demonstrates how thematic variation preserves the conceptual stability (Pavel 1993: 21).

“Non-compositionality” relates to the degree to which the phraseologic unit is semantically opaque and often lacunar. A string of words is considered to be partially non-compositional when the meaning of at least one of its constituents no longer corresponds to its prototypical or literal meaning. One can consider the meaning of a combination of words to be totally “non-compositional” when “the meaning arising from word-by-word interpretation of the string does not yield to the institutionalized, accepted, unitary meaning of the string” (Moon, 1998: 8).

4. Discussion

The last three above-mentioned internal features of PUs (syntactic, semantic and lexical criteria) should be considered in terms of continua. In other words, this would mean that there are degrees of syntactic fixedness, semantic non-compositionality and lexical restrictions, and that this variability differs from one PU to another. Combining these three graded criteria on one scale should thus arguably make it possible to classify word combinations from the freest to the most phraseologic (Bolly 2009: 24).

As the extraction in COCA of MWUs, PUs and non-fixed bigrams, trigrams and N-grams demonstrate that the contemporary language can be verified to the corpus-based approaches which are important for non-native English speakers and TEFL. We find the other worthy corpora as BYU – BNC (the British corpus), Strathy (Canadian corpus), iWEB or stand alone corpora like LOCNESS to dig in the further research for the sake of corpus and applied linguistics.

5. Conclusions

The derived MWUs demonstrate high frequency usage due to the Corpus based Analysis and Concordance Search, provided by COCA. The semantic shift of MWUs, set-phrases and idioms deviations take place at all levels of the contemporary discourse: in Fiction, Academic, Spoken, News, and Magazine. In order to eliminate the informative lacunae of the modified MWUs in the secondary and post-secondary deviations there is provided to use compare and contrast tool, wide context tool, idiomatic search engines, etymological background analysis enabling to reconstruct the intra- and inter- change of the formal level and semantic shift of the inner modifications bringing home the message.

The analysis of corpus data and concordance illustrations in COCA showed how phraseologic units were slightly changed and modified by

means of the secondary (subsequent) semantic derivation. Many MWUs in Spoken and Magazine, News endozones were high-frequency collocations of deviant nature in the modern AmEnglish comparing with their non-deviant source MWUs in EFL (English as Foreign Language). This is the reason why neology and phraseology are better learned and taught at the university level, where the self-education and corpus analysis are engaged.

The detailed study has revealed that most errors concerned the misuse of MWUs in TEFL are explained not by out-date resources EFL users use, but due to the rapid semantic change of MWUs making neology or rethought winged phrases or idioms in the real time regime, what is affirmed by corpora data from the current spoken, academic and magazine discourses. Some other questions concerning MWUs change, modification and transfer remain open to debate.

Bibliography

- Arntz, Reiner. 1993. Terminological equivalence and Translation. In: Helmi B. Sonneveld, Kurt L. (eds.). *Loening Terminology: Applications in interdisciplinary communication*. Amtrerdam / Philadelphia: John Benjamins. 5–21.
- Bolly, Catherine. 2009. The acquisition of phraseologic units by advanced learners of French as an L2: High-frequency verbs and learner corpora. In: Labeau, Emmanuelle, Myles, Florence (eds.). *The advanced learner variety: The case of French*, Peter Lang: Bern, Berlin, 1–24. URL: <http://hdl.handle.net/2078.1/83340> Retrieved: 16.11.2017.
- Colson, Jean-Pierre. 2017. The IdiomSearch Experiment: Extracting Phraseology from a Probabilistic Network of Constructions. In: *EUROPHRAS 2017: Computational and Corpus-Based Phraseology*. 16–28. DOI: 10.1007/978-3-319-69805-2_2
- Heid, Ulrich. 2007. Computational linguistic aspects of phraseology II. In: Dobrovot-skij, Dmitrij. Kühn, Peter and Norrick, Neal R. (eds.). *Phraseologie / Phraseology Harald Burger*. Berlin / New York: Mouton de Gruyter. 2: 1036–1044.
- Heid, Ulrich. 2008. Computational phraseology In: Sylviane Granger und Fanny Meunier (eds.). *Phraseology. An interdisciplinary perspective*. Amsterdam / Philadelphia: Benjamins. 337–360.
- Heid, Ulrich. 1994. On Ways Words Work Together – Topics in Lexical Combinatorics. In: *Martin, W. et al.*, 226–257.
- Hünning, Matthias & Schlücker, Barbara. 2015. Müller, Peter O. Ohnheiser, Ingeborg, Olsen, Susan, Rainer, Franz (eds.). In: *Word Formation, An International Handbook of Languages of Europe* (HSK series). Band 40. Volume 1. Berlin: De Gruyter. 450–467.
- Lopez, Elena Sánchez. 2015. Phraseologization as a process of semantic change. In: *Catalan Journal of Linguistics*. 159–177.

- Masini, Francesca. 2005. Multi-word expressions between syntax and the lexicon: the case of Italian verb-parücie constructions. In: *SKY Journal of Linguistics* 18: 145–173.
- Masini, Francesca. 2009. Phrasal lexemes, compound: and phrases: A constructionist perspective. In: *Word Structure*. 2: 254–271.
- Mieder, Wolfgang. 1993. Yankee wisdom: American proverbs and the worldview of New England. In: Skandera, Paul (ed.). *Phraseology and Culture in English*. 205–235.
- Mieder, Wolfgang. 2009. “Yes We Can”: Barack Obama’s Proverbial Rhetoric, New York: Peter Lang. 352 p.
- Pavel, Silvia. 1993. Neology and Phraseology as Terminology-in-the-Making. in *Terminology: Applications in interdisciplinary communication*. Sonneveld, Helmi B., Loening, Kurt L. (eds.) Amsterdam / Philadelphia: John Benjamin. 21–33.
- Pintarić, Anita Pavić; Škifić, Sanja. 2016. Changes of Proverbs in Form and Meaning: The Case of Croatian ‘Odjeća ne čini čovjeka čovjekom’ (Clothes Do Not Make the Man). In: *Folklore*, 127(2), 210–228.
- Purcell, Susan. 2009. “The Law of Hobson-Jobson”. In: *English Today*. 25 (01): 60–64. doi: 10.1017/S0266078409000108. Retrieved: 11.05.2017.
- Sprenger, Simone. 2003. Fixed expressions and the production of idioms. Ph.D. dissertation. Nijmegen: Max Instituut voor Psycholinguïstiek.
- Szerszunowicz, Joanna. 2013. Phraseological Gaps as a Translation Problem. In *Research on Phraseology Across Continents*. Volume 2. Bialystok: University of Bialystok Publishing House. 299–319.

Online Sources

- Corpus of Contemporary American English (COCA) – BYU corpora. URL: <https://corpus.byu.edu/coca/?c=coca&u=15255&k=40>. Retrieved: 18.11.2017
- Strathy. <https://corpus.byu.edu/can/>. Retrieved: 06.01.2018.
- The BNC. <https://corpus.byu.edu/bnc/>. Retrieved: 06.01.2018.
- The LOCNESS-CORPUS <http://www.learnercorpusassociation.org/resources/corpora/locness-corpus/>. Retrieved: 01.09.2018.
- New Modern random words. URL: <https://www.demilked.com/new-modern-funny-random-words-portmanteus/>. Retrieved: 17.11.2017.

Die Formänderungen und Bedeutungsverschiebungen der englischen MKEs

Zusammenfassung

Die moderne Linguistik verfügt über eine Vielzahl von Ansätzen für die Beschreibung der Bedeutungen. In unserer Forschung verwenden wir den Korpusansatz,

um eine besondere Sichtweise auf Bildung, Veränderungen und Umgestaltung von Wörtern als Multi-Komponenten-Einheiten (MKE) vorzuschlagen. Im Artikel wird die Verschiebung der ursprünglichen Bedeutungen von Multi-Komponenten-Einheiten (MKE) untersucht, die in Rahmen von Veränderungen ihrer kognitiven Szenarien beschrieben werden. An Beispielen aus der modernen amerikanischen Variante der englischen Sprache werden die Einsichten in die Veränderungen der Form und die Verschiebungen der Bedeutungen von MKE gegeben. Die abgeleiteten MKE, die in der Arbeit beschrieben werden, werden als sekundäre sprachliche Einheiten betrachtet, bei denen die tatsächliche Bedeutung geändert, getarnt, modifiziert oder gelöscht wird. Die Vorgänge der semantischen Abweichung werden als Verschiebungen und Neuaneordnungen beschrieben, die die Ähnlichkeit demonstrieren, wenn die ursprüngliche Bedeutung von MKE teilweise modifiziert oder gelöscht werden kann. Die semantischen Verschiebungen werden als sekundäre und postsekundäre Änderungen betrachtet. Im Verlauf der semantischen Abweichung verschwindet die alte Bedeutung des modifizierten bildhaften Ausdrucks und dabei entsteht ein neuer Begriff. In der Arbeit werden die Operationen der semantischen Ableitung beschrieben. Die kognitiven Szenarien werden als Verschiebung und Neuaneordnung sowie zweite und folgende Ableitung beschrieben. Die empirischen Verfahren und Methoden zeigen, wie sich die ungeordneten MKE in ihrer Form verändern und sich semantisch verschieben. Die semantischen Verschiebungen werden als die sekundären und postsekundären Änderungen betrachtet. Die anfängliche Bedeutung von MKE kann auch teilweise modifiziert, getarnt oder gelöscht werden. Im Prozess der semantischen Ableitung ändert sich die tatsächliche Bedeutung. Die modifizierten bildhaften Ausdrücke zeigen Ähnlichkeit mit den ursprünglichen MKE, die konvertierten MKE scheinen völlig neue konzeptionelle Einheiten zu sein. Die Interpretation des Umstrukturierungsschemas der MKE wurde durch verfügbare empirische Verfahren vorgenommen, um sekundäre und postsekundäre Abweichungen der MKE hervorzuheben. Einige Fragen, die sich auf die sprachlichen Einheiten mit vielen Komponenten beziehen, sind weiterhin für Diskussionen offen.

Damien Villers

Université de Toulouse-Jean Jaurès

France

<https://orcid.org/0000-0002-6382-4731>

Phraseological Blunders: When New Phrasemes Are Born from Errors

Abstract. This article first presents a typology of phraseological “blunders” – which include both errors and mistakes that accidentally modify the standard form, usage, or meaning of a phraseme – and then outlines the causes and the types of interference behind them, such as language pathologies, poor linguistic skills, or mere absence of mind. It then proceeds to study the notion of variation in order to draw a line with the notion of error with the help of criteria such as frequency ratio, communicational efficiency, and semantic coherence. Finally, the article presents cases of phraseological accidents that have lexicalised due to various cognitive biases, thus becoming new phrases. In order to account for this seemingly paradoxical phenomenon, the memetic approach is selected to build a presentation of the “phraseme genesis” process (or “phraseologisation”) and the selection criteria that facilitate it.

Key words: *phraseme, error, variation, propagation, memetics*

1. Introduction

As a result of their prefabricated nature, phraseological units (PUs) exhibit some formal stability, and are thus prone to various modifications. It is important to stress from the very start that this study will not deal with deliberate or premeditated modifications. Such modifications have been studied extensively and are generally called “creative idiom modifications”, “anti-idioms”, “phraseological puns”, “idiom parodies”, “twisted phrasemes” and so forth. This study will only tackle phraseological modifications that are the result of unintentional deviation and that do not match the attested form, usage, or meaning of a phraseme. As for the term “phraseme”, it will be used

interchangeably with “phraseological unit”¹ to refer to all² preconstructed³ polylexical units. This definition implies that the traditional view was chosen in this study. Let’s consider the following inappropriate utterances:

- (1) I tell people, let’s don’t fear the future, let’s shape it!* (G. W. Bush, Omaha, 2006).
- (2) They misunderestimated me* (G. W. Bush, Bentonville, 2000).
- (3) Being from a family of outlaws makes you a social leopard* (*The Law of Finders Keepers*, 2018).

Some scholars have recently argued in favour of a wider conception of phraseology that includes even polylexemic units, i.e. patterns and compounds. Under that view, the mistakes in examples (1) and (2) should be incorporated in the present study. The former is a blend of *let’s not* and *don’t* while the latter seems to be a mix of *misunderstand* and *underestimate*. However, under the traditional view, only example (3), where *leper* was replaced with its near homonym *leopard*, will be considered a phraseological blunder. As for examples (1) and (2), they will be considered syntactic and lexical blends.

Bergstrom (1906), Bolinger (1961), Cohen (1987), and Legallois (2013) are among the very few who published articles and volumes on syntactic blends – and also lexical blends in the case of Bergstrom. Works on “phraseological errors” are more recent and appeared in the 2000s due to the boom of phraseodidactics, the study of phraseological competence for language teaching and learning. Among notable works, it is possible to cite Nesselhauf (2003), Osborne (2008), Paquot (2008), Thiessen (2008), or Wang & Shaw (2008). They all focus on erroneous collocations among L2 English learners and the influence of their native language (L1) on such errors. From a less didactic and more linguistic viewpoint, Polguère (2007) studied the nature of “collocational grafts” (or “blends”), a type of phraseological error; while Liudmila Liashchova gave a presentation at Europhras 2018 about the “phraseological errors” made in Russian by a highly fluent US journalist.

¹ Other famous general labels include “formulaic language”, “fixed expressions”, “multi-word units”, “set-phrases”, “idioms” (which may also refer to a non-compositional subclass of phraseme), or “lexical bundles” among computational linguists.

² A minority of scholars, especially in Russia and Slavic countries, use the term to refer to figurative set phrases only. This is due to the influence of A.V. Kunin’s works and terminology, which were greatly influenced by Charles Bally’s *Traité de Stylistique Française* and its terminology.

³ From a cognitive viewpoint, other common terms include “reproducible”, “formulaic”, and “prefabricated”. From a structural point of view, they correspond to the notions of “frozenness” and “fixedness”, or the less aggressive notion of “stability”.

She analysed from a linguistic-pragmatic viewpoint some of his mistakes and the reasons behind them. Finally, it is worth noting that a website dedicated to idiom-based blends – “malaphors⁴” – appeared in 2012 (malaphors.com). It serves as a database and a forum where users list and discuss idiom blends gathered in the media.

2. Typology of phraseological blunders

Several typologies of phraseological errors have been formulated by scholars. According to James (1998), phraseological errors may be classified into two general categories: “grammaticality errors” (incorrect form) and “acceptability errors” (inappropriate usage). Thiessen (2008: 5–6) proposes a classification that is based on (some) phraseme types: lexical collocation errors, grammatical collocations errors, errors in idiom-like phrases, and phrasal verb errors. Finally, Liashchova (2018) notes that the major deviations are excessive use and inadequate use. She subdivides the latter into erroneous synonymity, loan translation or calque, comprehensible blending and incomplete semantization. While very useful, these typologies are either very general or focus on very specific approaches (phraseme type or semantics) and do not account for all existing types of phraseological accidents. Therefore, the following classification is proposed:

- Substitution

(3) Being from a family of outlaws makes you a social leopard* (*The Law of Finders Keepers*, 2018).

- Permutation

(4) Look who’s calling the pot black!* (*NCIS*, 3–18).

- Expansion

(5) I’m so fed up with being the escape goat for all the problems of my bloody family* (wehavekids.com).

- Omission

(6) I personally could care less if the shirt is made in the U.S. or not. That doesn’t matter to me* (www.mintees.com).

⁴ The term itself is a blend of *malapropism* and *metaphor*.

- Fusion or “blending”

(7) The whole Transformers thing isn’t my kettle of tea* (eurobricks.com).

- Inter-language calque

(8) You’re selling the bear’s skin before you’ve killed it!*

- Inappropriate contextual use.

(9) I know my lesson off the top of my head*.

- Wrong interpretation (absence of comprehension, semantic blends, literal reading etc.)

- Overuse and underuse.

- Mixed type

(10) Families is where our nation finds hope, where wings take dream* (G. W. Bush).

Substitution, exemplified by (3), is probably the most common type of phraseological blunder. In this example, the idiom *to be a social leper* (to be shunned) is modified by substitution of the term *leper* with its near homonym *leopard*. Such Homonymy-based errors are often called “eggcorns⁵”. Example (4) is a modification of *to call the kettle black*, which is a variant of *the pot calling the kettle black* and refers to hypocritical criticism. In this case, the terms *pot* and *kettle* are substituted, reversing their order. Such permutations are a rare type of malaphor and could be seen as a subtype of substitution. In addition, it must be noted that instances (3) and (4) are errors from the character’s viewpoint only since they are intentional for the writer. In (5), the word *scapegoat* is turned into a collocation that does not exist, thus increasing its lexical volume. Naturally, the expansion category does not only contain words that are turned into phrasemes, but may also concern phrasemes that are made even longer by adding letters or words to them. It is the case with *to exact revenge* and *to set foot on*, which are often misused as *to extract revenge** and *to step foot on**. Example (6) is the exact opposite, as the conversational routine *I couldn’t care less* was made shorter by omitting the term *not* or the phonemes /n(t)/. Another common mistake is *for all intensive purposes**, which is a lexical reduction of *for all intents and purposes*. Blending⁶ or “fusion”, exemplified by (7), is a common type of error, at least the most studied

⁵ The term is generally attributed to linguist Geoffrey Pullum, who reportedly suggested the label after reading an article about a patient who substitutes the term *eggcorn* for *acorn*.

⁶ Aarts (2007: 189), who studies syntactic blends, further distinguishes between the more integrated “blends” and “mergers”, where two distinct components can still be identified.

in literature. In this instance, the idioms *kettle of fish* (a tricky situation) and *not my cup of tea* (to dislike something) were blended together. Utterance (8) is a case of inter-language calque, where a phrase is translated word by word into a language in which it does not exist. Here, the phrase in question is *vendre la peau de l'ours avant de l'avoir tué*, whose correct English equivalent is *to count your chickens before they've hatched*.

Unsurprisingly, PUs – mostly non-compositional ones – may also be misunderstood or misused. Such “acceptability errors” (James 1998) are extremely frequent, especially among learners, which explains why the vast majority of studies on phraseological errors are language-learning oriented. Inappropriate usage is exemplified by (9), where the idiom *off the top of my head* (without thinking) is used in a wrong context, where *by heart* would have been more suitable. The next category, wrong interpretation, is probably the most common type, especially in translation exams. For instance, second-year students at the university of Toulouse were given a news article to translate into French for their exam. It contained an excerpt about a big company's performance on the Dow Jones industrial average, in which the said company was referred to as a “blue-chip conglomerate”. Out of 23 students, 21 failed to get the correct meaning in their translations, opting instead for omission, loan translations, nonsensical phrases, false meanings, barbarisms, and so forth. This shows that wrong interpretation could be further divided into several types that include the absence of comprehension as well as erroneous comprehension, which, in turn, comprises semantic blends with another phraseme or literal reading. Another type of blunder is overuse and underuse. Paquot (2008) gives the example of *for instance* and *for example*, which are overused among L2 English learners when expressing exemplification. Finally, the typology would be incomplete without a category for hybrids. Some blunders, as in (10), are a mix of several types. Let's suppose for a moment that George W. Bush meant “where dreams grow wings”. This would imply that he mixed up *to grow wings*, *to take flight* and reversed the word order. Similarly, *to step foot on** may be viewed as an addition (of one letter), or as a blend of *set foot on* and *step on*, or even as a substitution.

3. The causes behind phraseological blunders

As stated before, the majority of studies on phraseological errors deal with language teaching and learning. Therefore, they focus on errors made by learners and non-natives (e.g. Nesselhauf 2003, Osborne 2008, Paquot 2008, Thiessen 2008, or Wang & Shaw 2008). Unsurprisingly, these studies show

that their native tongues have a lot of influence on errors in L2 English and that insufficient knowledge of a language is the main cause for phraseological errors among learners. However, other causes may be put forward: as Granger (2004: 135) puts it, “advanced interlanguage is the result of a very complex interplay of factors: developmental, teaching-induced, and transfer-related”. In her presentation on errors made in Russian by a non-native US journalist, Liashchova (2018) also noted that incorrect use is due to insufficient learning. As for phraseme overuse, she posits that it is the result of personal affection for idioms, a desire to make use of their pragmatic functions, as well as a desire to sound authentic. In his PhD thesis on “frozenness”, Misri (1987: 414) claims that accidental modifications are common “among children who have not yet mastered the linguistic system, foreigners who have limited knowledge of frozen units, and patients with language-related pathologies”.

Many studies⁷ confirm that proverbs and idioms are used in various tests to detect pathologies such as dementia, Schizophrenia, Right Hemisphere Damage, Alzheimer’s, or aphasia. These tests include IQ tests (*Wechsler Adult Intelligence Scale / WAIS-R, Stanford-Binet test*), personality tests (*Famous Sayings Test, Attitude Measurement Test*), executive functions tests (*Delis-Kaplan Executive Function System, Proverb Interpretation Task*), or psychopathology tests (*Gorham Proverb Test*). The omnipresence of proverbs – at least figurative⁸ ones – and idioms in such tests is due to their neurocognitive complexity. As explained by Honeck (1997: 220–222) and his DARTS model, both brain hemispheres are required to process them, which makes them very valuable from a medical point of view.

Nonetheless, reducing phraseological blunders to pathologies or insufficient learning is excessive, as they only account for a fraction of such blunders. As a matter of fact, the vast majority of native speakers, including experienced and healthy speakers, are prone to such errors. According to Bergstrom (1906), who studied syntactic and lexical blends, blunders are caused by some sort of “contamination”; while Polguère (2007), who studied collocational blends (or “grafts”), claims they are caused by an “interference”. This explanation may be applied to all types of phraseological blunders but these phenomena should be seen as intermediate causes, and not the root cause of the problem. Both scholars agree that contamination or interference

⁷ Most of these studies on neurological applications for proverbs and idioms are summarised in Van Lancker (1990) and Murphy et al. (2013).

⁸ Paremiologists are divided as to whether metaphor should be considered optional or obligatory in proverb definition. This point is debated in Villers (2014), along with other criteria.

can in turn be explained by the notion of *analogy*. In other words, phraseological blunders are due to a similarity with a sound, a lexical element, a theme, or a meaning found in another phraseme – including phrasemes from other languages. Once again, these phenomena may be seen not as the root cause of the problem, but as intermediate causes. Furthermore, analogy is a very common⁹ cognitive process, at the heart of language and idiom processing and learning. Since analogy is so omnipresent, more specific triggers need to be put forward. Such triggers have actually long been identified in cognitive and psycholinguistic studies¹⁰ on performance errors and are generally labelled “performance factors” after Chomsky’s *Aspects of the Theory of Syntax* (1965). They include lack of attention, memory lapses, tiredness, lack of interest, emotional state, drugs and alcohol, etc. Insofar as these factors apply to all types of speech, they are, of course, valid for phraseological blunders as well. In didactics, the label “error” generally implies systematic deviation owing to competence factors while “mistakes” entail temporary performance factors. Therefore, the term “blunder” will be used to encompass both “mistakes” and “errors”.

4. The blurred line between error and variation

Dealing with the notion of error instantly raises the difficulty of setting boundaries with the notion of variation. Phraseological blunders were previously described as utterances that are the result of unintentional deviation and that do not match the attested form, usage, or meaning of a phraseme. Yet, some deviations are hard to classify, especially the ones that occur frequently. In this case, two ideologies clash: descriptivists will consider that common or systematic deviations should not be viewed as errors but as idiosyncrasms, while prescriptivists will view them as erroneous. It is therefore necessary to find a more objective and intermediate stance that incorporates all relevant criteria in order to establish functional delimitations. In corpus phraseology, the minimum frequency of (co)occurrence for a string of words to be considered preconstructed or “reproducible” – and thus deserve the label

⁹ The same could be said of blends. According to the “conceptual blending” theory, designed by Fauconnier and Turner (2002), blends are an omnipresent mental process; they are at the centre of how we create meaning and how we think. This theory is similar to George Lakoff’s “conceptual metaphor” theory, according to which we see the world through metaphors.

¹⁰ Among major studies on performance errors, one may cite Crain & Thornton (1998, chap. 15), Gleason & Ratner (1993), or Kamhi (1988).

of “phraseme” – is generally described as being higher “than would be expected by chance”. This distinction could not be applied to errors related to preconstructed language units, as both erroneous and standard forms are bound to be recurrent. When it comes to errors derived from free combinations, there is no known quantitative threshold either. This means that phraseological blunders cannot be singled out by means of frequency alone. Other criteria must be taken into account, as will be revealed through the examples in the table.

Table 1. Examples of variants and errors with their frequency in online corpora

Utterances and possible variants/errors	Google Books corpus (189 bn words)	iWeb ¹¹ corpus (14 bn words)
(3a) A social leper	5,770/610	32
(3b) A social leopard*	151	0
(4a) (Pot) Calling the kettle black	49,600	695
(4b) (Pot) Calling the pot black*	2,560	25
(5a) A scapegoat	3,440,000	9415
(5b) An escape goat*	2,600	88
(6a) I couldn't care less	58,620	1,111
(6b) I could care less*?	66,700	3,526
(11a) The early bird gets the worm	8,170	367
(11b) The early bird catches the worm	13,100	257
(12a) Nothing ventured, nothing gained	44,900	118
(12b) Nothing ventured, nothing had*	132	0
(13a) For all intents and purposes	561,000	7,663
(13b) For all intensive purposes*	2,990	381
(14a) To exact revenge	71,900	1,600
(14b) to extract revenge*	4,420	92
(15a) To set foot (on)	88,000	14,190
(15b) to step foot (on)*	4,510	3,973
(16a) First come, first served	858,000	3,535
(16b) First come, first serve*?	70,800	3,847
(17a) A leopard cannot ¹² change its spots	7,466	19
(17b) A tiger cannot change its stripes	1,099	2
(17c) A zebra cannot change its stripes	257	1
(17d) A zebra cannot change its spots*?	494	2

Source: own research.

¹¹ Can be found on corpus.byu.edu along with other corpora compiled by Mark Davies.

¹² All full and contracted forms were included: *cannot*, *can't*, *does not*, *doesn't*.

The first three examples confirm that the frequency criterion alone is not reliable since even the most ludicrous mistakes are significantly frequent. Moreover, frequency numbers need to be taken with a pinch of salt insofar as some utterances may correspond to quotations or titles, not to mention the margin of error inherent to Google Books. This is precisely why (10), “wings take dream”, now a famous Bushism¹³, cannot be studied, as it was quoted and commented on at length. It even inspired book titles. In fact, the *ratio* between the standard form and the deviation should be considered the most relevant frequency-related criterion.

In the case of (5), where the ratio is 1320 to 1 and 106 to 1 in favour of (5a), this criterion is enough to declare (5b) erroneous. The ratio is much higher in the case of (3), with a proportion of 38 to 1, but it is still low enough to make (3b) a phraseological blunder, unless uttered in a documentary on leopards. As a matter of fact, the ratio for both (3b) and (5b) is so low compared to (3a) and (5a) that their frequency curves are not visible on the n-gram viewer, the analytical tool based on Google Books. From a semantic viewpoint, it is interesting to note that even if these errors are very amusing, they are not entirely illogical. Not only are they close homonyms, they also contain a certain degree of semantic motivation: leopards do tend to be avoided, and goats might allow one to escape. Besides, *scapegoat* was diachronically derived from *escape goat*. Example (4) is different: although its low ratio of 19 to 1 and 28 to 1 (cf. Figure 1) is sufficient to claim that (4b) is erroneous, another filter may be added: pun probability. As it turns out, a good proportion of contexts involving (4b) indicate that the deviation – which implies the repetition of *pot* – is deliberate, either for emphasis or humorous intent. When taking this factor into account, the ratio appears to be even lower for genuinely deviant uses. The same may be said of (3b) and (5b), which is why the figures obtained from corpora need to be put into perspective. To conclude, (4b) should be considered erroneous if it accidentally deviates from the idiom in (4a).

Examples (11) and (12) are probably the easiest cases. Both proverb variants in (11) are correct and attested in numerous dictionaries and occur quite proportionally in both corpora. By contrast, (12b) is not an attested variant of the proverb in (12a) due to its very low ratio (340 to 1). Deliberate modification is very unlikely in the case of (12b), which might account for such low figures. The next two examples remain fairly easy to classify but

¹³ This term commonly refers to a malapropism uttered by Georges W. Bush. His linguistic faux-pas were so common that a name was soon coined for them. There are even websites dedicated to them, and they even appear on tee shirts, mugs, posters, and so on.

Figure 1. Frequency curves for (4a) and (4b) on Google Books' n-gram viewer



Source: own research.

the frequency ratio is no longer the only relevant criterion. (13b) and (14b) both show a very low ratio¹⁴: 187 to 1 and 20 to 1 for (13), while (14b) is proportionally 16 and 17 times less frequent than (14a). From a semantic viewpoint, one may argue that the terms *intensive* and *extract* are nonsensical and illogical, even figuratively: how can a purpose be intensive; and why would revenge need to be extracted from something? Furthermore, these variants most probably originated in accidental phonological deformation or slips, as is often the case in idioms. All these criteria allow us to conclude that (13b) and (14b) are erroneous and should not be viewed as acceptable collocations, unless the deviation from standard form is intended as humour, sarcasm, a pun, or a quotation in order to comment on it.

Example (15b) is more difficult to analyse. First of all, its ratio varies greatly from one corpus to another. Thus, corpus type and speech quality should be taken into account. On the one hand, it shows a very low ratio (20 to 1) on a gigantic corpus of published books such as Google Books. On the other, it shows a much higher ratio (3.5 to 1) on a smaller corpus composed of websites and forums (iWeb), which are likelier to contain errors and loose language. It is also important to remember that some of the utterances from the corpora are meant to be sarcastic comments, which further balances this seemingly good ratio. In addition, the frequency curve obtained from the overall percentage for (15b) in the whole Google Books corpus is nearly inexistent on the n-gram¹⁵ viewer, compared to the curve for (15a) (cf. Figure 2).

¹⁴ With 1 representing [13b]. The potentially erroneous variant is always represented by 1 in the present article.

¹⁵ The n-gram viewer may show slightly skewed results due to its margin of error, as it is based on Google Books.

Figure 2. Frequency curves for *set foot* and *step foot* on Google Books' n-gram viewer



Source: own research.

Another argument is the semantic redundancy contained in *step foot*, given that the verb *step* already implies moving one's *foot* or *feet* and placing it or them on a new surface. Regarding its origins, the phrase is most certainly the result of an accidental graft or "blend" – between the collocation *to set foot* and the verb *to step (on)* – that was made possible due to phonological and semantic resemblance. For all these reasons, we argue that (15b) should be considered erroneous when used without satirical or humorous intent.

(16) poses a similar challenge as its ratio varies greatly from one corpus to another: 12 to 1 for Google Books and 1 to 1.1 for iWeb. Even when taking into account corpus type or speech quality and the rather low ratio of 12 to 1, the evidence against (16b) is scarce. A contextual analysis reveals that 58% of occurrences for (16b) include cases where the proverb is used as an adjectival phrase: *on a first-come first-served basis*. This suggests that one of the main factors that led to its emergence is phonological economy, in order to avoid specific phonemic combinations (viz. /d/ and /b/ without pause). The semantic criterion is of great help in this case as it reveals that (16b) is semantically incoherent since it states something different from what is meant (serving other people instead of being served). One may object that the active voice may be coherent from the perspective of a salesperson or a merchant ("serve those who have come first"), to which could be replied that the perspective of the beneficiary is the more logical one for the latter is the logical subject. Ultimately, this lack of semantic clarity violates the Gricean maxim of manner and may create misunderstandings or ambiguity, which is why it should be considered erroneous. For the same reasons, (17d) should not be considered a valid variant of (17a) given that zebras, as a rule, have no spots. Besides, the fact that it has more hits in corpora than

Figure 3. Frequency curves for (6a) and (6b) on Google Books' n-gram viewer



Source: own research.

a bona fide variant such as (17c) is attributable to a famous mistake, made by Al Gore, which was heavily commented on.

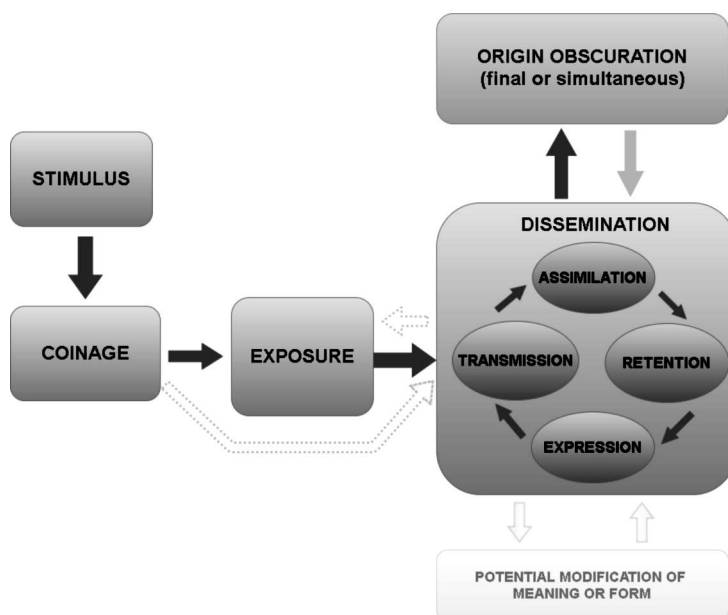
The most challenging case was saved for the end. In (6), the deviant and phonetically reduced form *I could care less* is more frequent than the standard one, or seems¹⁶ to be. While its ratio works in favour of (6b) in both corpora, the n-gram viewer – based on Google Books – shows contradictory results (cf. Figure 3). Moreover, a quick contextual analysis reveals that a large number of uses of (6b) turn out to be extracts where its grammaticality is debated. But even when balancing this ratio by filtering out grammatical comments and only keeping actual uses, *I could care less* is still very frequent, especially in the United States. The main criterion usually put forward to deny its grammaticality is the semantic one: the phrase actually says the very opposite of what it actually means. Unlike the previous examples from the table, (6b) is defended by several scholars, who make little of this semantic argument owing to the non-compositional nature of idioms. Among them is Stephen Pinker, who claims in *The Language Instinct* (p. 377) that *I could care less* is sarcastic, or John Lawler and Mark Lieberman who claim on their blogs that the phrase is, rather, a case of “negation by association”, where a construction retains its negative force without negation markers. While prosody might sustain these arguments – in some cases only, the notion of communicative efficiency is more important. In line with this, (6b) should be considered erroneous as it may create ambiguity and misunderstandings when used figuratively to mean the opposite of what its literal level or “surface structure” reads.

¹⁶ Naturally, both contracted and full forms were taken into account.

5. Phraseme genesis and selection criteria

The fact that illogical or erroneous variants can catch on and lexicalise no longer needs proving. It now needs explaining. The answer to this paradoxical phenomenon lies in the phraseme genesis process, which was described at length in Villers (2015, 2016, 2018).

Figure 4. The stages of phraseme genesis, adapted from proverb genesis



Source: own research.

In a nutshell, a stimulus (a specific situation) triggers the coinage of a phrase which in turn needs to be exposed to a sufficient number of hosts (speakers) – or vice versa – with the help of a propagating agent (a film, a website, a book, etc.). By repeating the phrase, the speakers perpetuate the cycle and allow the phrase to propagate in a virus-like pattern. This cycle may be broken into separate stages, inspired by Heylighen (1998) and its memetic¹⁷ approach: in order to spread, the phrase first needs to be noticed, understood, and accepted (assimilation). It then has to be remembered and

¹⁷ Memetics is the study of how cultural units (or “memes”, such as trends or customs) replicate by imitation. This discipline, which is the cultural equivalent of genetics, was heavily criticised for its lack of results and was soon buried, although breakthroughs came later (neurological evidence and concrete applications).

used through a medium (i.e. voice, ink, a signal, etc.). The chances of replication – and therefore of survival¹⁸ – are very low for most phraseme candidates, as the majority of potential hosts (speakers) do not repeat the cycle. In fact, the odds greatly depend on “selection criteria”. Heylighen (1998) describes these selection criteria for the replication of memes and cultural units in general, which were modified and adapted to phrasemes in Villers (2018). Thus, the main selection criteria in phraseme genesis and replication are:

- Novelty: a phraseme is likelier to replicate if it is perceived as being new, especially among younger age groups.
- Originality: the odds for replication are better if it has no direct “competition”.
- Stylistic markers: a phraseme that “stands out” from free combinations by signalling its preconstructed nature has more chances to be noticed, used, and remembered.
- Simplicity: a PU will be probably be used more often and will spread faster if it is easily pronounced or understood.
- Usefulness: a phraseme is likelier to catch on if it is applicable to many situations.
- Authority: a PU associated with a famous person or work will spread faster, even if that source is progressively forgotten.
- Conformity: a PU needs to be accepted in order to be repeated, which implies that it cannot go against the beliefs or knowledge of hosts.
- Publicity: most importantly, the phraseme candidate needs mass exposure in order to reach an optimal number of hosts (speakers). Publicity greatly influences its scope of dissemination (local, national, etc.).

Naturally, some criteria have more weight than others. In theory, the “conformity” criterion should prevent or limit the propagation of erroneous variants. In fact, it is in competition with other criteria. Some phraseological blunders might, for instance, be deemed more appealing due to their apparent novelty, at least among some speaker groups. Some might also be deemed simpler, as phonological reductions can make them easier to pronounce (e.g. 16b). Publicity may, of course, help a blunder “catch on”, since a mistake with a lot of exposure is likelier to be repeated, as with (17d). Even the “conformity” criterion might cause a host to favour an erroneous variant over a standard one, as with (5b), which might seem more logical to speakers who have never heard the word *scapegoat*. In other words, the whole process and its

¹⁸ The replication process is very selective and has a low-success rate, in a “survival-of-the-fittest” manner.

selection criteria leave room for error insofar as the “hosts” or “vehicles¹⁹” responsible for the replication of phrasemes are human speakers, who are prone to cognitive biases and may therefore not detect the deviation during the assimilation phase, or deem it acceptable.

6. Conclusion

What transpires from studies on phraseological errors is that scholars focus on specific categories of phrasemes (collocations) and certain categories of errors (blends and substitutions). However, the typology that was presented revealed that phrasemes may be accidentally modified in numerous ways, be it their form, their usage, or their meaning. Phraseological blunders are attributable to several core reasons: absence of mind and emotional state are the most common triggers for mistakes among experienced speakers and natives, while the lack or loss of linguistic competence is the most common reason behind phraseological errors among learners and people with language-related pathologies. The study of several recurrent phraseological blunders highlighted the difficulty of drawing a line between error and variation; a set of criteria was hence used to assess and classify them: frequency ratio between standard form and deviation, context (to filter out quotations and deliberate modifications), semantic coherence, and communicative efficiency.

The examples under study also demonstrated the possibility for “erroneous” variants to lexicalise, leading to what may be seen as a case of corrupt innovation. This seemingly paradoxical phenomenon was explained with a memetic approach of the process of phraseme genesis and propagation, during which cognitive biases are not always filtered out by the various selection criteria. Indeed, grammaticality, logic, or coherence are sometimes non-essential factors to human hosts – to deviate is human. This phenomenon may be compared to the propagation of fake news or rumours, where coherence and truth and not necessarily deemed important by the hosts who propagate them. Ultimately, the existence of phraseological blunders, some of which are recurrent, was to be expected – not only from a statistical perspective, but also owing to human nature.

¹⁹ This term is the one used in memetics, while “host” is the used in epidemiology. It corresponds to speakers.

Bibliography

- Aarts, Bas. 2007. *Syntactic gradience*, Oxford university press, Oxford.
- Bergstrom, Gustaf .1906. *On blendings of synonymous or cognate expressions in English*. Lund: Hakan Ohlsson
- Bolinger Dwight. 1961. Syntactic blends and other matters. *Language* 37, 366–381.
- Crain, Stephen & Thornton, Rosalind. 1998. *Investigations in Universal Grammar*, Cambridge: MIT.
- Cohen, Gerald. 1987. *Syntactic Blends in English parole*. Verlag Peter Lang.
- Fauconnier, Gilles & Turner Mark. 2002. *The Way We Think. Conceptual Blending and the Mind's Hidden Complexities*, Basic Books, New York.
- Gleason, Jean Berko & Ratner, Nan. 1993. *Psycholinguistics*, New York: Harcourt Brace Jovanich.
- Granger, Sylviane. 2004. Computer learner corpus research: current status and future prospects. In Connor, Ulla; Upton, Thomas. (eds.). *Applied Corpus Linguistics: A Multidimensional Perspective*. Amsterdam & Atlanta: Rodopi. 123–145.
- Heylighen, Francis. 1998. What makes a meme successful? Selection criteria for cultural evolution. *Proc. 15th Int. Congress on Cybernetics*, 418–423.
- Honeck, Richard. 1997. *A Proverb in Mind*, LEA, London.
- James, Carl. 1998. *Errors in Language Learning and Use*. London and New York: Longman.
- Kamhi, Alan G. 1988. A Reconceptualization of Generalization and Generalization Problems. *Language, Speech, and Hearing Services in Schools* v19, 304–13.
- Liashchova, Liudmila. 2018. Phraseological errors in a foreigner's talk discourse and their reasons. Europhras conference 2018 (Bialystok).
- Legallois, Dominique. 2013. Les greffes phraséologiques – ou quand la syntaxe se compromet. *Langages* n° 189, 103–120.
- Misri, Georges. 1987. *Le figement linguistique en français contemporain*, PhD diss., Université Paris V.
- Murphy, Patrick *et al.* 2013. Impairments in proverb interpretation following focal frontal lobe lesions, *Neuropsychologia*, n° 51, 2075–2086.
- Nesselhauf, Nadja. 2003. The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics* 24 (2), 223–242.
- Osborne, John. 2008. Phraseology effects as a trigger for errors in L2 English: The case of more advanced learners. In Meunier, Fanny; Granger, Sylviane. (eds). *Phraseology in Foreign Language Learning and Teaching*, Amsterdam/Philadelphia: John Benjamins, 67–83.
- Paquot, Magali. 2008. Exemplification in learner writing: A cross-linguistic perspective, in Granger, Sylviane; Meunier, Fanny (eds). *Phraseology: An Interdisciplinary Perspective*. Amsterdam and Philadelphia: John Benjamins.
- Polguère, Alain. 2007. *Soleil insoutenable et chaleur de plomb : le statut linguistique des greffes collocationnelles*, *Academia* (revised official version).

- Thiessen, Jenifer. 2008. The phraseological errors of French, German, and Spanish-speaking EFL learners: Evidence from an error-tagged learner corpus, In Associação de Estudos e de Investigação Científica do Isla-Lisboa (eds). *Proceedings from the 8th Teaching and Language Corpora Conference (TaLC8)*, 300–306.
- Van Lancker, Diana. 1990. The Neurology of Proverbs, *Behavioural Neurology*, 169–187.
- Villers, Damien. 2014. *Le Proverbe et les Genres Connexes*, Presses Académiques Francophones, Sarrebruck.
- Villers, Damien. 2015. Proverbiogenèse et obsolescence, la naissance et la mort des proverbes, *Proverbium : Yearbook of International Proverb Scholarship*, n° 32, 383–424.
- Villers, Damien. 2016. Proverbiogenesis: the phases of proverbial birth, *9th Interdisciplinary Colloquium on Proverbs – Proceedings, Tavira: International Association of Paremiology*, 369–380.
- Villers, Damien. 2018. *The cake is a lie*: witnessing the birth of a modern phraseme, In Filatkina, Natalia & Stumpf, Sören (eds). *Konventionalisierung und Variation*, Peter Lang: Berlin.
- Wang, Ying & Shaw, Philip. 2008. Transfer and universality: Collocation use in advanced Chinese and Swedish learner English. *ICAME Journal* 32, 201–228.

Gaffes phraséologiques: quand de nouveaux phrasèmes naissent d’erreurs

Résumé

Le présent article propose en premier lieu une typologie des erreurs phraséologiques, où le locuteur modifie accidentellement la forme, l’usage ou le sens standard d’un phrasème. Sont ensuite étudiées les causes principales de ces erreurs, telles que les troubles langagiers, la faible maîtrise de la langue ou le manque de concentration. Les notions de variation et d’erreur sont alors différenciées lors de l’étude d’exemples problématiques à l’aide d’une combinaison de critères tels que la fréquence proportionnelle, la cohérence sémantique ou l’efficacité communicationnelle. Malgré leur caractère erroné, il est indéniable que certaines combinaisons subissent une lexicalisation et se propagent en raison de divers biais cognitifs. C’est à travers le processus de phraséogenèse que les réponses à ce phénomène a priori paradoxal sont apportées, et plus précisément les critères de sélection qui influencent la dissémination ou « répliation » des phrasèmes.

PART III

VARIABILITY OF FIXED EXPRESSIONS

Chen Ting

Minsk State Linguistic University

Belarus

<https://orcid.org/0000-0001-5783-6739>

Types and Degrees of Variation in English and Chinese Phraseological Units

Abstract. The paper examines different classifications of phraseological units (PhUs) in English and Chinese and states that these languages reveal different structural-semantic groups of phraseological units. Universal reasons for variation in phraseological units are presented and their universal results are revealed (lexical, constructional, grammatical and pragmatic types of variation). The author argues that though PhUs allow a considerable amount and diversity of variation in every language, typologically different languages, like English and Chinese, exhibit specific tendencies in types of variation (for example, grammatical variation of PhUs is not characteristic of Chinese), and that variation in PhUs is more characteristic of English than Chinese. The model of potential degrees and levels variation of word combinations, including PhUs, is worked out.

Key words: *phraseological units, contrastive study, types and degrees of phraseological variation*

1. Introduction

It is well known that in contrast to free word combinations, phraseological units (PhUs) tend to be frozen in form and meaning not to allow changes. However, it is almost accepted today that phraseological units as multi-word structures are not as stable and fixed as it was believed earlier: they permit a considerable amount and a continuum of different types of variation, both in language and speech. As John Sinclair said, “fixed phrases” are not in fact fixed (Sinclair 1996: 83). Moreover, recent studies have illustrated that variation can occur even with nondecomposable idioms (Geeraert 2017: 80).

However, structural and typological specific features of different languages may seriously influence phraseological variation, promoting it or preventing from it.

The **aims of the research** are to examine the reasons for phraseological variation in a language, to identify the peculiarities of phraseological units in Chinese in comparison with English and to reveal specific features of PhUs variation in these languages.

These languages are not related genetically – they belong to different families (Sino-Tibetan and Indo-European families accordingly) and to different typological language groups (the isolating and inflected ones), and, in addition, the societies speaking these languages did not have close socio-cultural contacts in the past. That is why these languages are of special interest for linguists. It should also be mentioned that variation of PhUs in English has often been the subject of linguistic investigation but it has not been studied yet, to the best of our knowledge, in Chinese. Neither has it been the object of contrastive study yet.

The **material of the study** are 64 English and 199 Chinese phraseological units with a lexical component *stone* in English and 石 [shí] ‘stone’ in Chinese presented in reliable dictionaries and corpora.

2. Reasons for variation in phraseological units

There are three major reasons for PhU variation:

- **Cognitive reason**

PhUs are typically based on metaphors, and metaphors as mental images are not stable but easily modifiable, which also contributes to the variation of PhUs.

- **Subjective reason**

According to the systemic functional linguist M. Halliday (1994: 37), languages evolve as systems of “meaning potential” or as sets of resources which influence what the speaker can do with language in a particular social context. As meaning potential, a language constantly allows speakers to make choices from a set of options. The selection of a language item, including a phraseological unit, is based primarily on communicative needs. Phraseological units as language tools should also meet the creative and expressive purpose of speakers. Thus, the use of a PhU is largely subjective to the speaker who may use it as a ready-made unit or change the form and meaning of the original linguistic sign according to his/her pragmatic needs, and thus cause a PhU variation in different forms (e.g., lexical substitution or syntactic rearrangement of words).

- **Objective (structural) reason**

All PhUs are not totally opaque – all of them are still perceived to be multi-word structures. That is why they are all open to syntactic and conceptual analysis.

Speakers work with PhUs as they do with any other language material: they adapt them, combine them, change parts of them. Like any components of a language structure, or of any other structure in general, the components of a PhU can be replaced and changed, thus allowing lexical, constructional or grammatical changes.

In different PhUs, these changes may be different and happen to a different degree. Some phraseological units can accommodate interchangeable synonyms (or closely related words with different senses), for instance *at any rate – at all rates; at any cost – at all costs; burst into tears – burst into crying*. Some PhUs allow structural changes as in *to break the ice – the ice is broken*.

3. Phraseological units in English and Chinese

Though the reasons for phraseological changes are universal, their results in different languages may be different due to the language divergences in structure, origin, and cultural heritage. Even the types of phraseological units in English and Chinese are different, and that could also tell on the character of their variation.

3.1. Phraseological units in English

English phraseological units are exclusively varied in origin and nature, and this fact causes problems for classifying them according to a single criterion. Different scholars proposed a number of approaches to their classification (structural, semantic, contextual, functional, etc.) to embrace the diversity.

In the Soviet linguistics on the material of Russian and English, the following classifications of phraseological units were elaborated by Vinogradov, Smirnitkiy, Amosova, Kunin and others:

- 1) **Structurally** phraseological units may be viewed as *word-equivalent* as *to kick the bucket* ‘to die’ or *sentence-equivalent* as *Life is not a bed of roses* ‘something is not always good or easy’.
- 2) **Semantically** they may be:
 - *non-motivated, non-transparent* (such PhUs are usually referred to as idioms) as in the *wet blanket* ‘a bore’;

- *partially motivated* with limited semantic transparency as *to kill two birds with one stone* ‘to solve two problems with one action’, and
 - *fully motivated* if their interpretation does not cause any difficulties as in *hard as a rock* ‘very hard’.
- 3) **Contextually** phraseological units may be divided into:
- *idioms* where the meaning of each word in a PhU is unique and contextually dependent (*red tape* ‘bureaucracy’) and
 - *phrasemes* where some of the words in a PhU may be contextually dependent and some are used in their regular dictionary senses as in *small hours* ‘early hours’.
- 4) **Functionally** phraseological units may be divided into:
- *nominative* (*to kick the bucket* ‘to die’);
 - *communicative* (*Life is not a bed of roses* ‘something is not always good or easy’);
 - *nominative-communicative* (*to break the ice* ‘to begin’ = *The ice is broken*) and
 - *interjectional* PhUs (*Goodness gracious!*)

Though all these phraseological units are varied in meaning, structure, function, and origin some scholars believe that they share the following features: “stability of content and structure, replication in speech, semantic integrity, structure solidity, and visualization”.

3.2. Phraseological units in Chinese

In the Chinese linguistics, the term *phraseology* as the study of, first of all, *fixed* expressions, appeared only in 1950-ies under the influence of the Soviet linguistics. Classification of PhUs in Chinese is, however, different and takes into account peculiarities of this language.

According to Ma Guofan (1985: 78), one of the most well-known linguists in Modern China who contributed much to the development of the Modern Chinese phraseology, PhUs can be classified into the following three groups:

1) **惯用语** [guàn yòng yǔ] ‘locution’ – this type of a PhU, often word-equivalent and usually performing the function of a noun, is based on certain collocations, fixed through their repeated usage by groups of people and usually used as a *figurative* device, for example, **垫脚石** [diàn jiǎo shí] – ‘stepping stone’, lit.: ‘pad foot stone’, i.e., ‘a person or things used to advance one’s career’; **绊脚石** [bàn jiǎo shí] – ‘stumbling stone’, lit.: ‘stumble foot stone’, i.e., ‘obstacles’; **试金石** [shì jīn shí] – ‘touch stone’, lit.: ‘test gold stone’, historically it referred to a kind of black stone which was used to

examine the quality of gold; now it refers to any criterion with the help of which other things can be evaluated.

More obvious differences between English and Chinese PhUs are observed in other groups of Chinese PhUs reflecting peculiarities of the language syntax and culture. They are:

2) **成语** [chéng yǔ] – lit.: ‘ready-made language’. **Chengyu** are four-character Chinese phrases that are highly characteristic of Chinese. They may be both *word-like* and *sentence-like* phraseological units. These phrases or expressions go back into history, are concise, deep, incisive and insightful in meaning, have a fixed structure and orderly syllables. They are usually referred to as *idioms* by European linguists. Usually they are composed of four words, like **坚若金石** [jiān ruò jīn shí] – lit.: ‘hard like gold/metal stone’, i.e., ‘as hard as a stone’; **一石二鸟** [yì shí èr niǎo] – lit.: ‘one stone two birds’, i.e., ‘to kill two birds with one stone’; **匪石匪席** [fěi shí fěi xí] – lit.: ‘not stone, not mat’, i.e., ‘it cannot be turned like a stone, it cannot be rolled up like a mat’ – this phrase is used to show someone’s determination and loyalty, etc.

3) **歇后语** [xiē hòu yǔ], **xiē hòu yǔ**, or **allegorical sayings**, are also characteristic of Chinese. They are two-part allegorical folk sayings performing either nominative or communicative functions. The first part of *xiē hòu yǔ* is like a riddle to be solved and the answer to it lies in the second part of it. When pronounced, there should be a pause between the two parts, like in **茅坑里的石头 – 又臭又硬** – lit.: ‘the stone in a privy – hard and stinking’, i.e., about somebody who is too stubborn and not cooperative; another example is **鸡蛋碰石头 – 坐输** – lit.: ‘just like eggs hurled against stone – an obviously lost game’, i.e., about some unwise choice leading to failure.

There are also 3 types exclusively sentence-like phraseological units:

- **谚语** [yàn yǔ] ‘**proverbs**’ – they sum up the collective wisdom of the community, a popular truth or a moral lesson in a concise and imaginative way, for example, **墙倒众人推** ‘When a wall is about to collapse, everybody gives it a push’, i.e., everybody hits a man who is down. In China they mostly were passed on orally.
- **俗语** [sú yǔ] ‘**sayings**’ – they refer mainly to fixed multiword phrases in the *adjectival* function, **陈谷子烂芝麻** ‘something out of date’, **吹胡子瞪眼睛** ‘very angry or in a rage with somebody’, and the like;
- **引语** [yǐn yǔ] ‘**quotations**’, derive from literature or speeches of celebrities, like Darwin or Confucius.

From a stylistic point of view, *chengyu* and *quotations* are formal and mostly used in written form, while *xiehouyu*, *proverbs* and *locutions* are less formal and are frequently used in oral speech.

4. Types of variation in phraseological units in English and Chinese

As for types of variation observed in PhUs, scholars tend to distinguish four universal types: lexical, constructional, grammatical and pragmatic. They are usually studied as separate phenomena, but in actual utterances they may co-occur.

4.1. Lexical variation

There is a considerable amount of lexical variation within PhUs among nouns, verbs, adverbs and prepositions as their components in English and Chinese. The variation of one or even more lexical synonyms in a PhU usually does not change its semantic integrity, for example:

- (1) (as) *hard as iron/rock* (noun),
- (2) *kill/hit two birds with one stone* (verb),
- (3) *坚若如金石* 'as hard as/like stone' (conjunction),
- (4) *镂琢于(之)金石* 'carve/cut in (of) stone' (verb, preposition).

According to Moor (1998), lexical variation in English is reaching 40% of all phraseological changes. As for Chinese, lexical variation has not been studied yet. However, in our contrastive study, lexical variation of English PhUs with the component *stone* is observed in 25% cases, while in the Chinese PhUs with the component 石 'stone', it took place only in 16% cases.

4.2. Constructional variation

When the syntactic format of an English or Chinese PhU is changed due to passivization or shortening, we deal with its *constructional variation* as in:

- (5) *a rolling stone gathers no moss* (the full version of a PhU) – *a rolling stone* (a shortened version),
- (6) *a stone's throw away* (the full version) – *a stone's throw* (a shortened version),
- (7) *坚若金石* (the full version) 'as hard as stone' – *坚石* (a shortened version) 'hard stone',
- (8) *鸡蛋碰石头—自不量力* (the full version) 'just like eggs hurled against stone – an obviously lost game' – *以卵击石* lit. 'use egg hit stone', i.e., 'just like eggs hurled against stone' – about an unwise choice leading to failure.

In our material, the constructional variation in English is 14% and for Chinese, it is only 1,5%.

4.3. Grammatical variation

English grammar permits changes between the indefinite and the definite form of a noun, its single and plural forms, the active and passive form of a verb that may also be used in the present or past tense. This grammatical variation facilitates the contextual expression of the concrete idea of number, time and state more clearly, for example:

- (9) *a rolling stone – rolling stones*
- (10) *No one wants to break the ice, I guess I will be first. – Finally, the ice was broken, and people started talking.* (However, not all English verbal PhUs permit passivization, e.g., *to kick the bucket – *the bucket is kicked.*)

In Chinese, grammatical variation is not characteristic of PhUs due to the limited number of morphological forms of the grammatical categories there. The concept of number variation in PhUs is usually rendered by lexical means:

- (11) 一石两鸟 – 一石多鸟 ‘to kill two birds with one stone’– ‘to kill many birds with one stone’

So, grammatical variation in our Chinese material is not found, while in English according to Moor (1998), 14% of phraseological units have two or more variants of their main grammatical form.

4.4. Pragmatic variation

Pragmatic variation in PhUs may resemble lexical variation, but it occurs not due to close semantic relations of words in the language system which are called synonyms, but due to the intension of the speaker who substitutes the words belonging to different lexico-semantic groups performing in the course the same pragmatic function as in the following Chinese example:

- (12) 没有过不去的河/坎儿 ‘there is no **river/setbacks** that can’t be crossed’

River and *setbacks* are not synonyms, the words refer to different things but both of them in this context denote something that stops someone from going ahead.

In some cases, it is very difficult to distinguish lexical and pragmatic variation, especially if the words are used in their minor, or derived meanings, as in the example:

(13) *flat broke* – *stone broke*.

Here both the words *flat* and *stone* have the derived meaning ‘absolutely, completely’ and this meaning is registered in the dictionaries. That is why they may be regarded as synonyms and in these phraseological units one may state the presence of lexical variation. However, if we take into account only the main meanings of the words, the type of variation can be considered pragmatic one.

Another example of the similar situation is observed in the variants of the PhU:

(14) *break balls* – *break stones*.

The syntactic format and grammar of these constructions remain unchanged, and the varied components *balls* and *stones* in (14) are not synonyms in their main meanings. Yet, in the context, they are both used in their derived senses and may be interpreted as the case of pragmatic variation.

We have to admit, however, that in our material of phraseological units with the word *stone* or 石 [shí] ‘stone’, no obvious cases of pragmatic variation were determined.

So, in English and in Chinese, we observe lexical and constructional variation in some kinds of PhUs. Grammatical variation in our Chinese material is not found. On the whole, variation in PhUs is more characteristic of English than of Chinese.

5. Degrees and levels of variation of phraseological units in English and Chinese

The issue of degree of variation of PhUs is problematic and unsettled. First of all, the degree of phraseological variation may refer to *the quantitative aspect* of PhUs changes. It may refer to the total number of PhUs in a language undergoing variation in comparison with the number of absolutely frozen PhUs not permitting any changes at all. Such studies need a thorough corpora data analysis of PhUs alterations, and are the subject of future studies.

And then, the quantitative analysis of degree of PhUs variation may also refer to *the number* of the changed components in a certain PhU in comparison with the number of possible alterations there that the language permits and zero variation in absolutely frozen PhUs.

Consider as example a nominative English PhU *a stone’s throw*, which literally means ‘any distance that a person might throw a stone away’, i.e. ‘a very short distance’.

The canonical form of the English PhU *a stone's throw* is composed of three main elements – the deverbal noun *throw* used with the article *a* and the noun *stone's* as the modifier and the pivotal element of the PhU.

The acceptable language variation in this PhU may be:

- **lexical** (synonymic variation of the deverbal noun: *a stone's cast/throw*) (the synonymic change of 1 out of 3 components), and,
- **constructional** (like addition of an adverb after the PhU – 2 possible variations: *a stone's throw away/from*; addition of an intensifier before the PhU – 2 possible variations: *just/only a stone's throw*; addition of a preposition before the PhU or a: *within/at a stone's throw*); in case of addition the number of variations in the PhU grows to four possible alterations,
- in other nominative English PhUs, **morphological** changes may be observed, examples of which were given earlier in 4.2.

Pragmatic variation in speech of this PhU may be various and unpredictable as it depends on the speaker's intention.

In all the given examples of quantitative variations in the PhUs, the changes were not radical as they did not alter the semantics of the PhU and its core components. Such purely quantitative changes may be considered to be of the first, or **primary level**.

The **second, deeper level of variation** concerns *qualitative changes* of a phraseological unit.

The *first degree of qualitative level of variation* of a PhU is related to the alteration of the part-of-speech meaning of the whole PhU as in *a stone's throw* (n) → *stone-throwing* (adj) in the following sentence:

- (15) *The other situation was a large rioting crowd threatening troops at a stone-throwing distance.*

Still, the next, *second degree of qualitative level of variation* of a PhU observed in a nominative phraseological unit refers to the change of lexical meaning of the whole PhU when a new meaning is derived on the basis of the former one. It tends to be connected with paraphrasing involving additional constructional and grammatical changes. Thus, the PhU *to cast/throw the first stone* that has the meaning 'be the first to make an accusation (used to emphasize that a potential critic is not wholly blameless)' and Biblical allusion may have the following use:

- (16) *"No, I don't throw stones though I can't understand it," she said shrugging her shoulders.*

In this context, the construction *to throw stones* means ‘to hurl insults or criticisms (at someone or something)’. This syntactically and semantically changed construction is related to the original Biblical PhU *throw stones at smb*.

So, all in all, the following universal model of hierarchy of levels and degrees of a word-group variation may be postulated:

L0 – zero variation in completely frozen PhU,

L1 – first (quantitative) level of a PhU variation:

- a) *substitution of a word by its synonym* (lexical variation),
- b) *morphological change of a word* (grammatical variation),
- c) *addition or subtraction of a word (words)* (structural variation),

L2 – second (qualitative level) of a PhU variation:

- a) *alteration of a PhU in its part-of-speech meaning,*
- b) *change of the lexical meaning of a PhU* that may go along with its *paraphrasing involving constructional, lexical and morphological changes,*

L3 – third (almost unrestricted) level of variation in a free word group.

However, in different languages and even in different PhUs of the same language, the realization of this potential PhUs variation model may be different.

6. Conclusion

On the basis of the conducted analysis, we may state that English and Chinese reveal different structural-semantic groups of phraseological units but each of them undergo variations. Variations in PhUs may occur due to cognitive (related to the nature of cognition), structural (related to the language structure), and subjective (pragmatic, related to the speaker’s intention) reasons.

Phraseological units in the studied languages demonstrate various types of variation (lexical, grammatical, constructional, and pragmatic), lexical variation being the leading type in both the languages. In Chinese, it is more characteristic of colloquial short phrases – *locutions* and also of *proverbs* and *sayings*. Two-part allegorical sayings with a riddle and an answer inside (*xiē hòu yǔs*) tend to admit in speech only constructional variation by shortening. Yet, when reference is made only to the first part containing a riddle, some lexical components may be added there, too. Phraseological variations are not characteristic of four-character rhythmic *chengyu* groups, going back into long history, and of quotations, where conventionalization is very high. In English, we observe all the 4 types of variation while in Chinese we have

found only 3 of them: due to the lack of morphological forms of words we did not find in our material grammatical changes of PhUs.

The potential PhU variation model is suggested which takes into account different degrees of quantitative and qualitative phraseological changes on different levels from frozen phrases on the top to free phrases on the bottom.

Corpora

corpus.byu.edu. <https://corpus.byu.edu/>

国家语委现代汉语语料库 [Chinese National corpus]. <http://www.cnccorpus.org/>

北京大学中国语言学研究中心 [Center for Chinese Linguistics PKU (Peking University)]. <http://ccl.pku.edu.cn/index.asp>

Literature

Atkins, Beryl T; Rundell Michael. 2008. *The Oxford Guide to Practical Lexicography*. New York: Oxford University Press.

Cameneva, Zinaida; Bushnaq, Tatiana. 2008. Variation and Flexibility within Idiomatic Expressions. *Annales Universitatis Apulensis, Series Philologica*. 9: 107–114.

Fraser, Bruce. 1970. Idioms within a Transformational Grammar. *Foundations of Language*. 6: 22–42.

Geeraert, Kristina et al. 2017. Understanding Idiomatic Variation. In: *Proceedings of the 13th Workshop on Multiword Expressions*. Spain: Association for Computational Linguistics. 80–90.

Gerald, Gazdar et al. 1985. *Generalized Phrase Structure Grammar*. Cambridge, Mass: Harvard University Press.

Halliday, Michael; Matthiessen Christian. 2001. *An Introduction to Functional Grammar* (3rd ed.). London: Edward Arnold.

Heinonen, Tarja Ritta. 2007. Variation and Flexibility within Verb Idioms in Finnish. In: Marja, Nenonnen; Sinikka, Niemi (eds.). *The First Nordic Conference on Syntactic Freezes*. Joensuu: University of Joensuu. 146–159.

Moon, Rosamund. 1998. Frequencies and Forms of Phrasal Lexemes in English. In: A.P. Cowie (ed.) *Phraseology: Theory, Analysis, and Applications*. Oxford: Clarendon Press. 79–100.

Senaldi, Marco Silvio Giuseppe. 2015. *Casting Light on Idiom Flexibility: A Corpus-based Approach*. Ph.D. diss. University of Pisa. Pisa, Italy.

Sinclair, John. 1996. The Search for Units of Meaning. *Textus: English Studies in Italy*. 9: 75–106.

Лешева, Л. М. 2016. Лексикология английского языка = English Lexicology. Минск: Minsk State Linguistic University.

- 中国社会科学院语言研究所词典编辑室 dictionary editorial room, institute of languages, Chinese academy of social sciences. 2002. *现代汉语词典 (Modern Chinese Dictionary)* 北京: 商务印书馆
- 黄伯荣 Huang Borong; 廖序东 Liao Xudong. 1991. *现代汉语 (Contemporary Chinese language) (上册)*. 北京: 高等教育出版社出版.
- 马国凡 Ma Guofan. 1985. *成语 (Idioms)*. 呼和浩特: 内蒙古人民出版社.
- 邵敬敏 Shao Jingmin. 2007. *现代汉语通论 (Introduction to modern Chinese language)*. 上海: 上海教育出版社.
- 许肇本 Xu Zhaoben. 1980. *成语知识浅谈 (An elementary introduction to knowledge of idioms)*. 北京: 北京出版社.
- 张斌 Zhang Bin. 2002. *新编现代汉语 (New modern Chinese language)*. 上海: 复旦大学出版社

Tipos y grados de variación en unidades fraseológicas inglés y chino

Resumen

El análisis realizado nos permite afirmar que el inglés y el chino revelan diferentes grupos semánticos estructurales de unidades fraseológicas, pero cada uno de ellos experimenta variaciones. Las variaciones en las unidades fraseológicas pueden ocurrir debido a razones cognitivas (relacionadas con la naturaleza de la cognición), estructurales (relacionadas con la estructura de la lengua) y subjetivas (pragmáticas, relacionadas con el hablante).

Las unidades fraseológicas en los idiomas en cuestión demuestran varios tipos de variación (léxico, gramatical, constructivo y pragmático), siendo la variación léxica el tipo principal en ambos idiomas. En chino la variación léxica es más característica de cortas frases coloquiales: locuciones y también de proverbios y refranes. Los dichos alegóricos de dos partes que son un acertijo y su solución (*xiē hòu yǔs*) generalmente admiten en el discurso solo una variación constructiva por acortamiento, cuando se hace referencia solo a la primera parte que contiene un acertijo, aunque también se pueden agregar algunos componentes léxicos. Las variaciones fraseológicas no son características de los grupos rítmicos de cuatro caracteres *chengyu*, que se remontan a una larga historia, y de citas, donde la convencionalización es muy alta. En inglés observamos los 4 tipos de variación, mientras que en chino hemos encontrado solo 3 de ellos: debido a la falta de formas morfológicas de palabras no encontramos en nuestro material cambios gramaticales de unidades fraseológicas.

Se sugiere un modelo potencial de variación de unidades fraseológicas que toma en cuenta diferentes grados de cambios cuantitativos y cualitativos en diferentes niveles, desde frases congeladas en la parte superior hasta frases libres en la parte inferior.

Tomáš Jelínek

Charles University

Czech Republic

<https://orcid.org/0000-0002-8521-4715>

Multi-word Lexical Units with Repetition of Lexemes in Czech and Identification of Their Variants

Abstract. A study of the variability of multi-word lexical units with repetition of lexemes such as *Bůh dal, Bůh vzal* ‘the Lord has given, the Lord has taken away’ based on a large corpus is presented. Four case studies illustrate the extent of variability of such expressions. A database of multi-word expressions is described, with a special attention to expressions with repetition and their variability. The automatic identification of multi-word expressions with repetition in Czech texts is also explained.

Key words: *multi-word expressions, repetition of lexemes, variability, automatic identification of MWE in text*

1. Introduction

The variability of multi-word lexical units with repetition of lexemes is an interesting case of reproducibility of multi-word expressions which are recognized not only on the basis of a particular word or group of words but on the basis of their syntactic structure and the fact that a lemma is repeated in such a structure as well. For example, for the Czech saying *Bůh dal, Bůh vzal* ‘the Lord has given, the Lord has taken away’, we found many cases where the word *Bůh* ‘God’ is replaced by another word occurring twice in the expression, such as *stát dal, stát vzal* ‘the state has given, the state has taken away’. Such variants are easily recognized as referring to the original multi-word expression by native speakers. In Czech, we have found about 90 such multi-word lexical units so far, with higher or lower variability. Some of these multi-word expressions are simple binomials such as *den za dnem* ‘day after day’, *knih knih* ‘the book of (all) the books’ Čermák (2007: 414–

429), some are more complex. The phenomenon of repetition in language has been studied extensively, see for example Fisher (1995) or Frédéric (1985).

In this paper, we describe the variability of these multi-word lexical units in four case studies and show how we address this variability in the database of multi-word lexical units LEMUR we have developed (see below). We also explain automatic identification of variants of these multi-word expressions in texts.

2. Data sources

In order to study the variability of multi-word lexical units in use, large corpora are needed. Multi-word expressions (MWE) occur far less frequently than most single-word lexical units, even in their invariant form; finding their variants in individual texts or in smaller corpora is very difficult. In order to investigate the variability of multi-word lexical units, we need really large corpora containing billions of tokens. The above mentioned saying *Bůh dal, Bůh vzal* ‘the Lord has given, the Lord has taken away’ appears, for example, only 4 times in three different variants of one single type, in the corpus SYN2015, developed by Křen et al. (2015), which contains one hundred million words. In the SYNv6 corpus, compiled by Křen et al. (2017), which contains 4.8 billion tokens, we find 80 occurrences of this multi-word expression with significantly richer variability. 80 occurrences are perhaps not enough to base a research on, but at least it gives us a better picture of how this MWE is used and what the extent of its variability is.

The research presented in this article is therefore based on the SYNv6 corpus of written contemporary Czech. This corpus consists of 4.8 billion text positions, i.e. 4 billion words not including punctuation. The corpus is not balanced, the majority of texts are journalistic texts because of easier accessibility of this type of text in electronic form. It does not contain texts originating from the Internet (due to the higher proportion of non-standard language in the texts of this provenance).

3. Case studies

In this section, we present four Czech multi-word lexical units: *Bůh dal, Bůh vzal* ‘the Lord has given, the Lord has taken away’; *čistému vše čisté* ‘to the pure, everything is pure’; *podobný/podobat se jako vejce vejci* ‘as alike as two peas in a pod, lit. similar as an egg to an egg’; *já na bráchu, brácha na mě* ‘you scratch my back, I’ll scratch yours, lit. me to brother, brother to me’. The first three case studies will show how diverse the lexical variability found

in the corpus for this type of MWEs is, the last case is interesting because of its structure and the possibility of replacing all autosemantic words in the multi-word lexical unit.

3.1. **Bůh dal, Bůh vzal**

The expression *Bůh dal, Bůh vzal (jméno Boží budiž požehnáno)* ‘The Lord has given, the Lord has taken away (the name of the Lord be blessed)’ is of biblical origin, coming from the book of Job (Job, 1:21). It expresses reconciliation with some loss. A typical use of this phrase is illustrated by example (1) from the corpus:

- (1) „Zřejmě nakoupím nové vybavení a včelstva,” nevzdává se postižený včelař Luboš Machatý. Ztrátu chovu bere optimisticky: „Je to příroda. **Bůh dal, Bůh vzal.**”

“Obviously, I will have to buy new equipment and honeycombs,” says the affected beekeeper Luboš Machatý. He takes the loss of beehives optimistically: “That is nature. God has given, God has taken away.”

There are 80 occurrences of this MWE in the SYNv6 corpus. This saying is very variable, only in 40% (27 occurrences) we find it in its original form with the noun *Bůh* ‘God’. The reproducibility of this MWE is based on the use of the verbs *dát* ‘to give’ and *vzít* ‘to take away’ (usually in past tense) and on a structure with the repetition of the noun in the nominative case. The structure of the variable MWE can be expressed by the formula:

N_{nom.a} *dát*_past.part., N_{nom.b} *vzít*_past.part (lemma_a = lemma_b).

The verbs *dát* ‘give’ and *vzít* ‘take away’ are usually in past tense, we find only two occurrences with a different tense (*stát dal, stát vezme* ‘the state gave, the state will take away’). The past tense (expressed by the past participle of the verb) agrees in gender and number with the subject; in most cases (70 out of 80) the subject is masculine animate, the remaining cases are feminine (*voda dala, voda vzala* ‘water gave, water has taken away’), or plural (*politici dali, politici vzali* ‘the politicians gave...’). Exceptionally, in four cases, we find an inversed order of the verbs (*banka vzala, banka dala* ‘the bank took away, the bank gave’).

The greatest variability of this MWE is related to the pair of nouns in the nominative case with the same lemma. In 58 cases (72% of all occurrences), these nouns are lexical variations of the word God: *Bůh* ‘God’, *Pánbůh* ‘the Lord God’, *pán* ‘Lord’, *Hospodin* ‘Lord’, *Alláh* ‘Allah’; in 13 cases, it is institutions: *stát* ‘state’, *televize* ‘television’, *Apple*, *ČEZ* (a Czech power company);

in 5 cases influential people: *politici* ‘politicians’, *bolševik* ‘Bolshevik’, *gosudar* ‘tsar’, *sudí* ‘jury’, *Jandák* (a Czech politician); the remaining 4 cases are impersonal forces: *příroda* ‘nature’, *voda* ‘water’, *kostka* ‘cube’, *náhoda* ‘coincidence’. In one single case the saying occurs without the lemma repetition (i.e. the reproducibility of MWE is based only on the structure, verbs and the meaning of the words): *car dal, sověty vzaly* ‘the tsar gave, the Soviets took away’.

In 7 cases, the MWE is followed by the phrase *jméno Boží budiž požehnáno* ‘the name of God be blessed’ in several variants. Otherwise, no syntactic variability has been found; for example, we have not found any occurrence of an attribute modifying the noun.

3.2. Čistému vše čisté

The saying *čistému vše čisté* ‘to the pure everything is pure’ also comes from the Bible (Titus 1:15); it means that honest men also consider others to be honest, which may be naïve. In the SYNv6 corpus, there are 137 occurrences of this MWE. It is used as in example (2):

- (2) *Yoko Ono a John Lennon se k sexu stavěli velice otevřeně, přesně podle rčení „čistému vše čisté“.*

Yoko Ono and John Lennon have been very open about sex, precisely by the saying “to the pure, everything is pure”.

The structure of the variable MWE can be expressed by the formula:

A_{nom.a} vše_{nom}. A_{dat.b} (lemma_a = lemma_b).

There is significantly less variability than for the previous saying, 118 occurrences (86% of all occurrences) are in the invariant form *čistému vše čisté*; there are 5 occurrences of a variant *hebkému vše hebké* ‘to the soft, everything is soft’ which is a name of a musical performance cited by the journals; we find 3 occurrences of *mrtvému vše mrtvé* ‘to the dead everything is dead’; each of the other 11 variants appears only once: *živý* ‘alive’, *hloupý* ‘stupid’, *chytrý* ‘smart’, *dobry* ‘good’, *velký* ‘big’, *nízký* ‘low’, *tajný* ‘secret’, *sladký* ‘sweet’, *česnekový* ‘garlicky’, *bílý* ‘white’, *ruský* ‘Russian’. Seven times the saying occurs in the variant with the verb *být* ‘to be’ (only with the adjective *čistý* ‘pure’), as in example (3):

- (3) *Některé zvěsti, které kolují o Hrabalovi, věru nepatří do čítanek (ale čistému je vše čisté).*

Some rumors about Hrabal do not belong to textbooks (but to the pure, everything is pure).

3.3. *Podobát se / podobný jako vejce vejci*

The third of our case studies presents a fixed comparison: *podobat se / podobný jako vejce vejci* ‘as alike as two peas in a pod’, lit. ‘resemble/similar as an egg to an egg’. The meaning of this comparison is that two objects or persons are very similar, almost indistinguishable. The reflexive only verb *podobat se* ‘resemble’ or adjective *podobný* ‘alike’ can occur in the vicinity of the words *jako vejce vejci* ‘as an egg nom. an egg dat.’ (which have a fixed order), not necessarily just in an adjacent position (in the corpus, we were searching for cases where *podobat se* or *podobný* occur at a distance of maximum ten tokens to either side, regardless of sentence boundaries). Typical use is shown in example (4). 2180 occurrences of this MWE and its variants appear in the SYNv6 corpus.

- (4) *Webové stránky mnohých fakult jsou si podobné jak vejce vejci, přitom studium na nich je naprosto odlišné.*

The web sites of many faculties are similar as an egg to an egg, while studying at them is totally different.

The structure of the variable MWE can be expressed by the formula:

podobný/podobat se... jako/jak N_{nom.a} N_{dat.b} (lemma_a = lemma_b).

Whereas the lexical variability in this comparison is very small compared to the above-described expressions, the frequency in the corpus is, on the contrary, far higher. Out of the total number of 2180 occurrences, 2166 (99.4%) are examples of the invariant form (*podobat se / podobný ... jako vejce vejci*), with a uniform division between the verb *podobat se* ‘to resemble’ (1129) and the adjective *podobný* ‘alike’ (1046). There is a strong preference for the longer form of the conjunction *jako* ‘as’ (2050), compared with the shorter form *jak* ‘as’ (116).

Apart from the noun *vejce* ‘egg’, a number of other nouns repeated twice appear in this comparison, but with a very low frequency. There are altogether only 14 such occurrences in the SYNv6 corpus. The most common is *podobný ... jako oko oku* ‘alike as an eye to the eye’ (3 occurrences); then there are objects that appear in large numbers: *tráva* ‘grass’, *kapka* ‘drop’, *hvězda* ‘star’; animals: *pták* ‘bird’, *vrána* ‘crow’, *vlk* ‘wolf’, *blecha* ‘flea’; people: *dvojče* ‘twin’, *Číňan* ‘Chinese’ (probably because the Asians are difficult to distinguish for the Czechs), *Hitler* (used in a context about a court battle between two companies using a similar design resembling Hitler) and *telenovela* ‘soap-opera’ (used in a context pointing to small differences among TV shows).

For the basic form ... *jako vejce vejci*, 6 examples of a noun attribute modification can be found in the corpus, in five occurrences the same adjective is repeated, twice *dračí vejce* 'dragon egg', once *pštrosí vejce* 'ostrich egg', *kukaččí vejce* 'cuckoo egg', *zlaté vejce* 'golden egg'.

In the case of (5), a single such occurrence in the corpus, the adjectives differ in a contextually dependent use of the original fixed comparison where the complete similarity of compared objects is modified by unequal wealth (*chudé vejce* 'poor egg' vs. *bohatší vejce* 'richer egg').

- (5) *Pořady České televize se většinou podobají těm na soukromých stanicích jako chudé vejce bohatšímu vejci.*

Czech TV shows are usually similar to those in private stations as poor eggs to richer eggs.

Exceptionally (also only once in the whole four-billion word corpus), we find a different variant of the fixed comparison *podobný jako vejce vejci* in which only the first part is used to recognize the MWE, the second part is changed and lemmas are not repeated (6):

- (6) *Ten mladík tam je mi podobný asi jako vejce hromádce kuřecích kostí.*

The young man there is similar to me as an egg to a bunch of chicken bones.

3.4. Já na bráchu, brácha na mě

The last of our MWE case studies *já na bráchu, brácha na mě* 'you scratch my back, I'll scratch yours', lit. 'me to brother, brother to me' is an example of a complex MWE where its reproducibility is based primarily on lemma repetition and morphosyntactic structure, there are no autosemantic words that could not be changed in a variant. The saying is used as a negative assessment of cases where people (typically powerful) help each other, usually in an unethical or illegal way (cronyism), as in example (7). There are 521 occurrences of this MWE in the SYNv6 corpus (not including fragments).

- (7) *Máme tady z toho dojem klientelismu – pověstného já na bráchu, brácha na mě. Je to vidět na firmách, které v tomto obvodu vítězí ve veřejných zakázkách.*

There is an impression of cronyism – the proverbial you scratch my back, I'll scratch yours. It manifests itself with the companies that are winning in public procurement in this area.

The structure of this MWE is expressed by the formula:

P/N_{nom.a} *na* N_{acc.b}, N_{nom.c} *na* P/N_{acc.d}
 (lemma_a = lemma_d & lemma_b = lemma_c).

In the invariant form, the personal pronoun forms *já, mě* 'I, me' is repeated at the first and last position, the noun *brácha* 'bro, brother' (a colloquial form of the word *brother*) is found following the first preposition *na* 'on, to' and also in front of the second preposition *na*. In some variants, both the noun *brácha* and the pronoun *já* can be replaced by a noun (denoting a person). What is especially interesting about these variants, is the fact that with such a substitution, only two prepositions *na* 'on, to' and the parallel morphosyntactic structure (a noun in nominative – preposition *na* – a noun in accusative – comma – a noun in nominative – preposition *na* – a noun in accusative) are taken over from the original multi-word expression, yet it is recognized in an appropriate context without difficulty. Between the two parallel nominal groups, there is usually a comma, but the conjunction *a* 'and' or a dash can be there as well.

In total, 521 of these multi-word lexical units (with the whole structure, see below) occur in the SYNv6 corpus out, of which 464 occurrences (89%) appear in the basic, invariant form, 53 variants occur (10%) with the noun *brácha* being replaced (E.g.: *já na soudruha, soudruh na mě* 'me to comrade, comrade to me') and 4 occurrences can be found with both positions replaced by another noun. In case only the noun *brácha* is replaced (53 occurrences), the replacing noun often denotes another family member such as *ségra* 'sis, sister', *žena* 'woman, wife', *táta* 'dad', *děda* 'grandpa', *švára* 'brother-in-law', *kmotr* 'godfather' (22 occurrences in total) or friends: *kamarád* 'friend' (masc.), *kamarádka* 'friend' (fem.), *kámoška* 'friend' (fem., colloquial); the other occurrences are mostly influential persons and institutions: *ministr* 'minister', *vláda* 'government', *polda*, 'cop, policeman', or proper names (of politicians, etc.) as in example (8). One variant of this MWE appears 12 times: *já na Háchu, Hácha na mě*. It is a title of a historical theater play about *Hácha*, a Czech politician, the title is a word game using paronymy between the nouns *brácha* and the name *Hácha*.

- (8) *Já na Obamu, Obama na mě! (...) Já na bráchu, brácha na mě! Jak se zdá, tohle heslo výborně zafungovalo i v případě netradičního spojení zpěvačky Beyoncé Knowles (27) a příštího amerického prezidenta Baracka Obamy (47).*

Me to Obama, Obama to me! (...) Me to brother, brother to me! This slogan seems to have worked well in the case of the unconventional alliance of singer Beyoncé Knowles (27) and the next US president Barack Obama (47).

In all four cases where both the pronoun *já* and the noun *brácha* were replaced, the replacing nouns were always the names of politicians as in example (9):

- (9) *Klaus na Majora, Major na Klause. Český a britský premiér v sobě našli zalíbení.*

Klaus on Major, Major on Klaus. The Czech and British Prime Ministers have found mutual sympathy.

However, not all results matching the abovementioned formula are variants of the MWE *já na bráchu, brácha na mě*. It is necessary to carefully verify that there are no other reasons for using two parallel constructions with the preposition *na*: sometimes, this preposition is used as related to a verb, adjective or noun with valency *na* + accusative (frequent in Czech) as in example (10), where the use of the preposition *na* is motivated by the noun *žaloba* ‘lawsuit’ and the repetition of lemmas is due to the mutual relationship of both persons (*Zelníček* and *Vovsík*).

- (10) (...) *do souboje, v němž se nyní žalobou ohání Zelníček na Vovsíka a Vovsík na Zelníčka.*

(...) into a duel in which Zelníček Vovsík and Vovsík Zelníček now threaten each other with a lawsuit.

Apart from the abovementioned variants, the saying *já na bráchu, brácha na mě* presents yet another type of variability: the use of a fragment, in particular the first part of the MWE *já na bráchu* ‘me on brother’ (often with the punctuation mark of ellipsis indicating that the MWE is not complete). The fragment sufficiently represents the whole expression as in example (11). Such occurrences can be found in the examined corpus 180 times. The saying is even frequently condensed into one noun *jánabráchismus*, used as a less formal synonym for *klientelismus* ‘cronyism’; in the corpus we find 168 such occurrences.

- (11) *Firma, kterou prosadil starosta, udělala s městem obchod a vydělala za rok bezmála dvacet milionů. Obvyklý obchod stylem „já na bráchu...”.*

The company promoted by the mayor made a deal with the city and earned nearly twenty million in a year. The usual business style “me on brother...”.

4. LEMUR, a database of Czech multi-word expressions

The case studies presented in the third part of this paper illustrate the variability of multi-word lexical units with the repetition of lemmas and, more generally, the variability of Czech multi-word expressions. So as to record this variability and to make it accessible to users for both their own study and the use of language processing tools, we have developed a database of multi-word lexical units LEMUR (lexicon of multi-word expressions). Currently it contains about 5000 MWEs: sayings, proverbs, weather lore, fixed comparisons, multi-word prepositions etc., hundreds of which were manually annotated in detail. The database contains a range of information about each of the multi-word lexical units, including their basic form, definition, syntactic structure, example of use, variability, idiomaticity, etc., as described in other articles, especially by Hnátková et al. (2017).

The database records the variability or fixedness of each MWE on the levels of word forms, word order, syntax, and lexicon. Generally, the database assumes that the modifications (like word order changes due to topic-focus articulation, passivizations or nominalizations) typical for the same Czech constructions are possible for any MWE, unless explicitly stated otherwise in the database.

4.1. Variation and fixedness in the LEMUR database

The question of word form variation or fixedness relates only to some MWEs in which there is no free choice of word forms with the same grammatical categories. For example, in the MWE *podle nosa poznáš kosa* 'someone's character can be recognized from her/his face', lit. 'after the nose you recognize the blackbird', the word form *nosa* is an unusual form of genitive singular of the noun *nos* 'nose', used almost exclusively in this expression. No other form of the noun *nos* can be used here (because of the rhyme). Otherwise, we assume a free variability of word forms (for example, in the expression *já na bráchu, brácha na mě* 'you scratch my back, I'll scratch yours', the accusative form *mě* 'me' of the pronoun *já* 'I' can be replaced with another equivalent form *mne* 'me', such examples were found in the corpus).

Similarly, the MWE database contains expressions for which the order of components cannot be changed, or no other word can be inserted between the components (e.g. the word order in the saying *čistému vše čisté* 'to the pure everything is pure' must be preserved, and no word can be inserted between the components, except for the verb *být* 'to be': *čistému je vše čisté*).

We also note whether there are any limitations of syntactic modification or syntactic variability, e.g. in case a verbal MWE cannot be passivized or nominalized or if it is not possible to modify a component of a MWE by another word. For example, in the MWE *Bůh dal, Bůh vzal* ‘the Lord has given, the Lord has taken away’, no syntactic variability is possible: the MWE cannot be passivized or nominalized, none of the components can be modified by other sentence members.

The lexical variability within the MWE is, on the contrary, explicitly defined. If a component is partially or completely lexically variable, a list of possible lemmas is recorded or other restrictions (if any) are noted in the database entry.

If a given MWE can be represented by a fragment, that is, a part of the multi-word lexical unit that identifies it and is used independently, as in example (11): *já na bráchu...*, these options are also described.

4.2. A database entry

In the database entry, which is used both for editing records and (at least so far) for viewing, a “slot” is defined for each MWE component. This slot represents a position in the MWE that can be filled with either a fixed or variable lemma. For example, the entry for the MWE *čistému vše čisté* ‘to the pure everything is pure’ is represented by four slots:

1. *čistému*: in the invariant form, the adjective *čistý* ‘pure’ in dative masculine singular; otherwise any adjective in the dative case.
2. *je*: optional verb *být* ‘to be’ in the third person of the present tense.
3. *vše*: nominative singular of the pronoun *všechno* ‘all’: forms *vše* or *všechno*.
4. *čisté*: adjective in nominative singular neuter, the same lemma as in the first slot.

For the whole MWE, the impossibility of word order changes, insertion of words or component modification is recorded.

5. Automatic identification of multi-word units with repetition of lemmas

In order to study multi-word lexical units in real use, it is not only necessary to have an extensive database of MWEs, but such expressions have to be identified in texts (in corpora), marked and linked to the database. For automatic annotation, we use (for now) a modified version of the system FRANTA, described by Kopřivová & Hnátková (2014), which identifies

MWEs based on word forms, lemmas and morphosyntactic tags. When any MWE is found, it is assigned a special MWE lemma. Via these lemmas, the MWEs in texts are linked to the LEMUR database.

In most cases, multi-word lexical units with lemma repetition are identified only in their basic, invariant form; moreover, some more frequent lexical variants of such MWEs are also identified. A more general approach concerning the identification of a morphosyntactic pattern and testing for the lemma identity is very costly in terms of computer resources (it is slower by at least two orders of magnitude), therefore new variants are identified solely in a few frequent cases comprising important lexical variability.

In the future, we are planning to develop new software that would directly use the export of information from the database, so it would not be necessary to link database entries with lemmas manually entered into the FRANTA system. The FRANTA system is not flexible enough and needs to be manually edited when adding new MWEs.

For more complex MWEs, it will probably be always necessary to concentrate only on documented lexical variants. A typical example of such complex MWE is *já na bráchu, brácha na mě* (see part 3.4), in which all autosemantic words can be replaced, an automatic query for any such variants would have to be based only on a morphosyntactic pattern, a frequent preposition and a test of lemma repetition. Moreover, the morphosyntactic pattern (N_{nom.} na N_{acc.}, N_{nom.} na N_{acc.}) is not sufficiently distinctive, and may be motivated by the valency of another nearby word as in example (10), which does not exemplify the given MWE. In order to identify variants of this multi-word lexical unit with both lemmas replaced, it is therefore necessary to understand the text and thus to check it manually.

At present, we are unable to search automatically for new, not yet identified multi-word lexical units with lemma repetition in general (e.g. using association measures and testing for lemma repetition). The best way is to identify a new MWE using standard methods (association measures etc.), check it manually and, if there is a repetition of lemmas in the MWE, search for lexical variability preserving lemma repetition.

6. Conclusions

The variability of Czech multi-word lexical units with repetition of lemmas was presented. This type of MWE is a phenomenon deserving more detailed study: it is interesting to observe lexical variability related to MWE reproducibility, i.e. whether a multi-word expression is recognizable after

replacing the original repeated lemmas or to record what the extent of such variability is and how often such replacement occurs. MWEs of a similar character appear in many European languages, some expressions are used in many languages (*crows will not pick out another crow's eyes; corbeaux avec corbeaux ne se crèvent jamais les yeux; vrána vráně oči nevyklove; вóрон вóрону глаз не вбќлюет*), others are language specific (*a friend in need is a friend indeed; il faut manger pour vivre, et non pas vivre pour manger; дружба дружкой, а слўжба слўжкой*), so it is possible to examine the variability of similar multi-word expressions across languages.

We have shown that the variability of individual MWEs can be defined and recorded in a database that is both human-readable and computer-readable. Linking corpora in which MWEs are labeled and lexical databases containing a detailed description of such expressions will allow for a more precise and easier research of multi-word lexical units.

Acknowledgement

This paper is part of the project *Between Lexicon and Grammar* (2016–2018), supported by the Grant Agency of the Czech Republic, reg. no. 16-07473S. This project is a follow-up of the project entitled *The Grammar-Based Treebank of Czech* (2013–2015, cf. Skoumalová et al. 2014; Petkevič et al. 2015a, 2015b) and devoted to automatic parsing driven by a formal HPSG-like grammar of Czech.

Bibliography

- Čermák, František. 2007. *Frazeologie a idiomatika česká a obecná*. Prague: Karolinum.
- Fischer, Andreas. 1994. *Repetition*. Günter Narr Verlag, Tübingen.
- Frédéric, Madeleine. 1985. *La répétition: Étude linguistique et rhétorique*. Tübingen: Max Niemeyer Verlag.
- Hnátková, Milena et al. 2017. Eye of a Needle in a Haystack. Multiword Expressions in Czech: Typology and Lexicon. In: Mitkov, Ruslan (ed.). *Computational and Corpus-Based Phraseology: Second International Conference, Europhras 2017*, London, UK, November 13–14, 2017, Proceedings. Springer: Berlin–Heidelberg, p. 160–175.
- Kopřivová, Marie; Hnátková, Milena 2014. *From Dictionary to Corpus. Phraseology in Dictionaries and Corpora*. Maribor, Filozofska fakulteta Maribor, p. 155–168.
- Křen, Michal et al. 2017. *Corpus SYN, version 6 from 12/18/2017*. Prague: Institute of Czech National Corpus. <http://www.korpus.cz>.
- Křen, Michal et al. 2015. *Corpus SYN2015: a representative corpus of contemporary written Czech*. Institute of Czech National Corpus FF UK, Prague 2015. <http://www.korpus.cz>.

Wielowyrzowe jednostki leksykalne z powtórzonymi leksemami w języku czeskim i identyfikacja ich wariantów

Streszczenie

Niniejszy artykuł przedstawia wyniki badania zmienności jednostek wielowyrzowych z powtórzonymi leksemami, takich jak *Bůh dal, Bůh vzal* 'Bóg dał, Bóg wziął', opartego na dużym ilościowo korpusie. Cztery studia przypadków ilustrują zakres zmienności tego rodzaju połączeń wyrazowych. Opisano bazę danych związków wyrazowych, ze szczególnym uwzględnieniem ustabilizowanych konstrukcji z powtórzonymi komponentami i ich wariantów. Ponadto omówiono automatyczną identyfikację analizowanych frazeologizmów.

Marie Kopřivová

Charles University

Czech Republic

<https://orcid.org/0000-0001-7390-0753>

Variability of Czech Verbal Phrasemes: Case Study of *dát* ('to give')¹

Abstract. The paper is concerned with the topic of variability of Czech verbal idioms and its representation in a database of a multi-word expressions. In terms of material, it is based on SYN2015, a representative corpus of contemporary written Czech, which is equally divided into fiction, non-fiction, and newspapers and magazines. This corpus features an automatic annotation of multi-word units. The verb *dát* 'to give' serves as a case study, being one of the most frequent verbal components of Czech verbal idioms, right after the verbs *být* 'to be' and *mít* 'to have'.

Key words: *verbal idiom, corpus, multi-word expression, variability*

1. Introduction

The verb *dát* ('to give') is the third most frequently used verb in Czech phrasemes². It is a ditransitive verb with arguments in the dative and the accusative. It retains this valency in most phrasemes. Its imperfective aspect counterpart is the verb *dávat*, which ranks seventh, with the interesting property that 40% of its occurrences are in phrasemes.

The semantically opposite action is denoted by verbs *dostat* and *dostávat* ('to get'), which rank 12th and 240th, respectively. As phraseme components, they are less frequent than the verb to give (see Table 1). This frequency

¹ This study was written within the project Between Lexicon and Grammar, supported by the Grant Agency of the Czech Republic, reg. No 16-07473S.

² The term phraseme is used here especially with regard to the formal criteria for its definition (see Čermák, 2007, 83), which includes the term quasi-phraseme to (see Čermák, 2007, 104), which are verbo-nominal structures with an abstract noun.

disproportion is one of the signals that hint at the semantic bleaching of this verb, which is also manifested in collocations (cf. also the valency lexicon by Lopatková et al.).

Table 1 shows the most frequent verbs used in phrasemes in the SYN2015 corpus. It is possible to see here that the verbs *dát* and *dávat* (to give) have the highest proportion of occurrences within phrasemes compared to the other verbs.

For Czech corpora, automatic annotation of phrasemes can be performed using the FRANTA tool (Kopřivová – Hnátková, 2014), which is based on the Dictionary of Czech Phraseology and Idioms (Čermák et al., 2009). This annotation has so far been applied only to written corpora. Each collocation³ component has a collocation lemma and a collocation tag. This annotation should be improved, supplemented and clarified by the newly-developed database LEMUR (see Hnátková et al., Jelínek, 2019).

Table 1. The most frequent verbs occurring within phrasemes in the SYN2015 corpus

Rank	Verb		Occurrences			Different idioms
			Total	Phrasemes		
1	být	be	4,044,082	122,363	3%	3444
2	mít	have	734,066	94,341	13%	1417
3	dát	give	112,636	25,792	23%	529
4	říci	tell	180,576	15,555	9%	170
5	jít	go	160,655	14,059	9%	391
6	dělat	do	75,480	13,720	18%	339
7	dávat	give	29,075	11,983	41%	161
8	stát	stand	182,776	10,115	6%	239
9	vzít	take	44,928	9,060	20%	224
10	vědět	know	151,747	9,009	6%	176
11	nechat	let	54,787	8,311	15%	161
12	dostat	get	82,741	8,270	10%	307
240	dostávat	get	12,052	352	3%	27

Source: own research.

³ This annotation is also applied to other types of collocations, such as terms.

2. Data sources

For the analysis of the verb *dát* (to give) in corpus texts, we used two corpora of written Czech. The first corpus, SYN2015, is a representative corpus of contemporary written Czech, with balanced proportions of fiction, non-fiction, and newspapers & magazines (one third each). It contains 100 million word forms, lemmatization and POS and MWE tagging. Circa 4% of the word forms are marked as components of an MWE (mostly phrasemes). It was selected for analysis because it includes different types of texts in and it is balanced.

However, the range of attested phrasemes depends on the size of the corpus. This can be seen when automatic tagging of MWEs is applied to a large corpus, such as SYN_v6, which counts some 4 billion word forms. It is only in this larger corpus that some little-used phrasemes and proverbs appear, albeit with occurrence rates in the single digits. This unbalanced corpus of written language, with a predominance of journalistic texts, was the starting point for an analysis complementing the types and variants of verbal phrases with verbal component *dát* (to give).

3. The verb *dát* (to give)

The Dictionary of Czech Phraseology and Idioms catalogues 466 different phrasemes under the verb *dát* (to give). In the SYN2015 corpus, 529 different phrasemes are annotated, and in the SYN_v6 corpus, 712 different phrasemes. These numbers also include cases where the verb is associated with a reflexive pronoun (*se* or *si*). In the SYN2015 corpus, the numbers are separated: *dát* alone 349 different phrasemes, *dát* + *si* 106 different phrasemes, *dát* + *se* 74 different phrasemes.

For the purposes of the analysis, phrasemes were divided into nine formal groups. These groups were then analyzed in the SYN_v6 corpus. We only describe two groups in more detail.

3.1. Dividing the phrasemes into groups

1. verb + acc (abstract noun)

Example: *dát přednost* (to give priority) to prefer

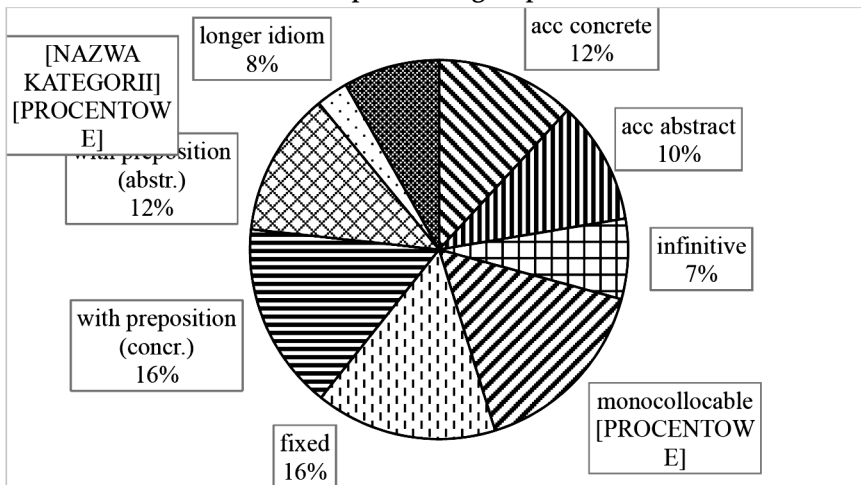
2. verb + acc (concrete noun)

Example: *dát pusu* (to give mouth – kiss) to kiss

3. verb + infinitive
Example: *dát vědět (to give to know)* *to alert*
4. verb + preposition + abstract noun
Example: *dát za pravdu (to give for the truth)* *to agree*
5. verb + preposition + concrete noun
Example: *dát k ledu (to give to the ice)* *put on ice*
6. verb + monocollocable word
Example: *dát bacha* *watch it/be careful*
7. verb in negation
Example: *nedat dopustit (do not give allow)* *to believe/to protect someone*
8. fixed idiom
Example: *to dá rozum (it gives sense)* *it is common sense*
9. longer idiom
Example: *dát ruku do ohně (put hand in fire)* *to be sure*

The distribution of these groups in the representative corpus SYN2015 is shown in Figure 1. The proportions are derived from the frequencies of all occurrences, not different types of phrasemes. The largest groups are fixed idioms, idioms with a monocollocable component and phrasemes with a preposition. For further detailed analysis, we chose the combination of the verb with the accusative.

Figure 1. Distribution of the *dát* phraseme groups in SYN2015



Source: own research.

3.2 Analysis of *dát* + acc

Using the SYN_v6 corpus, we searched for instances of the accusative following immediately after the verb and which were *not* annotated as phrasemes. This was the case of 61% of all occurrences. For a detailed analysis, the 115 most common nouns were selected, with their frequency ranging from 25,225 occurrences to 400 occurrences.

3.2.1. verb + acc (abstract noun)

A larger part of the analyzed sample consists of combinations of verb and abstract noun. These are classical verbo-nominal combinations (light verb constructions) which were missing in the dictionary, e.g.: *dát hlas* ('vote'), *možnost* ('possibility'), *odpověď* ('answer'), *příkaz* ('command'), *svolení* ('permission'), *podmínku* ('condition'). They are in competition with equivalent verbal expressions and their frequency increases at their expense.

The most frequent collocations are those that do not have a one-word equivalent, e.g.: *dát šanci* ('chance'), *prostor* ('space'), *průchod* ('passage'), *příležitost* (opportunity), *podnět* (impulse).

Their frequency is on the rise, in some cases perhaps under the influence of English⁴.

Some of these expressions are found in the dictionary with another verb, e.g.: *dát prostor – vytvořit* (space), *dát důvěru (vyjádřit důvěru)*, *dát maximum (vložit maximum)*. Together with the increasing frequency, there is a change of meaning, shifting to a more general plane, which confirms the process of grammaticalization of this verb. Individual expressions get simplified: more common vocabulary is used, which is less demanding and therefore corresponds to an informal style.

3.2.2. verb + acc (concrete noun)

With concrete nouns, the verb conserved its original meaning in many cases (e.g. to give a present, money, put your feet on the table).

In combination with the reflexive pronoun *si*, as in *dát si*, the meaning is "order/have some food or drink". The collocations *dát si panáka* (have an alcoholic drink) and *dát si jídlo* (have some food) are annotated. Other proto-

⁴ The English influence can be seen for example on the verb *dát* which newly occurs in spoken Czech with sole accusative valency, meaning "to handle something". These cases were previously covered in the Czech language by the verb *udělat* (to do), e.g. *dát zkoušku* (pass the exam).

typical representatives included the following 9: *pivo* (beer), *káva//kafe* (coffee), *čaj* (tea), *cigareta* (cigarette), *sklenička/sklenka* (little glass), *oběd* (lunch), *jídlo* (food). The shift that is captured by the corpus is not only from eating and drinking to other enjoyments such as smoking, but also towards metonymic equivalents of beverages (*sklenička* – glass). The glass here acts as a synonym for *panák* – which can express the amount of a drink or type of a drink – alcoholic.

Six other nouns come from sports news and are closely related to the second most common annotated collocation, *dát gól* (to score a goal). One is a synonym (*branku*), the remaining ones are either other football “terms” (*dát míč*, *balón* – pass the ball, *dát hatrick*, *penalty*) or pertain to another type of sport (*dát koš* – to score in basketball).

The remaining cases are verbo-nominal collocations which are used instead of their one-word verbal equivalent: *dát inzerát* – *inzerovat* (give an ad).

Conclusion

This partial analysis of the collocations of the verb *dát* (to give) has shown the importance of new corpus data. As it has been ten years since the *Dictionary of Czech Phraseology and Idioms* was released, a number of common phrasemes or their variants are missing. Using the corpus can reveal variants when verbs are alternated and determine which verb should be entered as default in the dictionary. It is necessary for users to be able to find all the variants if they know a less frequent one. The analysis showed the need to add new phrasemes to the repertoire.

The analysis also demonstrates the ongoing grammaticalization of the verb *dát*. Also, of course, it confirms that phrasemes with this verb are often used. With respect to the needs of the new LEMUR database and in the interest of better annotation of collocations in corpora of Czech, it is necessary to supplement this analysis of the verb data by analyzing the verb *dávat* (to give, imperfective) and verbs with opposite meaning *dostat*, *dostávat*. Together with the further analysis of phrasemes, this will form the basis for a better theoretical description of Czech phraseology.

Literature

- Čermák, František 2007. *Frazeologie a idiomatika česká a obecná*. Praha: Karolinum.
- Čermák, František et al. 2009. *Slovník české frazeologie a idiomatiky*. Vol. 1–4. Praha: Leda.

- Hnátková, Milena et al. 2017. Eye of a Needle in a Haystack. Multiword Expressions in Czech: Typology and Lexicon. In: Ruslan Mitkov (ed.). *Computational and Corpus-Based Phraseology. Second International Conference, Europhras 2017, London, UK, November 13–14, 2017*. Springer: Berlin–Heidelberg. 160–175.
- Kopřivová, Marie; Hnátková, Milena. 2014. From dictionary to corpus. In: *Phraseology in Dictionaries and Corpora*. Jesenšek, Vida; Grzybek, Peter (eds.). Maribor: Filozofska fakulteta Maribor. 155–168.
- Jelínek, Tomáš. 2021. Multi-word lexical units with repetition of lexemes in Czech and identification of their variants. (this volume).
- Kettnerová, Václava. 2017. Syntaktická struktura komplexních predikátů v češtině. In: *Slovo a slovesnost*, 78, 3–24.
- Sag, Ivan A.; Baldwin, Timothy; Bond, F.; Copestake, Ann; Flickinger, Dan. 2002. Multiword expressions: a pain in the neck for NLP. In: *Computational Linguistics and Intelligent Text Processing: Third International Conference, Cicing 2002*. Alexander Gelbukh (ed.). Berlin–Heidelberg: Springer. 1–15.

Corpora and tools

- Křen, Michal; Cvrček, Václav; Čapka, Tomáš; Čermáková, Anna; Hnátková, Milena; Chlumská, Lucie; Jelínek, Tomáš; Kovářiková, Dominika; Petkevič, Tomáš; Procházka, Pavel; Skoumalová, Hana; Škrabal, Michal; Truneček, Petr; Vondříčka, Pavel; Zasina, Adrian J. (2016). SYN2015: Representative Corpus of Contemporary Written Czech. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. Portorož: ELRA. 2522–2528.
- Křen, Michal; Cvrček, Václav; Čapka, Tomáš; Čermáková, Anna; Hnátková, Milena; Chlumská, Lucie; Jelínek, Tomáš; Kovářiková, Dominika; Petkevič, Tomáš; Procházka, Pavel; Skoumalová, Hana; Škrabal, Michal; Truneček, Petr; Vondříčka, Pavel; Zasina, Adrian J. *Corpus SYN, version 6 from 18.12.2017. Ústav Českého národního korpusu FF UK, Praha 2017*. <http://www.korpus.cz>.
- Lopatková, Markéta; Žabokrtský, Zdeněk; Kettnerová, Václava. VALLEX 2.5. *Valency Lexicon of Czech Verbs*. <http://ufal.mff.cuni.cz/vallex>.

Wariantywność czeskich frazemów werbalnych: Studium przypadku „*dát*” (‘dać’)

Streszczenie

Artykuł poświęcony jest wariantywności czeskich idiomów werbalnych i ich występowaniu w bazie połączeń wyrazowych. Materiał został zaczerpnięty z SYN2015, reprezentatywnego korpusu współczesnego pisanego języka czeskiego, w którym uwzględniono proporcjonalnie teksty fikcyjne, teksty niefikcyjne, prasę i czasopisma. Korpus umożliwia automatyczną anotację związków wyrazowych. Czasownik *dát* ‘dać’ posłużył jako przedmiot studium przypadku, ponieważ jest jednym z najczęstszych komponentów czasownikowych występujących w czeskich idiomach, zajmując trzecie miejsce po czasownikach *být* ‘być’ i *mít* ‘mieć’.

PART IV

**REPRODUCIBILITY FROM
A THEORETICAL PERSPECTIVE**

Wolfgang Eismann

Karl-Franzens-Universität Graz

Austria

<https://orcid.org/0000-0003-3681-7808>

Reproduzierbarkeit als unikales Kriterium zur Bestimmung von Phraseologismen. Zu den Verdiensten von Leonid Ivanovič Rojzenzon um die phraseologische Theorie

Abstract. Im vorliegenden Beitrag geht es um eine Würdigung der Verdienste Leonid Ivanovič Rojzenzons (1920–1977) um die Phraseologie. Rojzenzon war der Erste, der sich mit guten Gründen für eine linguistische Disziplin "Phraseologie" aussprach, deren einziges und ausschlaggebendes Kriterium die Reproduzierbarkeit ist. Diese fasste er als spezifische phraseologische Reproduzierbarkeit auf, die sowohl den Mehrwortcharakter als auch die Festigkeit des Phraseologismus mit einschließt. Der von Rojzenzon propagierte "dichotomische" Charakter der Reproduzierbarkeit ist allerdings nicht unumstritten. In einer Reihe von anderen Bereichen hat die von Rojzenzon mitbegründete phraseologische Schule von Samarkand jedoch Pionierleistungen für die phraseologische Forschung erbracht.

Key words: *Leonid Ivanovič Rojzenzon, Reproduzierbarkeit, phraseologische Schule von Samarkand*

1. Einführung

Da die Publikation, in die sich dieser Beitrag einreicht, unter dem übergeordneten Thema der Reproduzierbarkeit steht, sei es mir erlaubt, auf die Verdienste eines russischen Phraseologen hinzuweisen, der relativ früh dieses Kriterium als entscheidendes und einziges in den Mittelpunkt seiner phraseologischen Theorie gestellt hat.

Meine Ausführungen umfassen:

- kurze Erörterungen zu Leben und Werk von Leonid Ivanovič Rojzenzon (Kap. 2),

- einige Bemerkungen zu seinen Arbeiten und den diesbezüglichen Neuerungen im Bereich der Phraseologie (Kap. 3),
- eine Schilderung seiner Konzeption der Reproduzierbarkeit als einzigem Kriterium zur Definition von Phraseologismen (Kap. 4),
- sowie abschließend ein paar kritische Bemerkungen zu dieser Konzeption (Kap. 5).

2. Leben und Werk von Leonid Ivanovič Rojzenzon

Zu Leben und Werk von Leonid Ivanovič Rojzenzon gibt es mehrere kurze Darstellungen, die inhaltlich über den ihm nach seinem Ableben gewidmeten Nachruf (Nekrolog 1977) hinausgehen (Bulachov 1978, Zinin 2010). Rojzenzon wurde am 23.11.1920 in der Ukraine im Gebiet von Žitomir geboren und starb am 04.01.1977 in Samarkand. Nach dem Schulbesuch diente er in der sowjetischen Armee bis ins Jahr 1946 und erhielt mehrere Auszeichnungen für seine Teilnahme am Zweiten Weltkrieg. Danach studierte er am Institut für slawische Sprachen der Universität Lemberg, arbeitete nach Abschluss des Studiums (1950) zunächst als Dozent (prepodavatel') am Lehrstuhl für russische Sprache des Pedinstituts von Rivne (Рівне), russ. Rovno, poln. Równe, bevor er als Dozent an den Lehrstuhl für russische Sprache der Universität in Samarkand kam. Er doktorierte 1957 in Lemberg mit einer Dissertation zur historischen Syntax der tschechischen Sprache (kandidskaja dissertacija) und habilitierte sich an der Weißrussischen Akademie der Wissenschaften mit einer Dissertation über slawische Verbalpräfixe 1970 (doktorskaja dissertacija). 1972 wurde er zum Professor ernannt, war aber schon seit 1964 Leiter des Lehrstuhls für russische und allgemeine Sprachwissenschaft an der Universität von Samarkand. In Samarkand ist er auch gestorben. Rojzenzon hat eine Reihe von Arbeiten vorgelegt und sich dabei mit vielen Bereichen der Sprachwissenschaft, dem Russischen, den slawischen Sprachen, aber auch mit Kontakten und Einflüssen aus dem Gebiet der germanischen und der Turksprachen befasst. Neben Untersuchungen zu stilistischen Fragen, zur Morphologie und zur Syntax der russischen Sprache beschäftigten ihn vor allem die Syntax der slawischen Sprachen (vom Russischen bis zum Tschechischen und Sorbischen), lexikographische Fragestellungen, das slawische Verb, Fragen der historischen Sprachwissenschaft, Sprachstatistik, aber auch wissenschaftshistorische Probleme. Zu einem seiner Hauptarbeitsgebiete in Samarkand wurde mehr und mehr die Phraseologie. Sowohl im Bereich der russischen und slawischen als auch der allgemeinen Phraseologie hat er sich unschätzbare Verdienste erworben.

3. Rojzenzons Arbeiten im Bereich der Phraseologie

Rojzenzons phraseologische Forschungen begannen mit einigen Aufsätzen zur Phraseologisierung von Strukturen zusammengesetzter Sätze im Russischen und Tschechischen, die ihn zu allgemeinen Fragen des Problems der Phraseologisierung führten, das ihn immer wieder beschäftigte und das er später für ein psycholinguistisches Problem hielt, welches einer grundsätzlichen Aufarbeitung bedürfe (vgl. Rojzenzon 1973: 108). Darüber hinaus verfasste er eine Vielzahl von Arbeiten zu folgenden phraseologischen Themenbereichen: zur Entstehung und Entwicklung phraseologischer Einheiten, zum Verhältnis von Syntax und Phraseologie, zur Variation und zur Stilistik phraseologischer Einheiten, zum Vergleich von Phraseologismen in slawischen und anderen Sprachen, zur komparativen Phraseologie, zur Phraseographie, zu einzelnen Typen von Phraseologismen, zum Gebrauch von Phraseologismen in der Literatur und bei bestimmten Schriftstellern, zur Kontamination von Phraseologismen und zur Statistik von Phraseologismen. Er unternahm als Erster den Versuch einer detaillierten und vollständigen Klassifizierung der sog. Nekrotismen in der russischen Phraseologie (Rojzenzon 1973: 184–189) und definierte als einer der Ersten die Motiviertheit des Phraseologismus als seine "innere Form" (vgl. Zinin 2010: 83). Schon bevor in der westlichen Phraseologie von "Kinegrammen" die Rede war, hatte Rojzenzon gemeinsam mit Abramec diese Erscheinung untersucht und mit dem Terminus "sovmeščennaja omonimija" bezeichnet, was man in etwa mit "vereinte" oder "vereinigte Homonymie" wiedergeben kann.

Unter seiner Leitung und unter Mitarbeit von Julija Jul'evna Avaliani (einer Spezialistin für das Kurdische und die iranischen Sprachen) wurde Samarkand in den 70er Jahren des 20. Jhs zu einem der wichtigsten Zentren der Erforschung der russischen und der allgemeinen Phraseologie. Dies führte zur Herausgabe mehrerer Kongress- und Sammelbände. Der erste erschien nach einem Kongress 1959 im Jahre 1961 unter dem Titel "Voprosy frazeologii". Unter demselben Titel erschien ab 1965 eine Reihe von Bänden. Im Jahre 1972 kam dann erstmals das "Bjulleten' po frazeologii" in Samarkand heraus. Besondere Verdienste erwarb sich Rojzenzon auch im Bereich der bibliographischen Erfassung der Arbeiten zur Phraseologie und Parömiologie. Bereits 1964 hatte Vitalij Ivanovič Sergeev eine Bibliographie zur phraseologischen Forschung in der Sowjetunion von 1918 bis 1961 erstellt. Im zweiten Band der "Voprosy frazeologii" unternahm nun Rojzenzon 1965 gemeinsam mit M.A. Pekler den Versuch, eine allgemeine Bibliographie zu Fragen der Phraseologie zu starten. Diese Bibliographie war nach 25 Ka-

tegorien (teilweise mit Unterkategorien) gegliedert und enthielt 1507 Titel. Die Bände II und III wurden 1970 und 1974 von Rojzenzon und seinem Schüler A.M. Bušuj (Band III gemeinsam mit dem Bruder L.I. Rojzenzons) herausgegeben. In der Folge gab A.M. Bušuj die Bände IV, V und 6 in den Jahren 1976, 1979 und 1987 unter einem neuen Titel heraus. Die Bände II-6 enthielten jeweils ca. 2750, 2800, 3000, 3400 bzw. 1900 Titel und waren detailliert nach Kategorien und Sprachen geordnet. Bis 1987 ergaben sich somit insgesamt ca. 15.350 Titel. Man muss eventuell noch die Bibliographien zur Parömiologie Uzbekistans von Bušuj u.a. (1978, 1980) dazu rechnen, obwohl es hier manche Überschneidungen gibt. Immerhin handelte es sich um die bemerkenswerte Anzahl von über 15.000 Titeln. Wenn man das mit der hervorragenden Bibliographie von Wolfgang Mieder aus dem Jahr 2009 vergleicht, die ca. 10.000 Titel enthält, so finden sich hier bereits 12 Jahre früher insgesamt ca. 5.000 Titel mehr. Die Bibliographien wurden größtenteils im Rotaprintverfahren publiziert und erschienen oft in für die Sowjetunion sehr geringen Auflagen von 500 Exemplaren. Von der phraseologischen Forschung in den osteuropäischen Ländern wurden sie zwar wahrgenommen¹, doch kann man Wolfgang Mieder nicht vorwerfen, dass diese Autoren sowie deren Wirken und vor allem deren Bibliographien in seine exzellente Bibliographie keine Aufnahme fanden. Rojzenzon war auch der Erste, der zusammen mit anderen eine Geschichte der Phraseologie bzw. ihrer Entwicklung zu einer linguistischen Disziplin in der Sowjetunion von 1917 bis 1945 erarbeitete, wobei deutlich wurde, dass es lange vor Vinogradovs berühmten Arbeiten schon viele Bemühungen auf diesem Gebiet gab (Rojzenzon, Malinovskij, Chajutin 1975).

4. Das Kriterium der Reproduzierbarkeit

Rojzenzon sprach sich relativ bald für die Reproduzierbarkeit als einzigem Bestimmungskriterium der Phraseologismen aus. Dies hat er in mehreren Aufsätzen (vgl. z.B. Rojzenzon 1972) und dann vor allem in seinen beiden Monographien (1973: 101–115 und 1977: 8–10) immer wieder deutlich gemacht und dabei versucht, den vielen kritischen Einwänden, die auf seine Festlegung hin erfolgten, mit entsprechenden Argumenten zu entgegnen.

Seine letzte Definition dieses entscheidenden Merkmals zur Bestimmung von Phraseologismen lautete (Rojzenzon 1977: 9f):

¹ Man vgl. z.B. nur die kritische Rezension von V.M. Mokienko (1983).

In jeder Sprache – sei es eine alte (oder gar uralte) oder eine moderne, und unabhängig davon, ob diese von Hunderten von Millionen oder nur einzelnen Individuen gesprochen wird, entstehen unausweichlich bestimmte sprachliche (Wort-)verkettungen, die später in den Sprechakten dieses Kollektivs von Sprechenden als fertige Einheiten reproduziert werden. Folglich werden derartige Einheiten von den Sprechenden nicht jedes Mal neu geschaffen, nicht konstruiert, sondern als bereits vor dem Moment der Rede fertig geschaffene verwendet. Das grundlegende Merkmal dieser Einheiten oberhalb der Wortebene ist ihre Reproduzierbarkeit in der Rede. Dabei schließt der Begriff der Reproduzierbarkeit als komplexer Begriff Begriffe wie den Mehrwortcharakter (sverchslovnost') (d.h. die Wendung darf in jedem Fall aus nicht nur einem Wort, sondern muss aus mindestens zwei Wörtern bestehen – Autosemantika oder Synsemantika oder aber aus deren Verbindung), die allgemeine Bekanntheit (d.h. die Wendung in dieser Form und dieser Bedeutung muss einem bestimmten sprachlichen Kollektiv bekannt sein), die Festigkeit/Stabilität (d.h. reproduziert werden soll ein- und dieselbe sprachliche Einheit, wenn auch ihre Variation in bestimmten Grenzen möglich ist) mit ein.

Für die phraseologische Reproduzierbarkeit ist im Unterschied zu anderen möglichen Typen der Reproduzierbarkeit das Merkmal der Dichotomie charakteristisch. Das heißt, dass jedem festen Wortkomplex (USK = ustojčivij slovesnyj kompleks) im Bewusstsein des Sprechenden eine analoge Verbindung von Lexemen einer freien syntaktischen Konstruktion gegenübergestellt ist (direkt oder indirekt, real oder irreal). Der Begriff der Reproduzierbarkeit verfügt über keine rein linguistische Eigenschaft, das ist ein Begriff eher psycholinguistischen als rein sprachlichen Charakters und in diesem Sinne ist er nur den festen Wortkomplexen eigen.²

Diese bereits früher immer wieder ähnlich geäußerte Definition hat viel Kritik hervorgerufen, die hier nicht in Einzelheiten wiederholt werden soll. Einer der Haupteinwände, der in einer ausführlichen Kritik von Valerij Michajlovič Mokienko (1976), aber auch von vielen anderen vorgebracht wurde, betraf den Umfang der Phraseologie, die ausgehend von diesem einzigen Kriterium unermesslich viele Einheiten einschließen würde. Seit den 50er Jahren hatte es ja in der Sowjetunion eine Unterscheidung zwischen enger und weiter Phraseologie gegeben, die Rojzenzon mit seinem "Einheits" Kriterium bewusst aufgeben wollte (Rojzenzon 1973: 114f). Dabei wurden die Unterschiede zwischen den vielen Einheiten, die anhand dieses Kriteriums zusammengefasst wurden, von ihm nicht gelehnet, sondern, im Gegenteil, deutlich herausgearbeitet. Ihm ging es darum, ein auf alle Einheiten der Phraseologie zutreffendes Kriterium zu definieren. Daher verwendete er viel Energie darauf, auch Sprichwörter, sprichwörtliche Redens-

² Diese Übersetzung, wie auch alle anderen, stammt von mir (W.E.).

arten³ und geflügelte Worte, aber auch die im Russischen häufigen Appositive, die wegen ihrer Schreibung auch als Einwortkonstruktionen behandelt werden, und zusammengesetzte Termini dazu zu zählen. Diese Form der weiten Phraseologie entspricht in vielerlei Hinsicht den Ansätzen der Konstruktionsgrammatik, jedoch mit einer wichtigen Einschränkung: Während Anhänger der radikalen Konstruktionsgrammatik weitestgehend dem *idiom principle* von Sinclair (1991: 110) folgen, das darauf beruht, dass Sprache vornehmlich durch das Aneinanderfügen von idiomatischen Elementen aus dem Lexikon (Idiomen, Kollokationen oder längeren bedeutungstragenden Einheiten) produziert (in Rojzenzons Sinn "reproduziert") und verstanden wird und bei ihnen das *open choice*-Prinzip (bei dem einzelne Wörter zu Sätzen nach den abstrakten Regeln der Grammatik zusammengefügt werden) eine geringe bzw. gar keine Rolle spielt, ist für Rojzenzon das *open choice*-Prinzip ohne Zweifel ebenfalls wirksam, wenn auch der Anteil der Phraseologie (der reproduzierten festen Wendungen) sehr groß ist. Er grenzt sich allerdings ab von der situativ bedingten Reproduziertheit und auch von der Reproduziertheit "auf wahrscheinlichkeits-distributiver Basis" (Rojzenzon 1973: 113), wie wir sie aus der Konstruktionsgrammatik in Verbindung mit der Korpuslinguistik kennen, und ebenso von expliziten Zitaten durch das Kriterium der Dichotomie (siehe unten). Denn man hatte ihm vorgeworfen, dass es auch durch die außersprachliche Realität bedingte "Reproduziertheit" geben könne, wie z.B. in einer Situation (die sich wiederholt ergeben kann), in der jemand ausruft *kakoj vkusnyj sup!* (1973: 107)⁴, oder dass auch solche gewöhnlichen und oft wiederholten üblichen freien Wendungen wie z.B. das englische *blue sky* eigentlich nicht freie, sondern reproduzierte Einheiten seien. Ebenso würden auswendig gelernte Zitate oder auch längere Texte nicht frei produziert, sondern reproduziert. Von diesen Formen der Reproduzierbarkeit unterschied Rojzenzon jedoch die phraseologische Reproduzierbarkeit anhand des Kriteriums der Dichotomie. Darunter verstand er, dass jedem festen Wortkomplex im Bewusstsein des Sprechers eine analoge freie Wortverbindung gegenübergestellt sei bzw. entsprechen würde. Der Kritik, dass

³ Gelegentliche Fehler in der Zuordnung, wenn er z.B. die Sprichwörter *ovčinka vydelki ne stoit* und *mal zolotnik, da dorog* zu den sprichwörtlichen Redensarten zählt (Rojzenzon 1973: 124f), sollte man ihm nachsehen.

⁴ Andererseits hat gerade Rojzenzon viele Beispiele für "situativ gebundene Phraseologismen" (1977: 98f) (*dobroe utro, Ty ne steklannyj, U tebja čto otec – stekol'sčik?*) angeführt und darauf hingewiesen, dass diese großteils lexikographisch nicht erfasst seien. Auf den möglichen Unterschied zwischen einer "Kommunikationssituation", in der diese verwendet werden, und einer durch äußere Umstände häufig wiederholten Situation, in der man vielleicht *kakoj vkusnyj sup!* äußert, ist er nicht eingegangen.

das in vielen Fällen sehr schwierig nachzuvollziehen sei, entgegnete er mit folgenden Argumenten: Die phraseologische Dichotomie sei eine besondere Form der Dichotomie, über die andere bedeutungstragende Einheiten nicht verfügten. In den Fällen, in denen es keine "freien" Doppelgänger der festen Wortkomplexe gebe, läge eine besondere Form des Auftretens der Dichotomie vor, denn im Falle des Leugnens der Dichotomie dieser Einheiten müsste man davon ausgehen, dass sie "aus dem Nichts" entstanden seien. Daher müsse man in diesen Fällen von einer "indirekten Dichotomie" ausgehen (1977: 10). Die indirekte Dichotomie unterscheide sich von der gewöhnlichen Dichotomie dadurch, dass in letzterer die Inhalts- und Ausdrucksebenen der gegenübergestellten Einheiten auf die ein oder andere Weise übereinstimmen (ganz oder teilweise), während im Falle der phraseologischen Dichotomie nur die Form übereinstimmen würde, nicht aber der Inhalt. Diese Aussage Rojzenzons bleibt widersprüchlich, da nicht ganz klar wird, ob sich der Begriff "indirekte Dichotomie" auf alle Phraseologismen bezieht oder nur auf die, zu denen keine freien Wortverbindungen in der Sprache existieren.

5. Abschließende Kritik und Würdigung

Was Rojzenzons phraseologischen Ansatz betrifft, so liegt diesem ein spezifisches (wenn auch nachvollziehbares) Verständnis von "Dichotomie" zugrunde. Es bleibt unklar, ob sich die "Indirektheit" auf alle Phraseologismen bezieht oder nur auf diejenigen, zu denen es keine freien Doppelgänger gibt. Falls Letzteres der Fall ist, wofür vieles spricht, ist das nur vor dem Hintergrund eines Abbildrealismus zu verstehen, den man nicht unbedingt teilen muss. So sind einige sog. Unsinnphraseologismen wie das vielzitierte *ot žiletki rukava* oder deutsch *weniger als nichts*, frz. *trois jours après jamais* und viele andere wohl kaum dadurch zu erklären, dass ihnen einmal eine freie Wendung entsprochen habe, die die Grundlage für die phraseologische Wendung gebildet haben soll. Ähnliche Schwierigkeiten gibt es bei der Bildung von Termini wie z.B. *centr tjažesti* (Schwerpunkt), dessen sekundäre Phraseologisierung Rojzenzon natürlich nicht leugnen würde, dessen primäre Phraseologisierung aber höchstwahrscheinlich in einem bewussten Akt der Terminologiebildung begründet ist, wie bei vielen anderen Termini. Für Rojzenzon ist das Fehlen von "Doppelgängern" bei diesen Phraseologismen nicht etwa mit dem Fehlen der Dichotomie zu erklären, sondern nur mit der besonderen Form ihres Erscheinens. Er verwendet dafür den Begriff der indirekten Dichotomie, weil es für ihn sonst bedeuten würde, dass diese festen Wendungen aus dem "Nichts" kämen, und das würde heißen, dass "die Leugnung des di-

chotomischen Merkmals bei einigen Klassen von festen Wendungen uns unabweichlich von materialistischen Positionen in die Richtung des Idealismus führen würde.“ (1977: 10) Die Erklärung für Phraseologismen, deren Entstehung auf Sprachspiel, Kalauern oder bewusster Terminologiebildung beruht (freilich macht sie erst der Usus zu Phraseologismen), die selbstverständlich, wie alle Phraseologismen, immer “ein Stück der Wirklichkeit bezeichnen“ (Rojzenzon 1973: 102, 120), bleibt für Rojzenzon ein psycholinguistisches Problem. Das liefert aber keine Erklärung für die besondere Form der Dichotomie bei diesen Wendungen. Dass es auch hier zu Modellbildungen kommen kann, einzel- und zwischensprachlich, steht auf einem anderen Blatt, hat aber mit der indirekten Dichotomie nichts zu tun (vgl. *am St. Nimmerleinstag*, kroat. *na sveto nigdarjevo*, russ. *na tureckuju paschu, na russkij bajram*). Ohne hier einen radikalen Konstruktivismus oder einen absoluten Nominalismus vertreten zu wollen, muss man doch anerkennen, dass es offenbar Möglichkeiten gibt, mit Mitteln der Sprache “Begrifflichkeiten“ oder auch “unsinnige“ Wendungen zu konstruieren, für die es keine freien Wendungen gibt, die ihnen gegenüberstehen, die also aus dem “Nichts“ geschaffen werden oder auf menschlicher Kreativität beruhen. Dass dieses “Nichts“ aus z.T. vorgefundenen Strukturen (existierenden Wörtern, wobei es aber auch neugeschaffene Unsinnswörter gibt, die hierfür herhalten können) geschaffen wird, mag ein schwaches Argument zugunsten Rojzenzons indirekter Dichotomie sein.

Grundsätzlich lässt sich sagen, dass Rojzenzon bislang der einzige war, der den Versuch unternommen hat, eine linguistische Phraseologie unter einem einzigen entscheidenden Kriterium zusammenzufassen. Dabei hat er zwar auch den Usus einbezogen, aber sich nicht allein auf diesen gestützt, wie das die radikale *construction grammar* in Verbindung mit der Korpuslinguistik tut. Sein psycholinguistisches Kriterium der phraseologischen Reproduzierbarkeit, die durch Dichotomie gekennzeichnet ist, bleibt da vage, wo er diese Dichotomie als indirekt oder unreal bezeichnet. Zudem hat es Rojzenzon unterlassen, deutlicher auf die fließenden Grenzen zwischen den verschiedenen semantisch-lexikalischen Einheiten hinzuweisen, wenn er z.B. im Gegensatz zu vielen anderen Linguisten Bindestrichwörter, die in der russischen Sprachwissenschaft als Einwortkonstruktionen oder allenfalls als Komposita gelten, in die Phraseologie mit einbezogen hat, obwohl für ihn ein unerlässliches Kriterium für die Zugehörigkeit einer Einheit zur Phraseologie deren Mehrwortcharakter ist. Auf die fließenden Übergänge – einzel- und zwischensprachlich – zwischen den lexikalischen Einheiten vom Wort bis zum Sprichwort haben Peter Grzybek und ich selbst an anderer Stelle hingewiesen (Eismann, Grzybek 1994).

Zum Schluss bleibt mir nur, noch einmal festzustellen, dass Rojzenzon trotz dieser kritischen Bemerkungen in der Phraseologie vieles von dem vorweggenommen hat, was sich später in Westeuropa erst langsam zu entwickeln begann.

Literatur

- Bjulleten' po frazeologii* 1. 1972. *Trudy Samarkandskogo gosudarstvennogo universiteta imeni Ališera Navoi, Novaja serija*, vyp. 234. Samarkand: Izdatel'stvo Samarkandskogo Gosudarstvennogo universiteta.
- Bulachov, Michail Gapevič. 1978. Rojzenzon, Leonid Ivanovič. In: ders. *Vostočno-slavjanskije jazykovedy*, t. 3, Minsk. 168–176.
- Bušuj, Anatolij Michajlovič. 1976. *Bibliografičeskij ukazatel' literatury po voprosam frazeologii* vyp. IV, Samarkand: Samarkandskij Gosudarstvennyj Universitet imeni Ališera Navoi.
- Bušuj, Anatolij Michajlovič. 1979. *Bibliografičeskij ukazatel' literatury po voprosam frazeologii*, vyp. V, Samarkand: Samarkandskij Gosudarstvennyj Universitet imeni Ališera Navoi.
- Bušuj, Anatolij Michajlovič. 1987. *Bibliografičeskij ukazatel' po frazeologii*, vyp. 6. *Osnovnye voprosy teorii frazeologii*. Samarkand: Samarkandskij Gosudarstvennyj Universitet imeni Ališera Navoi.
- Bušuj, Anatolij Michajlovič, Salimova, R.Ch., Sattarov, R.T. 1978. *Bibliografičeskij ukazatel' literatury po paremiologii. Paremiologija Uzbekistana*, vyp. 1. Samarkand: Samarkandskij gosudarstvennyj pedagogičeskij institut imeni S. Ajni.
- Bušuj, Anatolij Michajlovič, Salimova, R.Ch., Inojatova, F.B., Mizarbaeva, E., Kari-mov, I. 1980. *Bibliografičeskij ukazatel' po paremiologii. Paremiologija Uzbekistana*, vyp. 2. Samarkand: Samarkandskij gosudarstvennyj pedagogičeskij institut imeni S. Ajni.
- Eismann, Wolfgang, Grzybek, Peter. 1994. Phraseologismus und Sprichwörtliche Redensart. Vom Mythos der Nicht-Trennbarkeit. In: Christoph Chlosta, Peter Grzybek, Elisabeth Piirainen (Hg.), *Akten des Westfälischen Arbeitskreises "Phraseologie/Parömiologie" 1991/1992 (= Studien zur Phraseologie und Parömiologie 2)*. Bochum. 89–132.
- Mieder, Wolfgang. 2009. *International Bibliography of Paremiology and Phraseology* Vol. 1–2, Berlin, New York: De Gruyter.
- Mokienko, Valerij Michajlovič. 1976. Rez. von: L.I. Rojzenzon. Lekcii po obščej i russkoj frazeologii: Učebnoe posobie. In: *Issledovanija po russkomu i slavjanskomu jazykoznaniju*, vyp. VII. Samarkand: 200–209.
- Mokienko, Valerij Michajlovič. 1983. Sovetskie bibliografičeskie ukazateli paremiologičeskoj i frazeologičeskoj literatury. In: *Sovetskaja etnografija*. 3: 156–161.
- Nekrolog 1977 = Leonid Ivanovič Rojzenzon (Nekrolog). In: *Leninskij put'*, 1977, 07.01. Samarkand.

- Rojzenzon, Leonid Ivanovič. 1972. Ešče raz o vosproizvodimosti kak priznake frazeologii. In: *Voprosy frazeologii*, vyp. V, č. 1. Samarkand: 17–21.
- Rojzenzon, Leonid Ivanovič. 1973. *Lekcii po obščej i russkoj frazeologii*. Učebnoe posobie. Samarkand. Samarkandskij Gosudarstvennyj Universitet imeni Ališera Navoi.
- Rojzenzon, Leonid Ivanovič. 1977. *Russkaja frazeologija*. Učebnoe posobie. Samarkand: Samarkandskij Gosudarstvennyj Universitet imeni Ališera Navoi.
- Rojzenzon, Leonid Ivanovič., Abramec, I.V. 1969. Sovmeščennaja omonimija v sfere frazeologii. In: *Voprosy jazykoznanija*. 2: 54–63.
- Rojzenzon, Leonid Ivanovič, Bušuj, Anatolij Michajlovič. 1970. *Materialy k obščej bibliografii po voprosam frazeologii*, vyp. II (pod red. prof. Kopylenko, M.M.). Samarkand: Samarkandskij Gosudarstvennyj Universitet imeni Ališera Navoi.
- Rojzenzon, Leonid Ivanovič, Bušuj, Anatolij Michajlovič, Rojzenzon, S.I. 1974. *Bibliografičeskij ukazatel' literatury po voprosam frazeologii*, vyp. III (pod red. Kopylenko, M.M.) Samarkand: Samarkandskij Gosudarstvennyj Universitet imeni Ališera Navoi.
- Rojzenzon, Leonid Ivanovič, Malinovskij, E.A., Chajutin, A.D. 1975. *Očerki istorii stanovlenija frazeologii kak lingvističeskoj discipliny*. Samarkand: Samarkandskij Gosudarstvennyj Universitet imeni Ališera Navoi.
- Rojzenzon, Leonid Ivanovič, Pekler, M.A. 1965. Materialy k obščej bibliografii po voprosam frazeologii. In: *Voprosy frazeologii*. 147–244.
- Sergeev, Vitalij Ivanovič. 1964. Bibliografičeskij ukazatel' literatury po frazeologii, izdannoj v SSSR s 1918 po 1961 g. In: *Problemy frazeologii. Materialy i issledovanija*. Moskva–Leningrad: Nauka. 255–317.
- Sinclair, John. 1991. *Corpus, Concordance and Collocation*. Oxford: Oxford University Press.
- Voprosy frazeologii*. 1961. *Trudy Samarkandskogo gosudarstvennogo universiteta imeni Ališera Navoi, Novaja serija*, vypusk 106. Samarkand: Izdatel'stvo Samarkandskogo Gosudarstvennogo universiteta.
- Voprosy frazeologii*. 1965. Taškent: Izdatel'stvo "Nauka" Uzbekskoj SSR.
- Zinin, Sergej Ivanovič. 2010. Leonid Ivanovič Rojzenzon (K 90-letiju so dnja roždenija). In: *Russkij jazyk v škole*. 2: 78–84.

Reproducibility as a Unique Criterion for Determining Phraseological Units. To the Merits of Leonid Ivanovič Rojzenzon for Phraseological Theory

Summary

Though Leonid Ivanovič Rojzenzon (1920–1977) is still mentioned in some reference works on phraseology even today, his merits in the field of general and Russian phraseology await their due appreciation. The general theme of this publication gives occasion to do so, for Rojzenzon was the first who argued with good reasons for a phraseological theory whose only and decisive criterion is reproducibility (vosproizvodimost').

The paper shows how Rojzenzon defines the term of reproducibility while including in his definition the multiple word character and the fixedness of the phraseologism. Furthermore it is discussed how Rojzenzon distinguishes reproducibility and quotability and what reasoning he uses to defend his conception of the dichotomic character of reproducibility (each reproduced unit corresponds to a potential identical produced unit) against critical objections. It shall not be concealed that his explanation for those cases in which he talks about indirect dichotomy is based on vulgar materialist arguments which may be explained by the context of the prevailing epistemological doctrine in the former Soviet Union. Reproducibility is seen by Rojzenzon as a psycholinguistic property and is closely related to the problem of phraseologization, to which Rojzenzon provided basic work. This paper tries to pay tribute to the achievements of Rojzenzon in the field of phraseology. In the '70s of the 20th century Rojzenzon turned Samarkand into a center of research in general and in Russian phraseology. This was reflected in anthologies and bibliographies on phraseology, but also in pioneering achievements in special areas of phraseological research. For example, Rojzenzon was the first who dealt with the problem of special phraseologisms which much later were called kinegrams or gestural phraseologisms by other scholars.

Marina Gutovskaya
Belarusian State University
Belarus

Phrasemes: Reasons for Reproducibility and Specificity of Sign Functions

Abstract. The paper examines mechanisms of phraseologization and reasons for reproducibility of nominative phrasemes on the material of the English and Russian metalanguage (denoting language and speech phenomena) phraseological units collected from dictionaries. It views two classes of reproducible nominative units: 1) idiomatic phrasemes – they arise as a result of phraseologization of free word combinations through developing figurative meanings and become reproducible due to the demand for expressive names conveying emotive, attitudinal meanings; 2) non-idiomatic phrasemes – in this case phraseologization of free word combinations is ensured by nominative accuracy and boils down to reducing variants of possible linguistic expression of a concept to one nominative unit, reproducibility is explained by the need for neutral names expressing factual meanings. The paper determines the proportions of the two classes of phrasemes in the research material and on the bases of the data obtained ranks the functions of phrasemes.

Key words: *phraseme, phraseologization, reproducibility, expressive function, nominative function*

Most of the phrasemes originate from free multilexemic expressions. The process of phraseologization of a multilexemic expression consists in acquiring the quality of reproducibility and can be caused by different reasons – the multilexemic expression may take on a figurative meaning or it may turn out to be an only or most accurate name of a concept. The paper on the material of the English and Russian nominative metalanguage (connected with language and speech) phraseology views the classes of phrasemes in accordance with the mechanisms of their formation and their functions: 1) idiomatic phrasemes the reproducibility of which is insured by the need for bright, vivid names of concepts (they are often secondary names functioning as expressive replacements of existing neutral monolexemic namings) – *a peg to hang something on* ‘a topic for discussion’ (БАРС II: 385), *звонить во все колокола* (lit. to toll all the church bells) ‘to talk about sth everywhere’

(ФСРЯ: 90): 2) non-idiomatic phrasemes which are reproduced because they function as one and only or most accurate neutral names of concepts often not having any monolexemic match – *to have (a) bad press* ‘to be criticized a lot by newspapers’ (LID: 272), *свобода слова* ‘the right to express one’s views’ (OnlineБТС). The paper determines the proportions of the two classes of phrasemes and on the basis of the data ranks the sign functions of phrasemes.

The research material has been taken from modern dictionaries which contain the reproducible nominative units of the mentioned types: «Longman Idioms Dictionary» (LID), «Фразеологический словарь русского языка» (ФСРЯ) and other paper and electronic phraseological and lexico-phraseological dictionaries of the English and Russian languages. The total corpus size is 1879 phrasemes, including 850 English phrasemes, 1029 Russian ones.

1. Phrasemes that arise as a result of constant use of multilexemic expressions in figurative meanings

The multilexemic expressions *breathe fire* ‘to talk in a way that shows you are very angry and determined to get what you want’ (LID: 121), *напустить туману* (lit. to blow the fog) ‘to make the story unclear, confusing’ (ФСРЯ: 296) became phraseologized because they acquired an integral figurative meaning. According to Ch. Bally, semantically integral phrasemes are termed phraseological unities (Bally 1936: 66–87). Phraseological unities are idiomatic – they contain lexemes (one or more) in non-usual (not registered in dictionaries) meanings. In the idiomatic phraseme *call off the dogs* ‘to tell sb to stop criticizing someone else’ (LID: 90) all the constituents actualize non-usual meanings. In the phraseme *крылатые выражения* (lit. winged expressions) ‘expressions that have become reproducible, popular’ (ФСРЯ: 126) the constituent *выражения* ‘expressions’ actualizes its usual meaning, but the constituent *крылатые* (lit. winged) has a special meaning ‘that have become reproducible, popular’ which is not given in any dictionary, and this makes the phraseme idiomatic in the constituent *крылатые*.

Among the semantically integral idiomatic phrasemes two groups of units are distinguished: transparent and non-transparent phrasemes.

The transparent idiomatic phraseme is characterized by bi-plane semantics – coexistence of actual figurative and etymological literal meanings. Components of such a phraseme agree semantically, and this ensures the possibil-

ity of understanding the literal meaning of the phrase forming the phraseme and of a subsequent transition to the figurative meaning. Literal semantization of the combination of words making the phraseme *тяжелая артиллерия* (lit. heavy artillery) draws the image of powerful firearms. The actual meaning of the phraseme 'the most authoritative and indisputable arguments you resort to when your own proofs are exhausted in a dispute' (ФСРЯ: 12) fixes the connection between the two notions – of large-caliber guns used in a combat when light guns are deemed unsuitable to achieve the desired result, and of authoritative arguments involved in a dispute when the evidence provided turns out to be not convincing enough – on the bases of the semantic feature 'efficiency, effectiveness'. Literal reading of the phrase *tie sb's tongue* 'to force sb to keep silence' (АРФС, 1099) creates an image of a situation in which a person ties the tongue of another person. The association of depriving tongue of the ability to move freely with coercion to silence is based on the metonymic convergence of the tongue and human speech ability.

The non-transparent idiomatic phraseme has lost its etymological meaning, its actual meaning is unmotivated. According to V. V. Vinogradov, the unmotivated semantically integral idiomatic phraseme is termed a phraseological fusion (Виноградов 1977: 121). The components of the phraseological fusion do not agree semantically – synchronically it is impossible to understand the literal meaning of the combination of words forming the phraseme and to recognize the ideas that brought the phraseme to life: *curry favour* (with sb) 'to try to make sb like you by saying nice things about them or doing sth for them' (LID: 115), *благим матом кричать* 'to cry in a very loud voice' (ФСРЯ: 143). Non-transparency (lack of motivation of the actual meaning) of such a phraseme may be due to various reasons of extra- and intralinguistic nature: the circumstances in which the phraseme arose may be forgotten, a constituent lexeme or some meaning of a constituent lexeme may go out of use, some grammatical form of a constituent lexeme may become obsolete etc. In the phraseme *curry favour* the components semantically disagree – the literal semantization does not make sense, the motive of nomination is obscure. Non-transparency of the phraseme is caused by the change of its original form and the loss of the associations that generated it: the phraseme is an early 16th-century alteration of the Middle English *curry favel*, Favel being the name of a chestnut horse in an early 14th-century French romance who epitomized cunning and duplicity (OnlineFD). The phraseme *благим матом кричать* contains the obsolete noun *мат* with the meaning 'voice', the adjective *благой* realizes its obsolete meaning 'strong' (РФИЭС: 419).

The majority of synchronically unmotivated idiomatic phrasemes – phraseological fusions – consist of lexemes which are widely used and well-known (although they realize in the phrasemes the meanings different from those given in dictionaries): *cock and bull story* ‘a story, excuse etc. that you think is very hard to believe’ (LID: 65), *перемывать косточки* (lit. wash bones) ‘gossip about someone’ (ФСРЯ: 119). A few of them contain unique phraseologically bound lexemes which exist in the language only as components of one phraseme (or a few phrasemes) and are not very clear in meaning: *Thursday morning tippy tappys* ‘a person who criticizes or passes judgement with benefit of hindsight’ (OnlineFD), *прописать ижицу* (*ижица* is the name of the last letter (ѣ) of the Church Slavonic and Old Russian alphabet) ‘to reprimand someone severely’ (ФСРЯ: 96).

2. Phrasemes which come into common use as one and only or most accurate names of their concepts

The phrases *refuse to take no for an answer* ‘not to accept a refusal, to be persistent in demanding sth’ (OnlineFD), *слоговое письмо* ‘a writing system in which each symbol represents a spoken syllable instead of an individual sound’ (ФСРЯ: 264) have acquired the quality of reproducibility because these accurate wordings have been repeatedly used to refer to the actual concepts. The reproducible nominative units of this type are non-idiomatic – their semantics is constructed out of the usual (registered in dictionaries) meanings of their components. According to N. M. Shansky, they are termed phraseological expressions (Шанский 1996: 69).

Complete non-idiomaticity is rare for phraseology. The semantics of many phraseological expressions includes the semantic features of their components and some further features – semes which differ from those making the semantics of the constituent lexemes – and this creates some idiomaticity. The meaning of the phraseme *name names* ‘to mention the names of people involved when reporting bad behavior or making an accusation’ (LID: 241; OnlineFD) includes the meanings of the components *to name* ‘to mention’ (OnlineFD) and *a name* ‘a word by which a person is commonly and distinctively known’ (OnlineFD) and a further semantic feature ‘when reporting bad behavior or making an accusation’. The presence of further semes in the semantics of this and similar phrasemes indicates that they are idiomatic, although to a minimal degree (Mel’čuk 1995, p. 183).

Among the non-idiomatic phrasemes, a small group of units with unique phraseologically bound lexemes are distinguished: *subjunctive mood*

'a mood of a verb used to describe hypothetical or non-real actions' (OnlineFD), *верительные грамоты* (lit. letters credential) 'documents certifying the appointment of a certain person as a diplomatic representative in some state' (ФСРЯ: 26). The bound lexemes have a very narrow co-occurrence range – they occur together with one or a few lexemes. Most of them are obsolete words (historicisms, archaisms: *обиняком говорить* (lit. to speak with hints; *обиняк* is an archaic name of a hint) 'to speak with hints' (ФСРЯ: 175)) or poorly adapted borrowings (Latinisms and Church-Slavonicisms which appeared as inexact calques in translations from Greek: *genitive case* 'the grammatical case expressing possession, measurement, or source' (OnlineFD) is from Latin *casus genitives*; the word *genitivus* 'of or belonging to birth' was misused by Latin grammarians to render Greek *genikē ptōsis* 'the general or generic case', *genikē* expressing 'race or kind', *genikos* also meaning 'belonging to the family' (OnlineED)). The non-idiomatic reproducible units which combine in their structure free and unique phraseologically bound lexemes are known as phraseological combination (Шанский 1996: 68).

The bound lexemes can hardly be viewed as autonomous units of a language – each of them exists in a language and occurs in speech only as a component of a particular phraseme (one or a few), its meaning realizes only within a phraseme. Phraseologically bound status of such words is signaled in a special way in Russian lexicographic works. They are registered in dictionaries as parts of phrasemes and are not provided with definitions of their own – phrasemes that include them are defined instead: *взгадать* ◇ *ни вздумать, ни взгадать, ни пером описать* 'about someone, something so beautiful that it is hard to convey' (МАС I: 165). Phrasemes that include unique components with extremely narrow syntagmatics have a rather unclear semantics, and therefore they are on the periphery of usage. Phraseological combinations that contain poorly adapted borrowings are used mainly in professional spheres or in literature to create a specific effect. Phraseological combinations with obsolete lexemes are gradually fading from language.

Syntagmatic "bindingness" is a gradual property inherent (in varying degree) in phrasemes, collocations and free word combinations. "All combinations of words," writes Yu. D. Апресян, "are more or less unfree and, therefore, all lexical meanings are more or less bound, that is, dependent on semantic, lexical, syntactic or other context" (Апресян 1989: 111). It seems, however, that the three types of syntagmatic bindingness – bindingness of free word combinations, collocations and phraseological combinations – can be differentiated.

Bindingness of free combinations of lexemes is determined by the semantic relations of lexemes – the lexemes that have common semes (semantical and grammatical clamps, or ties) can be united (*crisp toast, in order to travel*), those that do not are never put together (phrases like **crisp jam, *in order travels* are impossible in speech).

In collocations one of the components is selected contingent on another one. The choice of intensifiers in the collocations *perfectly healthy, seriously wealthy, highly respected, deeply moving, painfully sensitive* is determined by the adjectives as keywords. The meaning ‘beginning’ in the collocations *завязать дружбу* (lit. to tie friendship) ‘to make friends’, *развязать войну* (lit. to untie war) ‘to start a war’, *подняться на борьбу* (lit. to rise to fight) ‘to begin fight’, *прийти в восторг* (lit. to come to delight) ‘to become delighted’, *впасть в панику* (lit. to fall into panic) ‘to become panic-stricken’ is expressed by verbs contingent on the noun keywords.

Bindingness of phraseological combinations is imposed by the norm of usage established in the literary language in defiance of the general selection and combination rules of the language. The adjective *курсорный* ‘cursory’ has a very narrow syntagmatics – it occurs in one and only phraseological combination *курсорное чтение* (lit. cursory reading) ‘rapid reading, without detailed analysis’ which is used in teaching foreign languages (МАС II: 154). According to the general rules of the Russian language the adjective *курсорный* with the meaning ‘performed rapidly and without attention to detail’ could be used to describe a variety of continued intellectual actions (in word combinations like **курсорное ознакомление* (lit. *cursory examination*), **курсорная проверка* (lit. *cursory inspection*)), but the norm secures it to one action – reading.

3. Main search results

Quantitative analysis of the metalanguage corpus collected from modern dictionaries has shown that the percentage of idiomatic phrasemes – phraseological unities and phraseological fusions – in the Russian (62%) and English (75%) material is one and a half – three times as high as the percentage of non-idiomatic phrasemes – phraseological expressions and phraseological combinations.

The data obtained suggest that the expressive function prevails over the purely nominative function in phraseology. Idiomatic phrasemes serve in language as the main means of conveying emotive, attitudinal content. The expressive function is realized by both phraseological and lexical units;

however, idiomatic phrasemes considerably exceed idiomatic (in figurative meanings) lexemes by number. The earlier study of the English and Russian metalanguage nominative units – lexemes and phrasemes – has shown that among the expressive idiomatic units the percentage of phrasemes is two and a half times higher than the percentage of lexemes (Гутовская 2018: 63–64). Idiomatic phrasemes also exceed lexemes by the degree of expressiveness. According to Teliya, a more powerful expressive effect of phrasemes is due to the fact that they “reflect a whole image-situation and act as a microtext in a text” (Телия 1996: 83): cf. *звонить* (lit. to toll) ‘to talk about sth everywhere’ (OnlineБТС) and *звонить во все колокола* (lit. to toll all the church bells) ‘to talk about sth to everyone, everywhere’ (ФСРЯ: 90).

Non-idiomatic phrasemes form quite a large class of phraseology. They nominate factual content together with non-idiomatic lexemes. The functions of lexemes and phrasemes in conveying factual content are delineated: lexemes specialize in expressing general factual meanings, phrasemes convey more specific factual meanings, nominate a variety of nuances that cannot be named by lexemes: cf. *word* and *household word*, *four-letter word*, *function word*. The study of the English and Russian metalanguage nominative units has shown that among the nominative units with general factual meanings the percentage of lexemes is five times as high as the percentage of phrasemes, among the units with more specific factual meanings the percentage of phrasemes is one and a half times as high as the percentage of lexemes (Гутовская 2018: 63). The majority of phraseological expressions and phraseological combinations are one and only names of their denotata and do not have any monolexemic match, which indicates the high status of non-idiomatic phrasemes in the system of nominative means of language.

The results obtained allow us to conclude that according to lexicographic sources the most active feeder of phraseological stock are idiomatic phrasemes which arise as a result of a constant use of multilexemic expressions in figurative meanings and fulfill an expressive function. The second active feeder are non-idiomatic phrasemes which come into common use due to their ability to name details of their denotata and perform a pure nominative function.

Literature

- Bally, Charles. 1936. *Traité de stylistique française*. Paris: Librairie C. Klincksieck.
- Mel’čuk, Igor. 1995. Phrasemes in language and phraseology in linguistics. In: *Idioms: Structural and psychological perspectives*. Everaert, Martin et al. (eds). Hillsdale, New Jersey: Lawrence Erlbaum Associates. 167–232.

- Апресян, Юрий Д. 1989. О работах И. Е. Аничкова по идиоматике. *Вопросы языкознания*. 6: 104–119.
- Виноградов, Виктор В. 1977. *Избранные труды: лексикология и лексикография*. Москва: Наука.
- Гутовская, Марина С. 2018. Тематическая организация лексико-фразеологического поля метаязыковых обозначений в русском и английском языках. *Журнал Белорусского государственного университета. Филология*. 1: 49–65.
- Теля, Вероника Н. 1996. *Русская фразеология. Семантический, прагматический и лингвокультурологический аспекты*. Москва: Языки русской культуры.
- Шанский, Николай М. 1996. *Фразеология современного русского языка*. Санкт-Петербург: Специальная литература.

Dictionaries

- LID – *Longman Idioms Dictionary*. 1998. Harlow: Pearson Education.
- OnlineED – *Online Etymology Dictionary*. www.etymonline.com.
- OnlineFD – *The Free Dictionary*. www.thefreedictionary.com.
- OnlineБТС – *Большой толковый словарь русского языка*. Кузнецов, Сергей А. (ред.). gramota.ru/slovari/info/bts.
- АРФС – Кунин, Александр В. 1956. *Англо-русский фразеологический словарь*. 2-е изд. Москва: Государственное издательство иностранных и национальных словарей.
- БАРС – *Большой англо-русский словарь АБВУЯ Lingvo*. 2007. 2 т. Москва: Русский язык-Медиа.
- МАС – *Словарь русского языка*. 1999. Евгеньева, Анастасия П. (ред.). 4 т. 4-е изд. Москва: Русский язык, Полиграфресурсы.
- РФИЭС – *Русская фразеология. Историко-этимологический словарь*. 2005. Мокиенко, Валерий М. (ред.). 3-е изд. Москва: Астрель; АСТ: Люкс.
- ФСРЯ – *Фразеологический словарь русского языка*. 2008. Тихонов, Александр Н. (ред.). 4-е изд. Москва: Русский язык-Медиа; Дрофа.

Phraseme: Gründe für die Reproduzierbarkeit und Specificität der nominativen Funktion

Zusammenfassung

Im vorliegenden Artikel werden Mechanismen der Phraseologisierung und Gründe für Reproduzierbarkeit nominativer Phraseologismen am Material von den Wörterbüchern entnommenen phraseologischen Einheiten der Metasprache (Bezeichnungen für Sprach- und Redephänomene) im Englischen und Russischen untersucht.

Betrachtet werden zwei Klassen reproduzierbarer nominativer Einheiten: 1) idiomatische Phraseologismen, die als Ergebnis der Phraseologisierung freier Wortverbindungen durch die Entwicklung übertragener Bedeutungen entstehen und reproduzierbar werden, um als expressive Namen emotive, evaluative Bedeutungen zu vermitteln; 2) bei nicht idiomatischen Phraseologismen wird die Phraseologisierung freier Wortverbindungen durch nominative Genauigkeit gewährleistet, sie reduziert Varianten des möglichen sprachlichen Ausdrucks für einen Begriff auf eine nominative Einheit, die Reproduzierbarkeit lässt sich dadurch erklären, dass neutrale Namen faktische Bedeutungen zum Ausdruck bringen sollen. Im Artikel wird das Verhältnis dieser zwei Klassen der Phraseologismen im Forschungsmaterial festgestellt: die idiomatischen Phraseologismen bilden 75 Prozent der englischsprachigen und 62 Prozent der russischsprachigen metasprachlichen Phraseologismen, während die nicht idiomatischen Phraseologismen 25 Prozent bzw. 38 Prozent ausmachen. Aufgrund der erhobenen Daten wird im Artikel der Schluss gezogen, dass expressive Funktion in der Phraseologie wichtiger ist als reine nominative Funktion; im Artikel werden besondere detaillierende spezifizierende Besonderheiten reiner nominativer Phraseologismen betont.

PART V

**IDENTIFICATION AND ANALYSES
OF PHRASEMES
ACROSS LANGUAGE VARIETIES**

Ewa Koziół-Chrzanowska

PAN

Kraków

<https://orcid.org/0000-0001-6308-7156>

The Problems and (some) Solutions of Identifying Key Multi-word Expressions (MWEs). The Case Study of Polish Newspeak

Abstract. The paper aims to indicate and solve problems with practical usage of methods created for identifying key MWEs. The analysis is carried out on the basis of linguistic material representing Polish Newspeak (the language of propaganda and its mass media in totalitarian period). The paper considers three challenges: preparing an initial list of units which are supposed to be key ones, collecting searchable linguistic data and choosing the criteria of selecting appropriate texts. These problematic decisions which have to be made before analysis are inspired by works by Anna Wierzbicka and Raymond Williams (the notion of key MWEs is understood analogical to the key words in the interpretation of these authors).

Key words: *keyness, multi-word expressions, MWEs, Newspeak, keywords*

1. Introduction

1.1. The notion of key multi-word expressions (MEWs)

In linguistic studies, notions like *keyness* or *key words* (*keywords*) are understood in many different ways. According to Stubbs (2010), three loosely related, derived from different academic tradition uses of the term *keyword* can be indicated: words and culture (Williams 1976/1983), words and texts (Scott and Tribble 2006), phrases and schemas (Francis 1993) (Stubbs 2010: 23–32). The theoretical basis for the given paper is the first group, i.e. the interpretation derived from the cultural studies carried out by Williams or Wierzbicka¹;

¹ According to these authors, *keywords* are "(...) focal point around which entire cultural domains are organized" (Wierzbicka 1997: 156) "(...) significant, binding words in certain activities and their interpretation; they are significant, indicative words in certain forms of thought" (Williams 1975: 15).

the term *key multi-word expressions* is understood by analogy to the *keywords* in this sense. Generally speaking, *key multi-word expressions* are the ones which are focal, significant for the given culture or type of discourse.

1.2. Characteristics of the linguistic material

Presenting the analysis of identifying key MEWs of Polish Newspeak requires at least brief characteristic of the phenomenon. The term itself was coined by George Orwell in his *Nineteen Eighty-Four* to name an artificial, official language. The contemporary meaning is 'ambiguous euphemistic language used chiefly in political propaganda'. From 1944 to 1989 the Polish Republic was a non-sovereign country, dependent on the USRR as far as political and economic sense is concerned. It is claimed that during this period the official discourse was dominated by the Newspeak.

The Polish Newspeak has a few features, represented by groups of linguistic means. One of these features are pragmatic and semantic manipulations. They consist in giving new values to the language units, changing the components of their meaning, using the words with too general or too detailed meaning, e.g. *suggestions* was always used in a positive (e.g. *soviet suggestions*), while *declarations* – always in negative sense (e.g. *American declarations*), while there is no such division in the standard Polish. The next feature is using the schemata: conventional and repetitive phrases, metaphors and metonymies, e.g. *dalsze zacieśnianie braterskiej współpracy* (lit. continued bringing closer brotherly cooperation), *nirozzerwalny sojusz* (lit. inseparable alliance). In the Newspeak the world is divided into two parts: "we" and "you". There is always an enemy, who is presented in an unambiguously negative way and has characteristic distinguishing marks like weakness and dispersion. Another feature is distortion. The aim of the Newspeak is to create the so-called information commotion. The information is usually incomplete, fragmentary or simply false. However, by using repetitive schemata, authorities create the impression of doing a lot of pivotal activities. This is the way of creating texts which are devoid of information, but are full of phrases showing the power of authority and its operations. Taboo is also characteristic for the Newspeak. Propaganda avoided some words and phrases by omitting or replacing them, e.g. instead of *strike* the *brakes in work* took place, political opponents were closed not in *prisons* but in *places of seclusions*. Another feature of Newspeak is simple syntax, stylistic monotony, lexical poverty (Markowski 2007: 90–94).

According to the Polish researchers, the Newspeak has a few functions: persuasive, distorting (disturbing), ritual, controlling functions as well as

functions consisting in making the so-called information commotion, manifesting the authority's presence and organizing social emotions (Markowski 2007: 87–90). Głowiński states that in Newspeak values dominate over the meaning. The language is subordinated to the rules of rituals, the magical thinking about it plays a pivotal role: the aim of the language is not to describe or to get to know the reality, but to create it by using words in a desired way (Głowiński 2001: 175).

1.3. Aims

The article aims to indicate challenges which come to light while using methods created for identifying key MWEs in practice. As a next step, the paper examines to what extent these methods can be useful with reference to the linguistic material of Polish Newspeak. The final interest of the article lies in providing some solutions to the indicated problems.

2. Analysis

2.1. First problem: an initial list

According to Scott (2009), methods of identifying keywords in texts can be divided into three main groups: relying on word frequency alone, basing on human identification and combinations of these both (Scott 2009: 2). It may be assumed that the same conclusions can be referred MWEs. All these groups are important scientific procedures and should be carefully judged in reference to the analyzed linguistic material. However, in the given paper, the methods based on human identification of key MWEs fall within the scope of the survey.

The well-known authors using these methods are probably Raymond Williams and Anna Wierzbicka. Williams identified the keywords intuitively, and then searched for empirical evidence of their historical shifts in meaning:

First, Williams identifies words intuitively, on the basis of his extensive scholarship. He then uses the attested citations in the 12-volume *Oxford English Dictionary* as empirical evidence that his keywords have undergone historical shifts in meaning which have led to complex layers of meaning in contemporary English. (Stubbs 2010: 23–24)

Wierzbicka claimed that keywords are very often centers of phraseological clusters and that they frequently occur in some special kinds of texts, like proverbs, sayings, songs and titles. The belief that these kinds of text

have special significance for the culture and that they reflect this culture is the basis for such an assumption:

(...) one may want to show that this word is at the center of a whole phraseological cluster (...). One may also be able to show that the proposed “key word” occurs frequently in proverbs, in sayings, in popular songs, in book titles, and so on. (Wierzbicka 1997: 16)

It should be stated that the abovementioned methods based on human identification represent two different models. The first one – based on Williams’ method – can be called an extraction model, because the action goes from data to the list of MWEs. In the second one (based on method by Wierzbicka), the direction of the action is opposite – that is why the method can be named as confirmation model. When Williams chose the keywords intuitively, he had some data – his own intuition, linguistic memory and competence which let him choose the words considered as key ones. This situation is analogical to automatic extraction of keywords from a corpus. The researcher has some data and as a result of the action he extracts from them a list of key MWEs. In the confirmation model the starting point is the list of MWEs. As a result of using a given method, the researcher gets the confirmation or rejection of the MWEs key status. Talking about methods of identifying key words or expressions is in fact a simplification, because some of these methods (representing the confirmation model) do not identify the words and expressions but confirm their key status. The distinction of two models is important, because using them incur slightly different practical problems.

The first challenge arising from the confirmation model is simply having the initial list of key MWEs. In other words, when the researcher wants to check if there are variants of expressions or if they occur in some kinds of texts, he needs to have these expressions first. Both, Williams and Wierzbicka, used their own intuition. Can this method be considered as a reliable one? To some extent, the answer may be positive. As Wierzbicka stated – if the researcher’s choice is wrong, he will not get any interesting results. Some confirmation procedures can be used for checking intuitional choices. The bigger threat here is missing some important units. If the researcher omits them, they will probably remain omitted. Another problematic situation is having no or almost no intuition. In the case of Polish Newspeak, carrying out the research is problematic for those who do not remember the totalitarian period well or even at all. In such cases the only linguistic intuition about key words or MWEs of the past can be based on an idea built by books, films, newspapers and so on. The obvious advantage of older researchers is

their better intuition, which they could build by being immersed in the real, everyday various and live discourse.

In the case of Polish Newspeak, these problems are partially solved, thanks to works by Głowiński. Among his many books on the Polish Newspeak, there are four of them² in which the author describes particular words and phrases which he found crucial, interesting, surprising and so on. These books are a kind of linguistic diary – they were created on a regular basis in the totalitarian period of Polish history. They cover almost all years from 1966 to 1989. The tables of content of these books can be used as an initial list of key words and phrases of Polish Newspeak. Unfortunately, in the Polish scientific literature, we do not have any similar papers on the previous period (before 1966). The only possibility is to find some more general papers by different authors and note down the expressions which they describe. This piece of advice can be treated as a general solution to the problem of completing the initial list needed as a base for the research in the confirmation model. If the researcher looks for such a list, one of the possibilities is to search for the examples in as many various scientific works on the subject as possible. We can assume that their authors used plenty of sources or their own intuition, which is different from our own. The access to these works may be very helpful in the process of completing a list which can be processed in research being a part of a confirmation model.

2.2. Second problem: systematic search

The next indicated challenge of identifying key MWEs concerns both shown models (extraction and confirmation ones). No matter if the researcher wants to confirm his own assumptions that a given expression was a key one in a given period or if he wants to extract such phrases, he needs the collection of data which is searchable.

In the case of the extraction model, this data is available – first of all, the National Corpus of the Polish Language. One of the filters lets the users search only press texts, which seem to be the best source of propaganda (in comparison with books or the spoken language). The periods of publications can be also limited by choosing the years in which the newspapers were printed. The shortcoming of available search engines created for National Corpus is the lack of possibility of automatic extraction of a list of collocations. The same problem regards another source of texts – Chronopress,

² Głowiński 1991, 1993, 1996, 1999.

the portal of Polish press texts from year 1945 to 1954. However, this source is a part of the CLARIN – European Research Infrastructure for the Social Sciences and Humanities, focusing on language resources (data and tools). It means that we can easily use Chronopress with tools available in CLARIN, such as MeWeX, which is created for extracting the collocations from corpora.

The situation is much more complicated in relation to the confirmation model. In the abovementioned quotation Anna Wierzbicka (Wierzbicka 1997: 16) mentions proverbs, sayings, popular songs and book titles as the texts which are important for confirming the status of keywords. This list can be easily extended, e.g. to posters or internet memes. However, the pivotal problem lies in the accessibility of the data and the possibility of searching them. There is no collection of such texts which would let the researcher easily look for a word or expression in popular songs, for example. It is impossible to search this kind of texts in the same way in which the corpora can be analyzed. There are two main possibilities of solving this problem first – search many scattered sources, second – assume that many of these texts are available on the Internet and use its standard search engines to do the research.

2.3. Third problem: selection of texts

The last of the indicated problems is the challenge of selecting appropriate texts as the basis of the research. The quoted method by Anna Wierzbicka (Wierzbicka 1997: 16) assumes that some texts – which can be called “significant texts” – have special importance.

This importance is based on two mechanisms. Either the MWEs is key, so it appears in significant texts or the text is so significant that it makes MWE a key one. These two types of relationship are represented by illustrations 1 and 2.

The first illustration is the example of a significant text. The propaganda poster from 1945 became a symbol of a communist terror and post-war persecution of soldiers from Home Army. These soldiers are compared to dwarfs, as the communist propaganda was accusing them of collaboration with Germans and objection to social reforms. This metaphorical comparison is completely absent in other propaganda texts – there is no sample of this expression in Chronopress corpus (the abovementioned collection of press texts printed from 1945 to 1954). However, the poster is so significant and well-known that the MWE created for it became key. The opposite situation is visible in illustration no. 2. The internet meme uses the slogan *Polak potrafi* (lit. A Pole can). This expression was used on the building site

Illustration 1. The propaganda poster *Olbrzym i zapluty karzeł reakcji* (lit. The giant and the spat dwarf of the forces of reaction)



Source: the Internet.

Illustration 2. The Internet meme *Polak potrafi* (lit. A Pole can), example 1



Source: the Internet.

of Ironworks Katowice – a huge venture which became a symbol of Edward Gierek's (the first secretary of the communist party) era. Its meaning can be described as 'a Pole is a smart person who can deal with many problems'. The slogan became very popular and it is still used nowadays, usually in ironic contexts (like in the abovementioned meme, where it is a caption of the absurd construction being a mix of a car and a tractor). It is a key MWE of the Polish language and that is why it is frequently used in the internet memes.

At the same time the given illustrations represent two other categories of texts: the ones that are special by their exposition (a poster, no. 1) and the others which represent counterspeech³ (an Internet meme, no. 2). Both categories are useful as data for extracting key MWEs or confirming their key status.

Some types of texts are constructed in a way that exposes some content. To this idea refers, among others, the notion of text clusters (Püschel 1997) which is used for example in the keywords analysis on the Internet. Commercials and press where titles, leads and covers of newspapers play a special role work according to similar rules. In order to check the importance of exposition factor in propaganda press, the analysis based on the Chronopress corpus was conducted. The research compared frequency of using the words in the newspapers in general and on their covers in 1945. Taking into consideration 1,000 most popular examples has shown that approximately 16 per cent of words most popular in general were not comparably popular on the covers of newspapers. In about 5 per cent the difference in popularity was bigger than 1,000 positions on the frequency list. For example *a church* was 586th most frequently used word in Polish press in 1945, but at the same time it was only 2489th on the covers. Such examples suggest that the frequency of word is not crucial. Can it be stated that the word or phrase is key if it is not exposed? Probably the answer should be negative. In other words, the fact that some words or expressions appear in the texts or parts of texts which are well exposed proves their key status (in the case of press the best exposed part of the text is definitely the cover).

The next criterion helpful in choosing the types of significant texts is their affiliation to counterspeech. It seems to be obvious that all examples of counterspeech are based on units which are well-known, established in a language, used at least by a small group of people. Otherwise, making

³ Counterspeech is a linguistic phenomenon of opposing the traditional forms of communication used in a given society at a specific time, e.g. antiproverb.

Illustration 3. The Internet meme *Polak potrafi*, example 2 (lit. A Pole can)



Source: the Internet.

Illustration 4. The Internet meme *Polak potrafi* (lit. A Pole can), example 3



Source: the Internet.

counter-units based on them would not have any sense at all. Moreover, a large number of variants is a sign of an important role the unit plays in a language and culture. It can be stated that a MWE being a base for many various counter-units is key itself. Illustrations from 3 to 5 are the internet memes based on the slogan *a Pole can*. They constitute only a small sample of the collection which can be easily found on the Internet. Their number and variability are a sign of the fact that this slogan is important for Polish Newspeak, at least from the contemporary perspective.

Illustration 5. The Internet meme *Polak potrafi* (lit. A Pole can), example 4



Source: the Internet.

3. Conclusion

The aim of the paper was to indicate challenges of identifying key MWEs of Polish Newspeak and provide at least some solution to these problems. The first problem – creating an initial list of key MWEs of Polish Newspeak – has already been partially solved by Michał Głowiński's works. The period which was not described by the author needs a list created separately. This goal can be achieved by searching examples from different scientific works devoted to the topic of totalitarian propaganda. The second problem – systematic search in significant texts like sayings, songs and so on – can be solved either by using the Internet, or by a detailed enquiry of various sources. The third of the indicated problems is the challenge of selecting appropriate texts as the basis of the research. The solution can be provided on the basis of two main criteria: exposition and counterspeech. According to them, the texts which are well-exposed and/or represent counterspeech are significant enough to establish a collection of texts useful for searching of key MWEs.

Literature

- Głowiński, Michał. 1991. *Marcowe gadanie 1966–1971*. Warszawa: PoMost.
 Głowiński, Michał. 1993. *Peereliada 1976–1981*. Warszawa: PIW.

- Głowiński, Michał. 1996. *Mowa w stanie oblężenia 1982–1985*. Warszawa: OPEN.
- Głowiński, Michał. 1999. *Końcówka 1985–1989*. Kraków: Wydawnictwo Literackie.
- Głowiński, Michał. 2001. Nowomowa. In: Bartmiński, Jerzy (eds.). *Współczesny język polski*. Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej. 173–183.
- Markowski, Andrzej. 2007. *Kultura języka polskiego*. Warszawa: PWN.
- Püschel, Ulrich. 1997. Puzzle-Texte – Bemerkungen zum Textbegriff. In: Antos, Gerd; Tietz, Heike (eds.). *Die Zukunft der Textlinguistik. Traditionen, Transformationen, Trends*, Tübingen. 27–41.
- Scott, Mike. 2009. Problems in investigating keyness, or clearing the undergrowth and marking our trails. In: Bondi, Marina; Scott, Mike (eds.): *Keyness in Text*. Amsterdam: John Benjamins. 43–59.
- Stubbs, Michael. 2010. Problems in investigating keyness, or clearing the undergrowth and marking our trails. In: Bondi, Marina; Scott, Mike (eds.). *Keyness in Text*. Amsterdam: John Benjamins. 21–43.
- Wierzbicka, Anna. 1997. *Understanding Cultures through Their Key Words: English, Russian, Polish, German, Japanese*. New York Oxford: Oxford University Press
- Williams, Raymond. 1975. *A Vocabulary of Culture and Society*. Oxford: Oxford University Press.

Internet sources

- ChronoPress: Chronologiczny Korpus Polskich Tekstów Prasowych [Chronological Corpus of the Polish Press Texts]. <http://chronopress-test.clarin-pl.eu/>.
- MeWeX. <https://mewex.clarin-pl.eu/>.
- NKJP: *Narodowy Korpus Języka Polskiego* [National Corpus of the Polish Language]. <http://nkjp.pl/>.
- Examples 1–4 – found by the Google Images search engine (as a result of searching every single MWE).

Los problemas y (algunas) soluciones para identificar expresiones multipalabras clave. El case de estudio: la Neolengua Polaca

Resumen

Indudablemente, algunas expresiones creadas por la Neolengua Polaca (el lenguaje creado por la propaganda y los medios de comunicación masiva en el periodo totalitario) están aún en uso en el idioma polaco (esto es después de la caída del comunismo y la República Popular Polaca). Su presencia en el lenguaje actual, así como sus cambios semánticos y pragmáticos, contribuyen a importantes problemas en el idioma y la cultura polaca. Sin embargo, un análisis detallado de estos problemas requiere responder a una pregunta básica: ¿Cuales de estas expresiones, a las

que llamamos MWEs, pueden ser reconocidas como palabras clave? Diferentes criterios para identificar estas palabras clave están señalados en diferentes textos y son analizados basándose en publicaciones de la época totalitaria en Polonia. Como resultado, este trabajo muestra los problemas fundamentales y ofrece algunas soluciones a ellos. Las conclusiones pueden ser útiles para el caso de estudio que aquí se considera (identificación de palabras claves MWEs en el Neologismo Polaco), así como para otros textos del mismo perfil.

Katie Ní Loingsigh

Roinn na Gaeilge

School of Languages, Literatures and Cultures, NUI Galway, Ireland

<https://orcid.org/0000-0002-2145-0591>

Irish-language Idioms: an Overview and Analysis of Idioms in the Published Work of Peadar Ó Laoghaire

Abstract. This paper provides an account of the principal features and characteristics of Irish-language idioms – specifically idioms collected from the published work of Canon Peadar Ó Laoghaire (Peter Ó’Leary) (1838–1920), the primary Irish-language prose author of the late nineteenth and early twentieth century. The chief features of the collected idioms are analysed and described which provides an insight into Irish-language idiom syntax and semantics. Due to the limited amount of research undertaken on Irish-language idioms, this paper provides an overview of a relatively new field of academic research in the Irish language. Additionally, it provides a foundation for further research, comparison between idioms in other dialects and a base for future studies in idiom-related research.

Key words: *idioms, Irish language, corpus linguistics*

1. Introduction

The study of Irish-language idioms in the field of phraseology is a relevantly new and underdeveloped area of research. Even though idioms have been collected and analysed as part of general lexicographic studies from the late nineteenth century onwards, there has been only one major academic study undertaken on Irish-language idioms to date, i.e. *A concordance of idiomatic expressions in the writings of Séamus Ó Grianna* (Ó Corráin 1979)¹.

¹ Ó Corráin’s research focuses on idiomatic expressions collected from the published work of the Irish-language prose author, Séamus Ó Grianna, and contains a wide collection of idioms, proverbs, verbal nouns and idiomatic phrases. This analysis was completed before the widespread availability of large scale corpora and specifically focuses on idiomatic expressions in the Ulster dialect of Irish.

This paper aims to analyse the principal features and characteristics of Irish-language idioms in the work of Canon Peadar Ó Laoghaire (Peter O’Leary) (1839–1920), the primary Irish language prose author of the late nineteenth and early twentieth century. Ó Laoghaire’s work left an indelible mark on the Irish language throughout the twentieth century and numerous idioms found in his work are still in current use.

This paper focuses on the primary results of doctorate research which involved the creation of a database of idioms from Ó Laoghaire’s published work (<https://www.gaois.ie/en/idioms/>). An idiom is defined here as a type of phraseme which has a figurative meaning in terms of its whole, or a unitary meaning that cannot be derived from the meanings of its individual components and whose components can only be varied within restricted definable limits. This description follows the definition of idioms as laid down in the literature (e.g. Abdou 2012; Hanks 2004; Howarth 1998 and Fernando 1996). This paper is organised as follows; section 2 provides a brief background to the Irish language and Irish-language idiom research along with an overview of Peadar Ó Laoghaire and his work. Section 3 focuses specifically on the methodology regarding the collection and analysis of idioms. The primary features of the collected idioms are presented in Section 4 and future work and conclusions are discussed in Section 5.

2. Background

The Irish-language is one of two official languages of Ireland, the other being English. It belongs to the Celtic branch of the Indo-European family of languages and is a verb-subject-object (VSO) language. There are three primary regional dialects of Irish in Ireland – Connaught, Munster, and Ulster dialects – which correspond respectively to the most westerly, southerly, and northerly areas of the country (Kilgarriff, Rundell and Uí Dhonnchadha 2006). These areas are referred to as *Gaeltacht* regions. This paper focuses on the Muskerry dialect of Irish which corresponds with the Munster dialect and southerly region of Ireland. Despite the interest and research being undertaken in the field of phraseology internationally, there has only been a limited amount of research carried out on Irish-language idioms to date. Ó Corráin’s (1989) collection of idiomatic expressions as found in the Ulster dialect is the only academic research undertaken in this area in recent years. During the intervening period, the development of tools to analyse and collect idioms has been greatly improved, especially the availability and use of corpora in the analysis of idioms.

Up until the late nineteenth century, Irish-language idioms were collected and recorded as part of general lexicographic studies. A renewed interest in the Irish language and in Irish Gaelic culture, including a specific interest in folklore, sports and the arts, sparked a nationwide revival in Irish-language scholarship, research and cultural associations from the late nineteenth century onwards. This period is referred to as *Athbheochan na Gaeilge* (the Gaelic Revival) and lasted until the end of the 1920s and was followed by a protracted decline. It was recognized during this period that there was a need to collect and analyse the spoken language to ensure the language could be used as an appropriate comprehensive and modern medium for written communication (Ó Háinle 1994). The *Gaeltachtaí*, or Irish-speaking regions, became the focus of this work due to the strong aural tradition that existed in these areas. It was during this period that the bilingual periodical *Irisleabhar na Gaedhilge* (1882) was established and it became a central medium for the transmission of new Irish-language literature. The publication of *Irisleabhar na Gaedhilge* coincided with the founding aims of *Conradh na Gaeilge* (The Gaelic League) which was established in 1893, i.e. (i) to protect the Irish language as a national language and to increase its use as a spoken language and (ii) to promote and develop a native literature (Mac Mathúna 1987).

2.1. The Gaelic Revival

It was during the Gaelic Revival that a concerted effort was made to collect native phrases and expressions found in the spoken language and publish these collections for the benefit of an increased number of language learners. These compilations focused specifically on the collection of items currently referred to as multi-word expressions, phrasemes, multi-word units, fixed expressions, etc. Hogan (1898) explicitly focused on younger learners of Irish and produced a collection of *c.*2,500 idiomatic phrases in his work, *A Handbook of Irish Idioms* (Hogan 1898). Each entry was specifically selected, “because they were unlike English or other foreign modes of expression” (ibid. xiv). Following on from the publication of Hogan’s work, several other collections were published during the early twentieth century, e.g. *Leabhar Cainte* (Ua Dubhghaill and Bairéad 1901); *English-Irish Phrase Dictionary* (McKenna 1911); *Cora Cainnte as Tír Chonaill* (Mac Maoláin 1933); *An Béal Beo* (Ó Máille 1937) and *Cortha Cainnte na Gaedhilge* (Mac Síthigh 1940). Similarly, Mac Clúin’s work, i.e. *Réilthíní Óir I* (Mac Clúin 1922a) and *Réilthíní Óir II* (Mac Clúin 1922b), focused on the Munster dialect and specifically concen-

trated on providing the learner with rich examples of native phraseology (Mac Clúin 1922a).

Tá saibhreas flúirseach fairsing ins an nGaoluinn bheo a dheinfeadh caocha is dalla ar scoláirí na leabhartha (Mac Clúin 1922a, Preface)²

2.2. Peadar Ó Laoghaire

It was during this period that Peadar Ó Laoghaire began his literary career. His self-professed and unique style of writing became known as ‘*caint na ndaoine*’ (the speech of the people). It was Ó Laoghaire’s profound belief that a new style of literature was needed to reflect the speech of native Irish-language speakers and that same literature should be based on the spoken language. He adamantly believed that literature should reflect the speech of the native speakers. It was this style that fell within the primary aims of *Conradh na Gaeilge* and which garnered Ó Laoghaire praise both nationally and internationally as a writer (Ó Macháin 2015).

Whether as a writer of Irish, or as an exponent of Irish idiom, Canon O’Leary’s work stands the crucial test of time. It is impossible to overstate the debt that Irish grammar owes to him. (O’Rahilly 1926: iii)

Peadar Ó Laoghaire was born in 1839 in Lios Carragáin in the Muskerry Gaeltacht in County Cork. This area consisted of a large population of native Irish-language speakers and the county of Cork included more than one fifth of the total Irish-speaking population of Ireland during Ó Laoghaire’s youth (Ó Cuív 1951: 26). Despite the number of Irish-language speakers in Ireland at that time, Ó Háinle (1994: 754) states that only a very limited number of these speakers were literate and Nic Pháidín (1982) notes that even though Irish was the main language of converse, it was not being widely used as a medium for written communication.

However, the establishment of *Irisleabhar na Gaedhilge* in 1882 and *Conradh na Gaeilge* in 1893 changed this. Borthwick (LNÉ Ls. G 1,285(1)) suggests that *Irisleabhar na Gaedhilge* was the first functional medium that Ó Laoghaire found for his work and Gaughan (1970: 81) submits that it was not until the foundation of *Conradh na Gaeilge* that Ó Laoghaire recognized that he “really began to live in a worthy sense”. Ó Domhnaill (1951: 10) notes that if Ó Laoghaire’s unending campaign to promote ‘*caint na ndaoine*’ as a written medium had not succeeded, the Irish language would no longer be a living

² There is a great richness to be found in the living language which would dumbfound the scholars. (All translations are the author’s own unless stated otherwise).

language. This new style, which focused on the publication of native speech instead of classical Irish, laid the foundation for a new approach towards the publication of Irish and influenced generations of scholars throughout the twentieth century.

His Irish is “racy of the soil,” it nowhere denies the Munstersman, but it is also the Irish of an educated man who has learned to express himself in the language. And it is the reason why I have made a point of procuring Father O’Leary’s writings, ever since my attention was first drawn to him thirteen years ago. (Zimmer 1911)

It was this style of writing that made Ó Laoghaire’s publications more accessible to the general population who enjoyed his work being read aloud (O’Leary 1994: 13–4). Ó Laoghaire published more than 500 individual pieces of work, including works of original prose, translations, dramas, journal and newspaper articles, and it was this body of work, which was populated with native Irish-language idioms, that increased his popularity as a writer. Ó Laoghaire made a concerted effort to include native phrases, idioms and utterances in each of his publications and it is this aspect of Ó Laoghaire’s published work that is the focus of this paper. Ó Laoghaire’s seminal work *Séadna* (Ua Laoghaire 1904) was praised not only as a work of literature but also as an exemplary basis for modern Irish literature.

The formative influence of *Séadna* is likely to be great... there is such a thing as the ‘style’ of a period, or the ‘style’ of a national literature... We think in *Séadna* that An tAthair Peadar points the way in which Irish writers should march. (Mac Piarais 1904: 8)

The idioms in Ó Laoghaire’s work provide an insight into the primary features and characteristics of Irish-language idioms. The methodology relating to the collection and creation of the idiom database will be briefly discussed in the following section and the primary features of the collected idioms will be examined in Section 4.

3. Methodology

To facilitate the search and extraction of idioms from Ó Laoghaire’s published work, a lexicon of idioms was manually compiled from the two primary Irish-language dictionaries of the twentieth century. Due to the limited amount of phraseological research undertaken on Irish-language idioms, specifically the lack of a comprehensive lexicon of Irish-language idioms or the availability of an idiom dictionary which could be used as a founda-

tion for research, a new approach to search and extract idioms was needed. A more indepth description of the compilation and the contents of this idiom lexicon is available in Ní Loingsigh (2016).

The most common lemmas from this lexicon were categorized in order of frequency and were used to search a corpus of Ó Laoghaire's published work which was compiled using *Sketch Engine* tools (Kilgariff et al. 2004) and a morphological analyser and a part-of-speech tagger (Uí Dhonnchadha 2009). The search methodology, which used "idiom-prone words" (O'Keefe, McCarthy and Carter 2007) or the most frequent lemmas from the idiom lexicon to search the corpus, will not, however, be examined in this paper. A detailed analysis of the search methodology itself is available in Ní Loingsigh and Ó Raghallaigh (2018). The collected idioms were recorded in a database which was created in *Léacsclann*, an on-line platform used for building dictionary writing systems and terminology management systems as well as other lexicographic and reference applications (Měchura 2012) and is used in various research projects developed in Fiontar & Scoil na Gaeilge, Dublin City University (Ó Raghallaigh and Měchura 2014). The collected idioms were recorded in canonical form and classified semantically and syntactically. Each idiom has up to three usage examples from the corpus of Ó Laoghaire's published work and is paraphrased in the database. A detailed overview of the database schema is available in Ní Loingsigh and Ó Raghallaigh (2016).

4. Results

An excess of 1,500 idioms were collected during the corpus search but only 420 idioms were recorded in the database in their baseform or canonical form. Each entry in the database contains up to three usage examples from the corpus. Even though the idioms collected and recorded in the database are dialect specific, references to 62 per cent of the idioms collected were found in other lexicographic resources – primarily the two main Irish-English dictionaries published during the twentieth century, *Foclóir Gaeilge-Béarla* (Ó Dónaill 1977) and *Foclóir Gaedhilge agus Béarla* (Dinneen 1927). The remaining 38 per cent of idioms collected were paraphrased following guidance from native speakers in the Muskerry region.

Two chief conclusions can be garnered from these results. Firstly, it is accepted that Ó Laoghaire's work left a lasting imprint on the Irish language and on a generation of writers during the twentieth century (Buttimer 2011). It could be suggested that Ó Laoghaire's work was central to the widespread

adoption of certain idioms and resulted in certain dialectal idioms being brought into more general usage. Alternatively, it could be argued that many of the idioms collected in the corpus search are non-dialect specific idioms, but general widespread idioms used throughout Ireland. However, further research is needed to confirm this and to confirm the number of idioms which are still currently in widespread use. Additionally, it could be suggested that the collected idioms are not examples of Irish-language idioms but examples of linguistic creativity or a stylistic device employed by Ó Laoghaire as a prose author. However, Ó Laoghaire's adamant conviction on using only native speech in his publications confirms that the idioms collected and analysed in this paper are native in origin and an authentic representation of Irish-language idioms.

Avoid Provincialisms. Not ! I am determined to write down most carefully every provincialism I can get hold of. Then I shall be sure to have the people's language at least in that province... A person who writes carefully and exact the living speech of any of our three Irish-speaking provinces writes Irish which comes a great deal nearer to the Irish of those others than Keating comes to any of them. (O'Leary 1900)

The following section of the paper will discuss the primary features and characteristics presented in the collected idioms.

4.1. Prepositional idioms

Each idiom in the database consists of a baseform or canonical form or "the simplest morfo-syntactic form that an expression can take" to activate its specific idiomatic meaning and function (Barkema 1996: 141). When evidence from the corpus suggested that certain verbs and prepositions formed an integral part of the idiom, these prepositions were included as part of the baseform. This reflects the strong semantic link that exists between certain prepositions and verbs in Irish-language syntax (Ó Baoill and Ó Domhmalláin 1975). When evidence from the corpus suggested that a preposition was central to the idiomatic meaning of the idiom being analysed, the relevant preposition was enclosed in parentheses.

...bíonn brí nó ciall ar leith leis an aonad seo den bhriathar agus den réamhfhocail. Cuirtear an bhrí seo in iúl go minic i dteangacha eile le briathra anois. Sa Ghaeilge cuirtear briathra agus réamhfhocail le chéile mar shaghas aonaid a mbíonn ciall faoi leith leis. (ibid. 7)³

³ ...this unit containing a verb and preposition has a unique meaning. This type of meaning is often conveyed in other languages with a verb. A verb and preposition are combined in Irish to form a type of unit that has a unique meaning associated with it.

This point is further reflected in Wagner's (1959) publication *Das Verbum in den Sprachen der Britischen Inseln*, and highlights the importance of the preposition in relation to the semantic meaning of certain units in the Irish language. The inclusion of the preposition in the baseform of the idiom was incorporated to ensure the proper semantic meaning of the idiom was recorded and displayed in the database.

- (1) *aghaidh na muc 's na madraí a thabhairt (ar)* (lit. to give the face of the pigs and the dogs to someone)⁴
- (2) *na píobáin a bhaint (as)* (lit. to take the pipes from someone)⁵

The collected idioms not only highlight the importance of prepositions in certain idioms but also provide several examples of language change and shift through the modification of prepositions in certain idioms. The change in use of certain prepositions does not affect the overall semantic meaning of the idiom and can be seen in (3) below.

- (3) *an lug ar an lag a thitim (ag)* (lit. the *lug* fell on the *lag* on someone)⁶

An alternative version of this idiom, *an lug ar an lag a thitim (ar)*, can be found in modern usage which reflects the modification of the preposition *ag* to *ar* but the retention of the fundamental idiomatic meaning of the idiom. Both versions of this idiom were examined in *Corpas na Gaeilge, 1882–1926* (<http://corpas.ria.ie/>), the primary Irish-language historical corpus available for this period, and *The New Corpus for Ireland (NCI)* (<http://corpas.focloir.ie/>), the primary corpus of modern Irish currently available. The original idiom containing the preposition *ag* was found in both corpora but the more current version of the idiom containing the preposition *ar* was only found in NCI. Additionally, the original idiom only appeared in publications by writers associated with the southern dialect of Irish. This suggests that Ó Laoghaire's work could possibly have helped the spread of this dialectal idiom and helped transfer it into the modern vernacular, albeit in varied form.

4.2. Numerals in idioms

Numerals feature as integral components in many of the collected idioms which reflects the importance of certain numbers in Irish-language folklore.

⁴ Idiomatic meaning: to rile someone.

⁵ Idiomatic meaning: to fight ferociously.

⁶ Idiomatic meaning: to lose courage, to give way to despair.

Like 3 and 9 treated as a potent and formal number; 7 crosses were used for consecrating altars (5 acc. to L. B.), and groups of ecclesiastical ruins are called “seven churches”; 7, as also 4, was a favourite Celtic number in public organisation (O’Leary). (Dinneen 1927, s.v. *seacht*)

The use of numerals in Irish-language idioms can be seen in many of the collected idioms and provide evidence of the cultural importance of specific numbers in the Irish language.

- (1) *ceithre shaol duine a fháil* (lit. to get four lives of a person)⁷
- (2) *ó chúig cúigt na hÉireann* (lit. from the five fifths of Ireland)⁸

4.3. Somatic idioms

Idioms containing body parts, so-called *somatisims* or *somatic* idioms, are a central class of idioms in many languages and have been widely researched (Piirainen 2016). The principal body parts included in the collected idioms examined in this paper include idioms containing a reference to the head, leg or foot and hand. Additionally, numerous references are also made to specific body parts relating to the head, e.g. eye, nose, ear and mouth. However, many more specific body parts, as seen in (3) and (4) including references to internal organs were also recorded. A more detailed analysis of this central category of idioms is needed to fully explore the various categories of somatic idioms found in Irish-language idioms.

- (3) *faoina fhiaccla* (lit. under his teeth)⁹
- (4) *d’ainneoin a chúlfiaccla* (lit. in spite of his backteeth)¹⁰

4.4. Idiom syntax

An indepth analysis of the syntactic structures of Irish-language idioms has yet to be undertaken on Irish-language idioms. However, certain syntactic structures and patterns occurred frequently in the idioms that were analysed in this paper. The primary features of the main syntactic structures found in the collected idioms are presented in the following subsections. Further analysis is needed to develop this area of research but these structures provide a base for future research on the syntax of Irish-language idioms.

⁷ Idiomatic meaning: to live very long.

⁸ Idiomatic meaning: from all over Ireland, from near and far.

⁹ Idiomatic meaning: to mutter something.

¹⁰ Idiomatic meaning: in spite of him.

4.4.1. Irreversible binomial idioms

- (5) *gan poll ná póirse a fhágáil gan chuardach* (lit. without leaving any hole or porch unsearched)
- (6) *scéal ná duain* (lit. neither story nor poem/song)¹¹
- (7) *tóin ná ceann* (lit. neither bottom nor head)¹²

4.4.2. Idioms beginning with a compound preposition

- (8) *ar muin na muice* (lit. on the pig's back)¹³
- (9) *ar nós na gaoithe* (lit. like the wind)¹⁴
- (10) *ar ór na cruinne* (lit. for the gold of the world)¹⁵

4.4.3. Idioms beginning with a verbal noun

- (11) *ag crith ina chraiceann* (lit. shaking/trembling in his skin)¹⁶
- (12) *ag stracadh na bpíobán as a chéile* (lit. tearing the pipes out of each other)¹⁷
- (13) *ag tarraingt uisce chun a mhuilinn féin* (lit. pulling water to one's own mill)¹⁸

4.4.4. Idioms containing open function slots (Barkema 1996)

- (14) *ó Shamhain go Bealtaine* (lit. from November until May)¹⁹
- (15) *ó Luan go Satharn* (lit. from Monday to Saturday)²⁰
- (16) *ó bhaithis go bonn* (lit. from the top of the head to the sole of the foot)²¹

¹¹ Idiomatic meaning: neither tale nor tidings.

¹² Idiomatic meaning: can't make sense of.

¹³ Idiomatic meaning: to be very happy, content.

¹⁴ Idiomatic meaning: to travel very fast.

¹⁵ Idiomatic meaning: something that one will never do.

¹⁶ Idiomatic meaning: to be very afraid.

¹⁷ Idiomatic meaning: to physically fight with each other.

¹⁸ Idiomatic meaning: actuated by self-interest.

¹⁹ Idiomatic meaning: for a long time.

²⁰ Idiomatic meaning: continually, for a long period.

²¹ Idiomatic meaning: to be completely covered.

4.5. Emphasis in idioms

Specific words form part of the baseform of idioms and place a stronger stress on the idiomatic meaning of the idiom in certain cases. It was noted during the analysis of the collected idioms that certain words were used more often than others to place an added emphasis of meaning in certain idioms, e.g. use of numerals, colours and additional elements to add stress to the idiomatic meaning of the idiom.

4.5.1. Numerals

(17) *ó chúig cúigí na hÉireann* (lit. from the five fifths of Ireland)²²

(18) *ceangal na gcúig gcaol* (lit. the binding/tie of the five limbs)²³

4.5.2. Colours, e.g. 'red'

(19) *ina chogadh dhearg* (lit. to be a red war)²⁴

4.5.3. Additional components

Several of the collected idioms analysed contain a variant of the general baseform which included an additional element to add emphasis to the idiom as in (20) and (21).

(20) *ní fiú biorán (gan tóin) é* (lit. it is not worth a pin (without an end/bottom))²⁵

The addition of *gan tóin* to the baseform of this idiom adds a further stress to the idiom and emphasizes the idiomatic meaning of the idiom. Similarly, the repetition and reversal of the baseform of the idiom in (21) emphasizes the length of time being described. However, in both these examples, the additional element does not affect the underlying idiomatic meaning of the idiom.

(21) *ó Shamhain go Bealtaine (agus ó Bhealtaine go Samhain arís)* (lit. from November to May (and from May to November again))²⁶

²² Idiomatic meaning: from all over Ireland, from near and far.

²³ Idiomatic meaning: to be bound hand and foot.

²⁴ Idiomatic meaning: a ferocious fight or battle.

²⁵ Idiomatic meaning: it is worthless.

²⁶ Idiomatic meaning: a very long time.

4.6. Fossilized idioms

As noted in section 4.1 of this paper, the baseform of an idiom, or specific components of an idiom's baseform, can change over time. Each of the idioms analysed in this paper are recorded in their baseform and in standardized Irish following *An Caighdeán Oifigiúil* (Rannóg an Aistriúcháin 1958), the first official grammatical standard published for the Irish language. However, in certain cases, specific idioms were not standardized but recorded in their fossilized non-standardized form. The decision to leave fossilized idioms in non-standardized form was a subjective decision.

Following discussions with native speakers from the Muskerry region, a number of idioms were left in non-standardized form as the standardization of certain components of idioms would lead to the loss of idiomatic meaning. The standardized baseform was recorded when lexicographic evidence existed of this form in published lexicographic works. In cases where no lexicographic evidence of the standardized form existed, the baseform was recorded in non-standardized form based on evidence and advice from native speakers from the Muskerry region.

(22) *ag stracadh na bpíobán as a chéile* (lit. tearing the pipes out of each other)²⁷

Non-standardized component: *stracadh* was used in the baseform instead of the standardized form *sracadh* reflecting dialectal pronunciation of this idiom.

(23) *an chú is an cat is an giorria ar aon urlár* (lit. the hound, the cat and the hare on one floor)²⁸

Non-standardized component: *an chú*²⁹ was used in the baseform instead of the standardized form *an cú* reflecting dialectal pronunciation of this idiom.

(24) *do ló agus d'óiche* (lit. by day and by night)³⁰

Non-standardized component: *ló*³¹ was used in the baseform instead of the standardized form *lá* reflecting dialectal pronunciation of this idiom.

²⁷ Idiomatic meaning: to physically fight with each other.

²⁸ Idiomatic meaning: everything in a mess, turned upsidedown.

²⁹ *Cú* (hound, greyhound) is a feminine noun in the southern dialect of Irish and takes a lenition mark. This lenition mark is retained in the baseform of this idiom to reflect pronunciation and use. *Cú* is a masculine noun according to the official standard and masculine nouns do not usually take a lenition mark.

³⁰ Idiomatic meaning: continually.

³¹ Obsolete dative form of *lá*.

- (25) *ó thigh an deamhain go tigh an diabhair* (lit. from the demon's house to the devil's house)³²

Non-standardized component: *tigh* was used in the baseform instead of the standardized form *teach* reflecting dialectal pronunciation of this idiom.

5. Conclusion

The aim of this paper was to provide an overview of the primary features of Irish-language idioms. A general outline of Irish-language idiom research was presented along with an insight into Peadar Ó Laoghaire's background and his work. The methodology in relation to the collection and classification of idioms was explained and the principal results were discussed in the latter section of the paper. This overview hopes to provide an insight into research undertaken to date on Irish-language idioms along with providing a deeper understanding of the principal features of Irish-language idioms. This paper provides a base for future analysis and research in this area of research and for future comparison of idioms in other Irish dialects as well as other languages.

References

- Abdou, Ashraf. 2012. *Arabic idioms: a corpus based study*. London: Routledge.
- Barkema, Herman. 1996. Idiomaticity and terminology: a multi-dimensional descriptive model. *Studia Linguistica*. 50(2): 125–160.
- Bergin, Osborn Joseph. 1911. Reply from Osborn J. Bergin, Ph.D., F.R.U.I. In: Irish Book Company (eds.) *A statement of the position of modern Irish in the National University*. Dublin: Irish Book Company.
- Buttimer, Cornelius. 2011. Séadna: saothar sagairt. *Feasta*. 64(10): 19–20.
- Dinneen, Patrick Stephen. 1927. *Foclóir Gaedhilge agus Béarla, being a thesaurus of the words, phrases and idioms of the modern language*. 2nd ed. Dublin: Irish Texts Society.
- Fernando, Chitra. 1996. *Idioms and idiomaticity*. Oxford: Oxford University Press.
- Gaughan, John Anthony. 1970. *Doneraile*. Dublin: Kamac Publications.
- Hanks, Patrick. 2004. The syntagmatics of metaphor and idiom. *International Journal of Lexicography*. 17(3): 245–274.
- Hogan, Edmund Ignatius. 1898. *A Handbook of Irish Idioms*. Dublin: Sealy, Bryers & Walker.

³² Idiomatic meaning: out of the frying-pan into the fire.

- Howarth, Peter. 1998. Phraseology and second language proficiency. *Applied Linguistics*, 19(1): 24–44.
- Kilgarriff, Anthony, Rundell, Michael and Uí Dhonnchadha, Elaine. 2006. Efficient corpus development for lexicography: building the New Corpus for Ireland. *Language and Resources and Evaluation*, 40(2): 127–152.
- Kilgarriff, Anthony, Rychly, Pavel, Smrz, Pavel and Tugwell, David. 2004. The Sketch Engine. In: Williams, G. agus Vessier, S. (eds.) *Proceedings of the 11th Euralex International Congress, 6–10 July 2004*, Lorient, France, 105–115. <https://tinyurl.com/3kja5cuk>. Retrieved: 3.11.2012.
- Mac Clúin, Seóirse. 1922a. *Réilthíní Óir. Cuid I. Baile Átha Cliath: Comhlucht Oideachais na hÉirean*.
- Mac Clúin, Seóirse. 1922b. *Réilthíní Óir. Cuid II. Baile Átha Cliath: Comhlucht Oideachais na hÉirean*.
- Mac Maoláin, Seán. 1933. *Cora Cainnte as Tír Chonaill*. Baile Átha Cliath: Oifig Díolta Foilseacháin Rialtais.
- Mac Mathúna, Liam. (ed.) 1987. *Séadna*. Baile Átha Cliath: Carbad.
- McKenna, Lambert Andrew Joseph. 1922. *English-Irish phrase dictionary*. Dublin: M.H. Gill & Son.
- Měchura, Michal Boleslav. 2012. Léacsclann: A platform for building dictionary writing systems. In: Fjeld, Ruth Vatvedt and Torjusen, Julie Matilde (eds.) *Proceedings of the 15th Euralex International Congress, 7–11 August 2012*, Oslo, Norway, 855–861. <https://tinyurl.com/29twwppw>. Retrieved: 23.11.2012.
- Nic Phóidín, Caoilfhionn. 1982. Irisleabhar na Gaedhilge 1882–1909. *Comhar*. 41(11): 4–16.
- Ní Loingsigh, Katie. 2016. Towards a lexicon of Irish-language idioms. In: *Proceedings of the Second Celtic Language Technology Workshop (CLTW)*, Paris, France, 69–78. <https://tinyurl.com/y8cs8vt5>. Retrieved: 14.08.1017.
- Ní Loingsigh, Katie and Ó Raghallaigh, Brian. 2018. Harping on a new string: the search for Irish-language idioms. In: Soutet, Olivier, Sfar, Inès and Mejri, Salah (eds.) *Phraséologie et discours*. Paris: Honoré Champion, 393–401.
- Ní Loingsigh, Katie and Ó Raghallaigh, Brian. 2016. Starting from Scratch – The Creation of an Irish-language Idiom Database. In: Margalitadze, Tinatin and Meladze, George (eds.) *Proceedings of the 17th Euralex International Congress, 5–10 September 2016*, Tbilisi, Georgia, 726–734. <https://tinyurl.com/jwccrrhch>. Retrieved: 20.09.2016.
- O’Keefe, Anne, McCarthy, Michael and Carter, Ronald. 2007 *From corpus to classroom: language use and language teaching*. Cambridge: Cambridge University Press.
- O’Leary, P. 1994. *The prose literature of the Gaelic Revival, 1881–1921: ideology and innovation*. Pennsylvania: Pennsylvania State University Press.
- O’Leary, Peter. 1900. Comhairle. Rosg Catha Bhriain. *An Claidheamh Soluis*, 13 January 1900. 700–701.

- O’Rahilly, Thomas Francis. 1926. Preface. In: O’Leary, P. *Papers on Irish Idiom*. Dublin: Browne and Nolan.
- Ó Baoill, Dónall and Ó Domhnaill, Tomás. 1975. *Réamhfhocail le briathra na Gaeilge*. Baile Átha Cliath: Institiúid Teangeolaíochta Éireann.
- Ó Corráin, Ailbhe. 1989. *A concordance of idiomatic expressions in the writings of Séamus Ó Grianna*. Belfast: The Institute of Irish Studies, Queen’s University Belfast.
- Ó Cuív, Brian. 1951. *Irish dialects and Irish speaking districts: three lectures*. Dublin: Dublin Institute for Advanced Studies.
- Ó Dónaill, Niall. 1977. *Foclóir Gaeilge-Béarla*. Baile Átha Cliath: An Gúm.
- Ó Háinle, Cathal. 1994. Ó chaint na ndaoine go dtí an caighdeán oifigiúil. In: McCone, Kim, McManus, Damian, Ó Hainle, Cathal, Williams, Nicolas and Breatnach, Liam (eds.) *Stair na Gaeilge: in ómós do Pádraig Ó Fiannachta*. Maigh Nuad: Roinn na Sean-Ghaeilge, Coláiste Phádraig, 745–793.
- Ó Macháin, Pádraig. 2015. Aitheasc na comhdhála. In: Ó Raghallaigh, Eoghan. (ed.) *Léachtaí Cholm Cille XLV: saothar an Athar Peadar*. Maigh Nuad: An Sagart, 7–9.
- Ó Máille, Tomás. 1937. *An Béal Beo*. Baile Átha Cliath: Comhlucht Oideachais na hÉireann.
- Ó Raghallaigh, Brian and Měchura, Michal Boleslav. 2014. Developing high-end reusable tools and resources for Irish-language terminology, lexicography, onomastics (toponymy), folkloristics, and more, using modern web and database technologies. In: *Proceedings of the First Celtic Language Technology Workshop*, Dublin, Ireland, August 2014, 66–70. <http://www.aclweb.org/anthology/W14-4610>. Retrieved: 24.09.2016.
- Piirainen, Elizabeth. 2016. *Lexicon of Common Figurative Units. Widespread Idioms in Europe and Beyond. Volume II*. New York: Peter Lang.
- Rannóg an Aistriúcháin. 1958. *Gramadach na Gaeilge agus litriú na Gaeilge: an caighdeán oifigiúil*. Baile Átha Cliath: Oifig an tSoláthair.
- Ua Dubhghaill, Séamus and Bairéad, Stiophán. 1901. *Leabhar Cainte*. Baile Átha Cliath: Connradh na Gaedhilge.
- Ua Laoghaire, Peadar. 1904. *Séadna*. Baile Átha Cliath: Irish Book Company.
- Uí Dhonnchadha Elaine. (2009). *Part-of-speech tagging and partial parsing for Irish using finite-state transducers and constraint grammar*. Ph.D. diss., Dublin City University.
- Wagner, Heinrich. 1959. *Das Verbum in den Sprachen der britischen Inseln; ein Beitrag zur geographischen Typologie des Verbums*. Tübingen: Niemeyer.
- Zimmer, Heinrich. 1911. Appendix: reply from the Rev. Richard Henebry, Ph.D. In: Irish Book Company (eds.) *A statement of the position of modern Irish in the National University*. Dublin: Irish Book Company, 11–22.

Unpublished manuscripts

- LNÉ (Leabharlann Náisiúnta na hÉireann/National Library of Ireland). Ls. G 1,285 (1) 3.3.1.1 (7), 26 March 1920 (“Shán Ó Cuív Papers”).

Irlandzkie idiomy: przegląd i analiza idiomów z opublikowanego dzieła Peadara Ó Laoghaire

Streszczenie

Artykuł przedstawia główne cechy i właściwości irlandzkich idiomów – zwłaszcza jednostek zebranych w opublikowanym dziele Canon Peadar Ó Laoghaire (1838–1920), najważniejszego irlandzkiego prozaika z przełomu XIX i XX wieku. Zanalizowano i opisano główne cechy zebranych idiomów. Analiza przeprowadzona na potrzeby badania daje obraz składni i semantyki irlandzkich idiomów. Do tej pory niewiele uwagi poświęcano irlandzkiej idiomatyce, więc niniejszy artykuł przybliży stosunkowo nowy obszar badań naukowych nad językiem irlandzkim oraz stanowi podstawę dla kolejnych badań i porównań z idiomami z innych odmian terytorialnych, jak również przyszłych badań powiązanych z idiomami.

Daniela Prutscher

Universität Jena

Deutschland

<https://orcid.org/0000-0003-4693-5832>

Phraseme und Wortverbindungen in der deutschen Wissenschaftssprache¹

Abstract. Wissenschaftssprache zeichnet sich unter anderem durch die Verwendung von spezifischen Wortverbindungen und musterhaften Formulierungsweisen aus. Anhand eines Korpus, das aus 500 wissenschaftlichen Artikeln aus den Jahren 2010 bis 2017 besteht, werden aktuelle Kollokationen und Textroutinen für die Bereiche Germanistik, Glottodidaktik und Medizin ermittelt und nachgewiesen.

Schlüsselwörter: *Wissenschaftssprache, Wortverbindungen, Kollokationen, Phraseme, sprachliche Muster*

1. Einleitung

Das von der Deutsch-Polnischen Wissenschaftsstiftung geförderte und an der Universität Ermland-Masuren in Olsztyn und der Friedrich-Schiller-Universität in Jena angesiedelte Projekt "D/P-IPHRAS: Interphraseologie als Element der Wissenschaftssprache"² widmet sich dem Phänomen von (mehr oder weniger) festen Wortverbindungen in der deutschen und polnischen Wissenschaftssprache.

Für den Wissenschaftsdiskurs beider Sprachen relevante und typische Wendungen werden hier anhand aktueller Texte korpuslinguistisch ermit-

¹ Dieser Beitrag konnte durch die Unterstützung der Deutsch-Polnischen Wissenschaftsstiftung entstehen, die das Projekt *D/P-IPHRAS: Interphraseologie als Element der Wissenschaftssprache* (Projektnummer 2017-07) finanzierte.

² Leitung: Dr. Bettina Bock (Jena); Mitarbeiterinnen: Dr. Monika Czerepowicka, Dr. Magdalena Makowska und Dr. Joanna Targońska (Olsztyn) sowie Francis Gieseke-Golembowski und Daniela Prutscher (Jena).

telt und über die Datenbank des schon bestehenden Portals "IPHRAS – Interphraseologismen für Studien- und Berufsmobile" (www.iphras.eu) zugänglich gemacht. Dieses Portal stellt bereits für den Bereich "Studium und Beruf im Ausland" Phraseme zum Thema "Bewerbung" und allgemeine Routineformeln in den Sprachen Bulgarisch, Deutsch, Englisch, Griechisch, Romani, Rumänisch und Türkisch zum Üben und Lernen zur Verfügung. Das aktuelle Projekt ergänzt diese Datenbank um das Polnische als einer weiteren im Kontext der Mehrsprachigkeit wichtigen Sprache Europas.

Die für das D/P-IPHRAS-Projekt ermittelten Daten sollen mit einem praxisnahen Tool eine Hilfe für alle sein, die deutsche oder polnische wissenschaftliche Texte verfassen. Der Fokus liegt dabei sowohl auf (polnischen) Studierenden der Germanistik und (deutschen) Polnisch-als-Fremdsprache-Studierenden als auch den muttersprachlichen Studierenden dieser Fächer. Deshalb wurden für das Korpus, das aus jeweils 500 wissenschaftlichen Artikeln in deutscher und polnischer Sprache besteht, Texte aus den für die Studienrichtungen Germanistik und Polonistik besonders relevanten Bereichen Sprachwissenschaft, Literaturwissenschaft und Glottodidaktik³ gewählt. Als Kontrast zu den geisteswissenschaftlichen Beiträgen wurden für beide Sprachen Artikel aus dem Fachgebiet Medizin hinzugenommen.

2. Wissenschaftssprache

Die alltägliche oder allgemeine Wissenschaftssprache stellt über die Fachtermini der einzelnen Disziplinen hinausgehend den "Anteil der für wissenschaftliche Zwecke verwendeten Sprache, der in allen Fächern bekannt, verwendbar und – mehr oder weniger frequent – auch im Gebrauch ist" (Graefen 2001: 191) dar. Sie ist "Bestandteil, Resultat und zugleich Voraussetzung der Wissenschaftskommunikation, also unabdingbar für jeden, der sich am deutschen Wissenschaftsbetrieb beteiligen will" (Ehlich & Graefen 2001: 373). Zu ihr zählen "die fundamentalen sprachlichen Mittel [...], derer sich die meisten Wissenschaften gleich oder ähnlich bedienen" (Ehlich 1993: 33). Als relevante Merkmale der Wissenschaftskommunikation gelten beispielsweise die vier Gebote *Ökonomie*, *Präzision*, *Origo-Exklusivität* und *Dis-*

³ Der Begriff *Glottodidaktik* wird vor allem in Polen verwendet und dort einerseits synonym zu dem im deutschen Sprachraum verbreiteten Terminus Sprachlehr- und -lernforschung verwendet und bezeichnet andererseits auch den Prozess des Unterrichts einer Fremdsprache (vgl. Targońska 2014: Anmerkung 18).

kussion.⁴ Andere für die Wissenschaftssprache wichtige Handlungsmuster sind *Fragen, Begründen, Argumentieren* oder *Demonstrieren* (Ehlich 1993: 26 ff.).⁵

Fest steht auch, dass “[e]s [...] sich bei der allgemeinen Wissenschaftssprache um ein relativ festes Repertoire von Ausdrucksweisen [handelt], das zwecks Erhaltung seiner Funktionalität nicht beliebig spontan erweiterbar oder variierbar ist” (Fandrych & Graefen 2010: 512), denn “[h]och differenzierte kulturelle Ordnungen wie [...] die wissenschaftliche Domäne zeichnen sich durch den Gebrauch eines spezialisierten Ausdrucksspektrums aus. Die Verwendung dieser Ausdrücke ist für jeden, der an der Kommunikation in der Domäne partizipieren will, sozial bindend” (Steinhoff 2007: 100 f.)

Die Wissenschaftssprache zeichnet sich daneben durch bestimmte Formulierungsmuster aus. Diese Musterhaftigkeit der Ausdrucksweisen kann auch durch Kollokationen deutlich werden (vgl. Feilke 2012: 17).

Kollokationen oder feste Wortverbindungen⁶ gehören auch in den Bereich der Phraseologie, in der es jedoch eine Vielzahl von Termini gibt, die – abhängig vom jeweiligen Forschungsschwerpunkt – unterschiedliche Phänomene bezeichnen können. Bereits 1981 hat Klaus-Dieter Pilz konstatiert, dass

[b]isher [...] über die Terminologie der Phraseologie sehr wenig reflektiert worden [ist]. Deshalb konnte sie (vor allem die deutschsprachige) so katastrophal ausufern. Überspitzt gesagt: Jeder, der sich mit phras.[eologischen] Phänomenen befasste, schuf sich eine eigene Terminologie. (Pilz 1981: 27)

Auch 20 Jahre später hat sich daran nicht viel geändert:

Wortübergreifende Phänomene werden als Phraseologismen **oder** als Kollokationen **oder** als Valenz **oder** als syntaktische Konkurrenz **oder** als Funktionsverbgefüge **oder** als Regelsystem semantischer Vereinbarkeiten **oder** als kommunikative Formeln betrachtet. (Steyer 2000: 102; Hervorhebungen im Original)

⁴ Die Umsetzung dieser Gebote kann durch Attribuierungen, nominale und pronominale Wiederaufnahmen, Passivkonstruktionen zur Deagentivierung und semantische Relationen mit für die Wissenschaftskommunikation typischen syntaktischen Funktionstechniken erfolgen (vgl. Cziczka et al. 2012: 5–6).

⁵ Ein allgemeines *Ich*-“Verbot” (neben einem *Erzähl*- und einem *Metaphern*-Verbot), wie es Harald Weinrich (1989: 132 ff.) postulierte und Kretzenbacher (1994: 26) spezifizierte, kann im vorliegenden Korpus nicht bestätigt werden: Dieses Personalpronomen wird in 307 der 500 Aufsätze 2.403 mal verwendet, wobei die Treffer auf jeweils 99 Dateien der Sprach- und Literaturwissenschaft, auf 82 der Glottodidaktik und auf lediglich 27 der Medizin verteilt sind (die geringe Frequenz in den medizinischen Artikeln lässt sich auch darauf zurückführen, dass in diesem Fachgebiet hauptsächlich AutorInnenkollektive publizieren). Bei der Betrachtung der *ich*-Verwendung sollte allerdings immer zwischen Verfasser-*ich*, Forscher-*ich* und Erzähler-*ich* unterschieden werden (vgl. z. B. Deml 2015: 56–61).

⁶ Das von Kathrin Steyer herausgegebene *Jahrbuch des Instituts für deutsche Sprache* 2003 trägt den Titel “Wortverbindungen – mehr oder weniger fest”.

Als ein Oberbegriff bietet sich “Phrasem” an. In der weiten Auffassung Harald Burgers aus dem Jahr 2015 bestehen Phraseme

[erstens] aus mehr als einem Wort, zweitens sind die Wörter nicht für dieses eine Mal zusammengestellt, sondern es handelt sich um Kombinationen von Wörtern, die uns als Deutschsprechenden genau in dieser Kombination (eventuell mit Varianten) bekannt sind, ähnlich wie wir die deutschen Wörter (als einzelne) kennen. (Burger 2015: 11)⁷

Während lange Zeit im Allgemeinen Einigkeit darüber bestand, dass sich der Untersuchungsgegenstand der Phraseologie durch *Polylexikalität*, *Einheitsstatus*⁸, *Stabilität*, und *Idiomatizität* auszeichne, erfolgte (auch unter dem Einfluss der Korpuslinguistik) in den letzten Jahren ein Perspektivwechsel “weg von den auffälligen Idiomen hin zu den unauffälligen Kollokationen und musterhaften Konstruktionen” (Handwerker 2010: 249). Sowohl die Konstruktionsgrammatik als auch Erfahrungen aus der Fremdsprachendidaktik haben zusammen mit empirischen Ergebnissen zu der Erkenntnis geführt, dass “[m]ehr oder weniger feste Wortverbindungen [...] keine Sonder-, sondern Normalfälle sprachlicher Zeichenbildung dar[stellen]” (Ágel 2004: 65).

Vor der Arbeit mit großen Korpora wurde Wortverbindungen die Stabilität häufig intuitiv zugeschrieben, nun können tatsächliche Frequenzen ermittelt werden (die unter Umständen dem eigenen Sprachgefühl widersprechen).

In den letzten Jahren wurden auch Forschungsfragen aus dem Bereich der Wissenschaftskommunikation vermehrt korpuslinguistisch bearbeitet und somit auf eine breitere empirische Basis gestellt.⁹ Mit solchen

⁷ Kürzer formuliert es Ken Farø: “Phraseme sind lexikalisierte (spatial) polylexikalische Sprachzeichen” (Farø 2006: 53).

⁸ Phraseme werden mental als Einheiten gespeichert.

⁹ Es muss jedoch angemerkt werden, dass bei bisherigen Korpusuntersuchungen fast ausschließlich Texte aus den Geisteswissenschaften und sogar oft nur aus der germanistischen Sprachwissenschaft bzw. Deutsch als Fremdsprache untersucht wurden. Eine Ausnahme ist jedoch beispielsweise die Untersuchung von Mirjam Weder, deren Korpus aus 100 Wissenschaftstexten die “Disziplinen Geschichte, Philosophie, germanistische Linguistik und Literatur sowie Medien-/Kulturwissenschaften” (Weder 2015: 206) umfasst. In Sarah Brommers Arbeit zu sprachlichen Mustern (2018) werden über 2.000 Texte aus der germanistischen Sprachwissenschaft und der Medizin untersucht. Beispiele für weitere aktuelle korpuslinguistische Untersuchungen sind etwa Franziska Wallners “Kollokationen in Wissenschaftssprachen” (aus dem Jahr 2014) oder die Arbeit Cordula Meißners zu “Figurative[n] Verben in der allgemeinen Wissenschaftssprache des Deutschen” aus dem gleichen Jahr. Aktuelle Projekte wie *GeWiss* (hier handelt es sich um ein mehrsprachiges Korpus zu den Bereichen wissenschaftliche und studentische Vorträge sowie Prüfungsgespräche; siehe <https://gewiss.uni-leipzig.de>; vgl. z.B. Fandrych 2017: 13–32 und Fandrych et al. 2012: 319–337) und *eurowiss* (Untersuchung deutscher

Arbeiten wird dem noch 2010 von Christian Fandrych und Gabriele Graefen bemängelten "Problem einer unzureichenden Datenlage"¹⁰ bei der Untersuchung von Wissenschaftssprache entgegengearbeitet.

Am richtigen Gebrauch von Kollokationen kann nicht nur der Grad der Beherrschung einer Fremdsprache festgemacht werden, auch im Bereich der Wissenschaftssprache geben feste Ausdrücke "wichtige Hinweise auf die wissenschaftssprachliche Kompetenz des Schreibers, indem sie die fachspezifischen Themen und Begriffe indizieren und den fachübergreifenden textlichen, sprachlichen und kognitiven Prozeduren Gestalt geben" (Feilke & Steinhoff 2003: 115). Diese Kompetenz muss in besonderem Maße von L2-Sprecherinnen und -sprechern erworben werden, aber auch MuttersprachlerInnen beherrschen diese Verbindungen nicht automatisch, da sie nicht immer den alltagssprachlichen entsprechen.

3. Das *D/P-IPHRAS*-Korpus

Für das *D/P-IPHRAS*-Projekt wurde ein Korpus erstellt, das aus insgesamt 1.000 wissenschaftlichen Artikeln in polnischer und deutscher Sprache aus den Jahren 2010 bis 2017 besteht. Für jede Sprache wurden je 125 Beiträge aus den Bereichen germanistische Sprachwissenschaft, germanistische Literaturwissenschaft, Deutsch als Fremdsprache bzw. Glottodidaktik und Medizin zusammengetragen, die aus jeweils 6 online zugänglichen Zeitschriften stammen – nur die medizinischen Artikel wurden aus 5 Fachmagazinen zusammengestellt. Die Texte wurden in der Korpusvorverarbeitung bereinigt, d.h. Abstracts, Fußnoten und Literaturverzeichnisse wurden ebenso wie längere Zitate und Literaturverweise im Fließtext entfernt. Die Metainformationen zu jedem Artikel beinhalten Zeitschrift, Jahr, Aufsatztitel, AutorInnenamen und – soweit ermittelbar – auch das Geschlecht¹¹. Das bereinigte deutsche Korpus hat einen Umfang von über 2 Millionen (genau 2.281.135) Token.

und italienischer Lehrveranstaltungen aus verschiedenen Wissenschaftsdisziplinen; vgl. Thielmann et al. 2014: 7–17) beschäftigen sich mit der gesprochenen, *GeSig* mit der geschriebenen Wissenschaftssprache – für alle drei wurden umfangreiche Korpora erhoben. *GeSig* verfolgt dabei z.B. das Ziel, "[d]as gemeinsame sprachliche Inventar der Geisteswissenschaften" (siehe <http://research.uni-leipzig.de/gesig/>) korpusmethodisch zu beschreiben.

¹⁰ "Zwar werden Korpora verwendet, diese stehen aber oft nicht für andere Analysen zur Verfügung, sind meist klein und insgesamt heterogen, also untereinander kaum vergleichbar" (Fandrych & Graefen 2010: 510).

¹¹ Vor allem bei den medizinischen Artikeln sind die Vornamen der AutorInnen häufig abgekürzt und das Geschlecht nicht immer ermittelbar.

Tabelle 1: Korpusaufbau

Bereich	Zeitschrift	Beiträge
Glottodidaktik	Acta Philologica	5
	Fadaf – Materialien Deutsch als Fremdsprache	5
	GFL – German as a foreign language	35
	InfoDaF – Informationen Deutsch als Fremdsprache	20
	ZiF – Zeitschrift für interkulturellen Fremdsprachenunterricht	40
	ZVPG – Zeitschrift des Verbandes Polnischer Germanisten	20
Literaturwissenschaft	Acta Philologica	10
	GJB – Goethe-Jahrbuch	20
	LiTheS – Zeitschrift für Literatur- und Theatersoziologie	25
	PBB – Beiträge zur Geschichte der deutschen Sprache und Literatur	25
	TP – textpraxis. Digitales Journal für Philologie	20
	ZVPG – Zeitschrift des Verbandes Polnischer Germanisten	25
Sprachwissenschaft	FRAGL – Freiburger Arbeitspapiere zur Germanistischen Linguistik	10
	PBB – Beiträge zur Geschichte der deutschen Sprache und Literatur	10
	Gesús – Sprache & Sprachen	15
	tekst i dyskurs – text und diskurs	30
	ZGL – Zeitschrift für germanistische Linguistik	35
	ZVPG – Zeitschrift des Verbandes Polnischer Germanisten	25
Medizin	Der Ernährungsmediziner	10
	Deutsche Zeitschrift für Sportmedizin	30
	GMS Journal for Medical Education	30
	Schmerzmedizin. Angewandte Schmerztherapie und Palliativmedizin	25
	ZfMER – Zeitschrift für Medizin-Ethik-Recht	30

Quelle: eigene Forschung.

Das Korpus ermöglicht es einerseits, *korpusbasiert* (*corpus-based*) zu arbeiten, d. h. für die Wissenschaftssprache als typisch angesehene Formeln auf ihr Vorhandensein hin zu überprüfen. Andererseits können damit auch Untersuchungen angestellt werden, die *korpusgesteuert* oder *datengeleitet* (*corpus driven*)¹² sind, denn so können beispielsweise bisher auch unbekanntes Wortver-

¹² "Korpusgesteuertes und korpusbasiertes Vorgehen stellen [...] zwei methodische Zugänge dar, die sich ergänzen: Aus datengeleitet ermittelten Beschreibungskategorien lassen sich gezielt

bindungen ermittelt werden. Von Interesse sind neben hochfrequenten Kollokationen auch bisher unbekannte formelhafte Sequenzen oder Paarformeln.

4. Datenanalyse

Die nachfolgenden Daten wurden zum einen mithilfe des Korpus-Analyse-Tools **LancsBox** (Lancaster University corpus toolbox)¹³ erhoben, das 2015 an der Universität Lancaster entwickelt wurde. Dieses Programm kann beispielsweise auch bei deutschen Texten ein Part-of-Speech-Tagging mit dem TreeTagger¹⁴ vornehmen, Sprachdaten visualisieren und Kookkurrenzen mit verschiedenen statistischen Parametern errechnen.

Daneben wurde auch das vom Institut für Deutsche Sprache (IDS) für das Projekt "Usuelle Wortverbindungen" entwickelte Analysewerkzeug zur Untersuchung syntagmatischer Strukturen auf der Basis von Korpusdaten **lexpan** (Lexical Pattern Analyzer)¹⁵ verwendet. Über KWIC-Listen können mit dem Programm explorative, korpusgesteuerte Untersuchungen von Mehrwortstrukturen und sprachlichen Mustern vorgenommen werden.

4.1. Fallanalyse 1: Das hochfrequente Lemma *Frage*

Im Korpus kann nun nach für die Wissenschaftssprache als typisch angesehenen Wortverbindungen gesucht werden. Als ein Beispiel sollen hier verbale Kollokationen zu der Basis *Frage* vorgestellt werden, die es ermöglichen, "Textäußerungen wissenschaftssprachlich angemessen zu artikulieren" (Feilke & Steinhoff 2003: 116). Es gelten jedoch nicht alle denkbaren verbalen Partner als "domänentypische Möglichkeiten" (ebd.):

Es ist nicht verständlich zu machen, warum "sich" in wissenschaftlichen Texten "Fragen stellen", sie "sich" aber nicht, was ja ebenfalls möglich wäre, "ergeben", warum man dort "Fragen nachgeht", ihnen aber nicht "folgt", oder auch, warum "sich" dort "Fragen aufdrängen", aber nicht "aufzwingen". (Steinhoff 2007: 88)

Hypothesen formulieren und korpusbasiert untersuchen. [...] Am Beginn der Untersuchung steht dabei die Beobachtung der Sprachdaten, wobei zunächst alle Evidenzen zu akzeptieren sind. Auf der Basis dieser Evidenzen werden dann Hypothesen gebildet und überprüft. Hierbei stehen induktives und deduktives, also korpusgesteuertes und korpusbasiertes Vorgehen im Wechselspiel" (Wallner 2014: 91).

¹³ <http://corpora.lancs.ac.uk/lancsbox/> (Brezina et al. 2015: 139–173).

¹⁴ 1995 von Helmut Schmid am Institut für maschinelle Sprachverarbeitung der Universität Stuttgart entwickelt: <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/> (vgl. Schmid 1995).

¹⁵ <http://uwv.ids-mannheim.de/lexpan/>.

“Verständlich” werden solche Präferenzen durch ihre Verwendung, denn “gebrauchte Formen werden zu Formen des Gebrauchs” (Feilke 2003: 2016). Oder um es anders zu formulieren: “Kollokationen sind Mehrworteinheiten, die sich durch usuellen Gebrauch konventionell verfestigt haben” (Belica & Perkuhn 2015: 213).

Was zeigt das Korpus? In 409 der 500 Artikel kommt das Lemma *Frage* 2.321-mal vor, wobei die meisten Treffer im sprachwissenschaftlichen Korpus (941-mal in 109 Texten) zu finden sind, gefolgt von der Glottodidaktik (655-mal in 114 Texten) und der Literaturwissenschaft (424-mal in 104 Texten). Am seltensten taucht das Lemma in medizinischen Artikeln auf (301-mal in 82 Aufsätzen). Die Kookkurrenz *Frage stellen* ist mit 289 Treffern (das entspricht 12,5% aller Verbindungen mit *Frage*) die frequenteste. Mit Abstand am häufigsten wird diese Verbindung in sprachwissenschaftlichen Texten verwendet, nämlich 119-mal in 60 Artikeln. In Beiträgen der Glottodidaktik kommt diese Kombination noch 69-mal (in 49 Artikeln) vor, in denen der Literaturwissenschaft 64-mal (41 Artikel). Nur 26 der 125 medizinischen Aufsätze bedienen sich dieser Formulierungssequenz 37-mal.

Einer *Frage nachgegangen* wird 57-mal; das sind nur noch 2,5%, dennoch spricht die Häufigkeit für eine Kollokation. Auch hier versammeln sich die meisten Treffer in Artikeln der Sprachwissenschaft und der Glottodidaktik (nämlich jeweils 26). Lediglich vier Treffer gibt es in der Literaturwissenschaft und nur einen in der Medizin.

In nur 8 Fällen *drängt sich eine Frage auf* (0,3%). Wie von Steinhoff postuliert, wird den Fragen nicht *gefolgt* und sie *zwingen sich* auch nicht *auf*. 5-mal *ergeben sich* jedoch Fragen, das sind zwar lediglich 0,2% der Treffer; nichtsdestotrotz dürfte eine wissenschaftssprachliche Kollokation vorliegen. Diese Annahme wird auch dadurch gestützt, dass in den Wissenschaftstexten des DWDS¹⁶-Kernkorpus aus den Jahren 1900 bis 1999 die Verbindung immerhin 70-mal¹⁷ vorkommt.

Wird nach den statistisch signifikanten verbalen Kookkurrenz-Partnern¹⁸ für das Lemma *Frage* gesucht, steht auf Platz 1 ebenfalls *stellen*, gefolgt von *beantworten* – dieser Kollokator erzielt 97 Treffer im Korpus (4,2%).

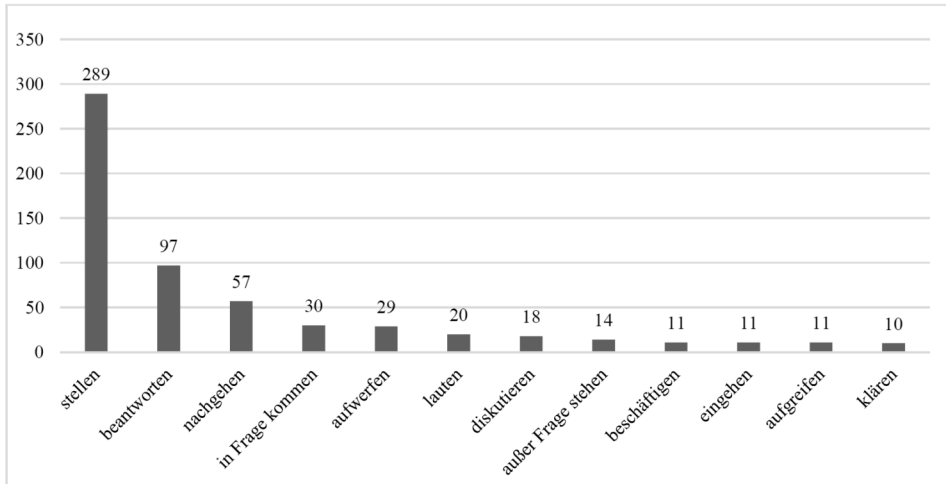
Einen Überblick über verbale Kollokationen, die mindestens 10-mal im Korpus auftreten¹⁹, gibt Abbildung 1:

¹⁶ Digitales Wörterbuch der deutschen Sprache (www.dwds.de/r).

¹⁷ Mit einem rechten und linken Suchabstand zwischen den beiden Lemmata von 15 Token.

¹⁸ Mit dem statistischen Kollokationsmaß der *Mutual Information* und einer Kollokationsweite von jeweils 5 Token auf der rechten und der linken Seite sowie einem statistischen Wert von 5.

¹⁹ Die in einem Korpus mit wissenschaftlichen Texten in hoher Frequenz zu erwartende

Abbildung 1. Verbale Kollokationen zum Lemma *Frage*

Quelle: eigene Forschung.

Als hochfrequentes Lemma erscheint “Frage” oft in ähnlichen oder gleichen Wendungen, die über die reine verbale Kollokation hinausgehen. Beispiele für solche Formulierungsmuster aus dem Korpus, die auch explizit auf das jeweilige Forschungsinteresse des Beitrags Bezug nehmen (vgl. Weder 2015: 208), sind:

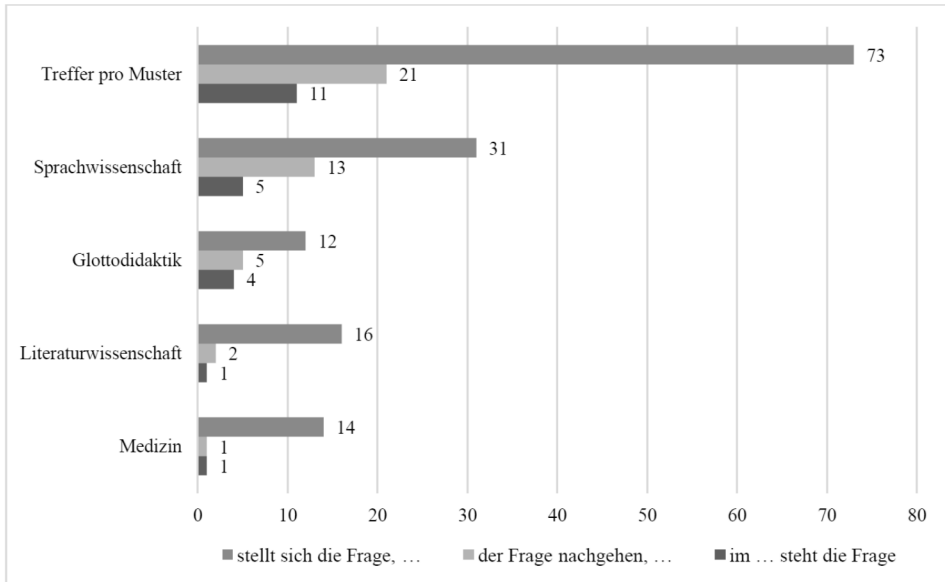
- [es] **stellt sich die Frage**, ob/welche/wie ...
- im Fokus/Mittelpunkt/Zentrum **steht die Frage** ...
- [es] **wird der Frage nachgegangen**, inwieweit/ob/was/warum/welche/wie ...

Hier zeigt sich, dass feste Formulationssequenzen an den rechten und linken Rändern wiederum formelhaft aufgefüllt werden können.

Knapp die Hälfte dieser musterhaften Verwendungen kommt im Unterkorpus mit den sprachwissenschaftlichen Artikeln vor (49 von 105). Abbildung 2 zeigt die Verteilung.

Bei der Suche nach diesen Mustern in den Textsorten Belletristik, Zeitung und Gebrauchsliteratur im DWDS-Kernkorpus des 20. Jahrhunderts (das ca. 92.196.000 Token umfasst, also über 40-mal größer als das vorliegende Wissenschaftssprachkorpus ist) ergab die Sequenz “es wird der Frage nachgegangen” überhaupt keine Übereinstimmung, die Formulierung “im Fokus/Mittelpunkt/Zentrum steht die Frage” wurde nur 3-mal gefun-

Kollokation *eine Frage behandeln* hat lediglich 5 Treffer (davon 3 in literaturwissenschaftlichen Texten und jeweils 1 aus der Sprachwissenschaft und aus der Glottodidaktik).

Abbildung 2. Musterhafte Verwendungen mit *Frage*

Quelle: eigene Forschung.

den und “[es] stellt sich die Frage” mit den oben erwähnten nachfolgenden Lexemen 28-mal. In den Wissenschaftstexten des DWDS-Kernkorpus (ca. 27.780.000 Token) wird 1-mal “der Frage nachgegangen” und 3-mal “steht die Frage im Fokus/Mittelpunkt/Zentrum”. 30-mal “stellt sich die Frage”. Dieser Befund unterstützt die Annahme, dass diese Muster tatsächlich wissenschaftssprachlich sind und in anderen Textsorten selten oder gar nicht auftreten.

4.2. Fallanalyse 2: Paar- und Mehrlingsformeln

Auch Paar- oder Mehrlingsformeln sind typische Phrasemtypen. Sie sind schon lange Forschungsgegenstand der Phraseologie; durch Erkenntnisse aus der Konstruktionsgrammatik rückt die Musterhaftigkeit jedoch besonders in den Blick: Sie werden immer nach dem Muster gebildet, dass “[z]wei Wörter der gleichen Wortart oder auch zweimal dasselbe Wort [...] mit *und*, einer anderen Konjunktion oder einer Präposition zu einer paarigen Form verbunden” (Burger 2015: 55) werden. Die Formelhaftigkeit zeigt sich hier nicht nur in der festen oder einer bevorzugten Reihenfolge (z.B. *klipp und klar* vs. **klar und klipp*), sondern auch in der lexikalischen Festigkeit (z.B. *ganz und gar* vs. **völlig und gar*; vgl. Gawel 2017: 25–43). Ob solche in der Umgangssprache

häufig verwendeten sprachlichen Muster auch in der Wissenschaftssprache Verwendung finden, kann ebenfalls korpuslinguistisch untersucht werden. Wird nicht nach den verschiedenen Arten von Paarformeln unterschieden, sondern lediglich eine Frequenz von mindestens 10 Treffern vorausgesetzt, ergibt sich für die Paarformeln des Korpus das in Tabelle 2 dargestellte Ergebnis (wobei die jeweils meisten Treffer grau hinterlegt sind):

Tabelle 2: Frequente Paarformeln

Paarformel	Gesamt	Glottodidaktik	Literaturwissenschaft	Sprachwissenschaft	Medizin
<i>mehr oder weniger</i>	182	61	40	78	3
<i>nach wie vor</i>	82	28	16	27	11
<i>nach und nach</i>	43	11	19	9	4
<i>mehr oder minder</i>	31	4	16	11	0
<i>mit und ohne</i>	24	7	1	8	8
<i>mehr und mehr</i>	19	6	9	3	1
<i>[im] Großen und Ganzen</i>	17	8	3	5	1
<i>Schritt für Schritt</i>	14	9	4	0	1
<i>hin und her</i>	14	3	10	1	0
<i>Hand in Hand</i>	11	4	5	2	0
<i>ganz und gar</i>	10	1	5	4	0
<i>Theorie und Praxis</i>	10	3	3	2	2
gesamt	457	145	131	150	31

Quelle: eigene Forschung.

Die mit Abstand häufigste Verwendung findet die mit dem disjunktiven Konnektor *oder* verbundene Formel *mehr oder weniger*, die allein schon 40% aller Treffer ausmacht. Insgesamt finden sich vergleichbar viele Ergebnisse in den Artikeln der Sprachwissenschaft, der Glottodidaktik und der Literaturwissenschaft (33, 32 und 29% der Treffer) – in den medizinischen Beiträgen werden die Paarformeln fast gar nicht verwendet (nur 7%).

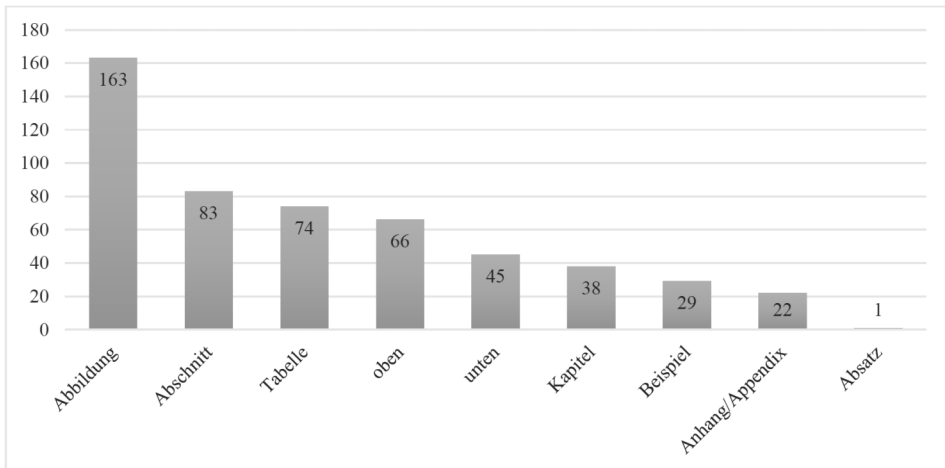
4.3. Fallanalyse 3: Metadiskursive Textroutinen

Als typisch für wissenschaftliche Beiträge gelten auch metadiskursive Textroutinen oder Metakomentierungen. Diese sprachlichen Routinen kommentieren die Struktur und Ziele des Textes und sind häufig musterhaft aufgebaut. Sie sind somit wichtige "Bausteine" der Wissenschaftssprache, die zum einen der Verständnissicherung beim Rezipieren dienen, zum anderen

aber auch den AutorInnen durch ihre textkompositorischen Funktionen bei der Textproduktion helfen (vgl. Olszewska 2013: 80).

Metadiskursive Verfahren wie das intra- und intertextuelle Verweisen treten in der Wissenschaftssprache häufiger auf als im nichtwissenschaftlichen Sprachgebrauch (vgl. Hyland 2005). Die einfachste Art eines Metakommentars ist die Verbindung mit *siehe* oder *vergleiche*. Diese Verben können zwar auch auf Sachverhalte außerhalb des Textes verweisen (z.B. auf Literatur), sie dienen aber ebenso der internen Textorganisation, indem auf oben oder unten Erwähntes oder auf zurückliegende und folgende Kapitel, Abschnitte, Seiten, Tabellen oder Abbildungen verwiesen wird. Im Korpus kommen 707 solcher Verweise vor. 45% davon werden in den sprachwissenschaftlichen Aufsätzen verwendet, 25 bzw. 24% in der Medizin und der Glottodidaktik und lediglich 6% in literaturwissenschaftlichen Beiträgen. Am häufigsten werden vorhergehende oder nachfolgende Abbildungen genannt, gefolgt von Abschnitten und Tabellen, die allgemeineren Verortungen *oben* und *unten* finden sich auf den Plätzen 4 und 5 (siehe Abbildung 3).

Abbildung 3. Intratextuelle Verweise mit *siehe* und *vergleiche*



Quelle: eigene Forschung.

Komplexere Verweise können metakommunikative/textkommentierende Äußerungen wie *“in diesem Abschnitt soll gezeigt werden ...”*, *“das folgende Kapitel geht auf X ein”*, *“im Folgenden geht es um ...”* und andere (mehr oder weniger) feste Wortverbindungen sein.

Textorganisierende Ausdrücke (vgl. Völz 2016: 14) dienen bei wissenschaftlichen Beiträgen oft dem *Initiieren*, *Sequenzieren* und *Schließen* (vgl. Olszewska 2013: 91) des gesamten Textes oder eines Textabschnitts.

Als Beispiel sollen hier sprachliche Muster für das Sequenzieren vorgestellt werden. Danuta Olszewska nennt hierfür die drei relationalen Positionen *zunächst*, *dann* und *schließlich*.²⁰ Als Beispiele für die erste Position in einer Sequenz sind etwa folgende Muster im Korpus zu finden:

- **zunächst** soll[en] [X] beschrieben/betrachtet/diskutiert/überprüft ... werden
- [es] ist **zunächst** zu bemerken/klären/konstatieren/zeigen, dass/was/wie ...
- **zu Beginn** des Abschnitts/Artikels/Aufsatzes/Beitrags/Kapitels ... wird/soll ...
- **am Anfang** der Analyse/des Beitrags/Kapitels/Textes ... steht

Beispielhafte Metatexteme²¹ für die nächste Position der Sequenz sind

- der **nächste** Abschnitt/Punkt/Schritt widmet sich ...
- das **nächste** Beispiel/Kapitel zeigt ...
- wir wollen/ich möchte **nun** auf [X] eingehen
- im **folgenden** Abschnitt/Beispiel/Beleg wird argumentiert/dargelegt/hergeleitet ...

Geschlossen wird eine Sequenz durch Muster wie

- **zusammenfassend** kann bestätigt/festgehalten/gesagt werden, dass ...
- **zusammenfassend** lässt sich festhalten/feststellen/sagen, dass ...
- **abschließend** ist zu betonen/festzustellen/zu konstatieren ...

Auch hier zeigt sich, dass es formelhafte Verwendungen gibt, die entweder identisch oder mit lexikalischen Varianten gebraucht werden. Solche Textschablonen können auch unterschiedlich komplex aufgefüllt werden, sie ändern ihre Struktur dabei aber nicht.

5. Zusammenfassung

Die Beispiele haben gezeigt, dass mit korpuslinguistischer Analyse und den verwendeten Programmen sowohl wissenschaftssprachliche Kollokationen als auch musterhafte Verwendungen im gesamten Korpus der Wissenschaftssprache nachgewiesen und entdeckt werden können.

²⁰ "Während des gesamten Textherstellungsprozesses hat der Autor zu entscheiden, in welcher Reihenfolge er das zu vermittelnde Wissen präsentieren soll. Diese Entscheidungen manifestieren sich häufig an der Textoberfläche in Form von Metatextemen mit zahlreichen ablaufkonstituierenden Ausdrücken, die die Position der einzelnen Wissenskomponenten in der linearen Abfolge bestimmen und ihre Relationierung verdeutlichen" (Olszewska 2013: 93).

²¹ "In der deutschen linguistischen Literatur werden dafür verschiedene Bezeichnungen gebraucht, z.B. *metakommunikative Mittel* (Göpferich 1995), *Textkommentare* (Graefen 1997), *metadiskursive Signale* (Mautner 2011). Aus ökonomischen Gründen verwende ich den Einzelwortterminus *Metatexteme*, der eine Übersetzung der polnischen Bezeichnung *metatekstemy* ist, die Gajewska (2004) in die polnische Linguistik eingeführt hat" (Olszewska 2013: 80).

Die Festigkeit der Wortverbindungen besteht zum einen in der genauen Wiederholung lexikalischer Abfolgen, aber auch in der Struktur von Mustern, die dann mit lexikalischen Varianten gefüllt werden (die wiederum aus einem relativ festen Repertoire bestehen). Auch wenn alle Unterkorpora Treffer aufweisen, sind besonders die sprachwissenschaftlichen Texte von Musterhaftigkeit und formelhaften Verwendungen geprägt.

Literatur

- Ágel, Vilmos. 2004. "Phraseologismus als (valenz)syntaktischer Normalfall". In: Steyer, Kathrin (Hg.). *Wortverbindungen – mehr oder weniger fest*. Berlin / New York: de Gruyter, 65–86.
- Belica, Cyril & Perkuhn, Rainer. 2015. "Feste Wortgruppen / Phraseologie I: Kollokationen und syntagmatische Muster". In: Haß, Ulrike & Storzjohann, Petra (Hgg.): *Handbuch Wort und Wortschatz*. Berlin / New York: de Gruyter, 201–225.
- Brezina, Vaclav / McEnery, Tom & Wattam, Stephen 2015. "Collocations in context: A new perspective on collocation networks." In: *International Journal of Corpus Linguistics* 20 (2), S. 139–173.
- Brommer, Sarah. 2018. *Sprachliche Muster. Eine induktive korpuslinguistische Analyse wissenschaftlicher Texte*. Berlin / Boston: de Gruyter.
- Burger, Harald. 2015. *Phraseologie. Eine Einführung am Beispiel des Deutschen*. 5., neu bearbeitete Auflage. Berlin: Erich Schmidt.
- Burger, Harald / Dobrovól'skij / Kühn, Peter & Norrick, Neal R. 2007. "Phraseologie: Objektbereich, Terminologie und Forschungsschwerpunkte". In: Burger, Harald / Dobrovól'skij, Dmitrij / Kühn, Peter & Norrick, R. Neal (Hgg.): *Phraseologie. Ein internationales Handbuch der zeitgenössischen Forschung* (Handbücher zur Sprach- und Kommunikationswissenschaft 35.1). Berlin / New York: de Gruyter, 1–10.
- Czicza, Dániel / Hennig, Mathilde / Emmerich, Volker & Niemann, Robert. 2012. "Zur Verortung von Texten zwischen den Polen maximaler und minimaler Wissenschaftlichkeit. Ein Operationalisierungsvorschlag". In: *Fachsprache* 1–2, 2–44.
- Deml, Isabell. 2015. *Gebrauchsnormen der Wissenschaftssprache und ihre Entwicklung vom 18. bis zum 21. Jahrhundert*. Dissertation. Universität Regensburg.
- Ehlich, Konrad. 1993. "Deutsch als fremde Wissenschaftssprache". In: *Deutsch als Fremdsprache* 19, 13–42.
- Ehlich, Konrad & Graefen, Gabriele. 2001. "Sprachliches Handeln als Medium diskursiven Denkens. Überlegungen zur sukzursiven Einübung in die deutsche Wissenschaftskommunikation". In: *Jahrbuch Deutsch als Fremdsprache 2001* (27). München: iudicium, 351–378.
- Fandrych, Christian. 2017. "Gesprochene-Sprache-Forschung und Korpuserschließung am Beispiel von GeWiss digital". In: Fandrych, Christian / Meißner, Cordula & Wallner, Franziska (Hgg.): *Gesprochene Wissenschaftssprache – digital. Ver-*

fahren zur Annotation und Analyse mündlicher Korpora. Tübingen: Stauffenburg, S. 13–32.

- Fandrych, Christian & Graefen, Gabriele. 2010. "Wissenschafts- und Studiensprache Deutsch". In: Krumm, Hans-Jürgen / Fandrych, Christian / Hufeisen, Britta & Riemer, Claudia (Hgg.): *Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch* (Handbücher zur Sprach- und Kommunikationswissenschaft 35.1). Berlin / New York: de Gruyter, 509–517.
- Fandrych, Christian / Meißner, Cordula & Slavcheva, Adriana. 2012. "The GeWiss corpus: Comparing spoken academic German, English and Polish". In: Schmidt, Thomas & Wörner, Kai (Hgg.): *Multilingual corpora und multilingual corpus analysis*. Amsterdam: Benjamins, 319–337.
- Farø, Ken. 2015. "Feste Wortgruppen / Phraseologie II: Phraseme". In: Haß, Ulrike & Storjohann, Petra (Hgg.): *Handbuch Wort und Wortschatz*. Berlin / New York: de Gruyter, 226–247.
- Farø, Ken. 2006. *Idiomatizität, Ikonizität, Arbitrarität – Beitrag zu einer funktionalistischen Theorie der Idiomäquivalenz*. Kopenhagen: Museum Tusulanum.
- Feilke, Helmuth. 2012. "Was sind Textroutinen? – Zur Theorie und Methodik des Forschungsfeldes". In: Feilke, Helmuth & Lehnen, Kathrin (Hgg.): *Schreib- und Text-routinen. Theorie, Erwerb und didaktisch-mediale Modellierung*. Frankfurt/Main u.a.: Peter Lang, 1–31.
- Feilke, Helmuth & Steinhoff, Torsten. 2003. "Zur Modellierung der Entwicklung wissenschaftlicher Schreibfähigkeiten". In: Ehlich, Konrad & Steets, Angelika (Hgg.): *Wissenschaftlich schreiben – lehren und lernen*. Berlin / New York: de Gruyter, 112–128.
- Gaweł, Agnieszka. 2017. "Zur Ikonizität deutscher Zwillingformeln". In: *Linguistik online* 81 (2), S. 25–43.
- Graefen, Gabriele. 2001. "Einführung in den Gebrauch der Wissenschaftssprache". In: *Materialien Deutsch als Fremdsprache* 58 ("Wie schwer ist die deutsche Sprache wirklich?"), 191–210.
- Handwerker, Britta. 2010. "Phraseologismen und Kollokationen". In: Krumm, Hans-Jürgen / Fandrych, Christian / Hufeisen, Britta & Riemer, Claudia (Hgg.): *Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch* (Handbücher zur Sprach- und Kommunikationswissenschaft 35.1). Berlin/New York: de Gruyter, 246–255.
- Hyland, Ken. 2005. *Metadiscourse. Exploring Interacting in Writing*. London/New York: Bloomsbury Publishing.
- Kretzenbacher, Heinz. 1994. "Wie durchsichtig ist die Sprache der Wissenschaften?". In: Kretzenbacher, Heinz & Weinrich, Harald (Hgg.): *Linguistik der Wissenschaftssprache* (= Forschungsbericht der Akademie der Wissenschaften zu Berlin Nr. 10). Berlin / New York: Walter de Gruyter.
- Olszewska, Danuta. 2013. "Auf der Suche nach einem *tertium comparationis*: Wissenschaftliche Texte im deutsch-polnischen Vergleich". In: *Studia Germanica Gedan-sia* 29, 79–99.

- Pilz, Hans-Dieter. 1981. *Phraseologie, Redensartenforschung*. Stuttgart: Metzler.
- von Polenz, Peter. 1981. "Über die Jargonisierung von Wissenschaftssprache und wider die Deagentivierung". In: Bungarten, Theo (Hg.): *Wissenschaftssprache. Beiträge zur Methodologie, theoretischen Fundierung und Deskription*. München: Fink, 85–110.
- Schmid, Helmut. 1995. Improvements in Part-of-Speech Tagging with an Application to German. *Proceedings of the ACL SIGDAT-Workshop*. Dublin, Ireland.
- Steinhoff, Torsten. 2007. *Wissenschaftliche Textkompetenz: Sprachgebrauch und Schreibentwicklung in wissenschaftlichen Texten von Studenten und Experten*. Tübingen: Niemeyer.
- Steyer, Kathrin (Hg.). 2004. *Wortverbindungen – mehr oder weniger fest* (Jahrbuch des Instituts für Deutsche Sprache 2003). Berlin et al.: de Gruyter.
- Steyer, Kathrin. 2000. "Usuelle Wortverbindungen des Deutschen. Linguistisches Konzept und lexikographische Möglichkeiten". In: *Deutsche Sprache* 28, 101–125.
- Targońska, Joanna. 2014. "Kollokationen – ein vernachlässigtes Gebiet der DaF-Didaktik?". In: *Linguistik online* 68 (6), 127–149.
- Thielmann, Winfried / Redder, Angelika & Heller, Dorothea. 2014. "Akademische Wissensvermittlung im Vergleich". In: Redder, Angelika / Heller, Dorothea & Thielmann, Winfried (Hgg.): *Eristische Strukturen in Vorlesungen und Seminaren deutscher und italienischer Universitäten*. Heidelberg: Synchron, S. 7–17.
- Völz, Irina. 2016. *Lexikalische Textgliederung beim wissenschaftlichen Schreiben in der Fremdsprache Deutsch. Eine empirische Untersuchung zum Erwerb und Gebrauch textorganisierender Ausdrücke durch internationale DaF-Studierende*. Kassel: university press.
- Wallner, Franziska. 2014. *Figurative Verben in der allgemeinen Wissenschaftssprache des Deutschen. Eine Korpusstudie*. Tübingen: Stauffenburg.
- Weder, Mirjam. 2015. "Fragen, zeigen, argumentieren? Metadiskursive Textroutinen zur Etablierung von Forschungspraktiken in wissenschaftlichen Einleitungen". In: Schmidlin, Regula / Behrens, Heike & Bickel, Hans (Hgg.): *Sprachgebrauch und Sprachbewusstsein. Implikationen für die Sprachtheorie*. Berlin: de Gruyter, 199–220.
- Weinrich, Harald. 1989. "Formen der Wissenschaftssprache". In: *Jahrbuch 1988 der Akademie der Wissenschaften zu Berlin*. Berlin et al.: de Gruyter, 119–158.

Phrasemes and Word Combinations in the German Scientific Language

Summary

Scientific language is characterized, among other things, by the use of specific phrases and exemplary formulations. Current collocations and text routines can be determined and proven using a corpus consisting of 500 scientific articles from the years 2010 to 2017 for the areas of German Studies, Glottodidactics and Medicine.