

DE GRUYTER

PRODUCTIVE PATTERNS IN PHRASEOLOGY AND CONSTRUCTION GRAMMAR

A MULTILINGUAL APPROACH

Edited by Carmen Mellado Blanco

FORMELHAFTE SPRACHE
FORMULAIC LANGUAGE

EBSCO Publishing : eBook Collection (EBSCOhost) - printed on 2/10/2023 2:32 AM
via
AN: 3290777 ; Carmen Mellado Blanco.; Productive Patterns in Phraseology and
Construction Grammar : A Multilingual Approach
Account: ns338141

Productive Patterns in Phraseology and Construction Grammar

Formelhafte Sprache

Formulaic Language

Edited by
Natalia Filatkina, Kathrin Steyer and Sören Stumpf

Advisory Board

Harald Burger (Zurich), Joan L. Bybee (New Mexico),
Dmitrij Dobrovolskij (Moscow), Stephan Elspaß (Salzburg),
Christiane Fellbaum (Princeton), Raymond Gibbs (Santa Cruz),
Annelies Häcki Buhofer (Basel), Claudine Moulin (Trier),
Jan-Ola Östman (Helsinki), Stephan Stein (Trier),
Martin Wengeler (Trier), Alison Wray (Cardiff)

Volume 4

Productive Patterns in Phraseology and Construction Grammar

A Multilingual Approach

Edited by
Carmen Mellado Blanco

DE GRUYTER

The publication of the current volume was generously supported by:

- Spanish Ministry of Science and Innovation. Research Project *Construction Grammar and Phraseology: German and Spanish Constructional Idioms in contrast through Corpora* (RETOS; Ref. FFI2019-108783RB-100)
- LITLINAL Research Group (Ref. GI-1954). University of Santiago de Compostela.

ISBN 978-3-11-051849-8

e-ISBN (PDF) 978-3-11-052056-9

e-ISBN (EPUB) 978-3-11-051860-3

ISSN 2625-1086

Library of Congress Control Number: 2021939867

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <http://dnb.dnb.de>.

© 2022 Walter de Gruyter GmbH, Berlin/Boston

Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Acknowledgements

This anthology brings together contributions to the *Construction Grammar* section of the *EUROPHRAS International Conference 2019 Productive Patterns in Phraseology*, held at the University of Santiago de Compostela (24–25 January 2019), which are published here in revised form. The volume has been expanded to include papers from colleagues whose work usefully complements the overall theme of the collection. The articles reflect on the phenomenon of idiomatic patterns and constructions in different languages and from the theoretical, methodological and empirical approaches of Pattern Grammar, Phraseology, and particularly Construction Grammar.

I would like to thank the EUROPHRAS Society, the LITLINAL Research Group of the USC, and the Spanish Ministry of Science and Innovation (projects FFI2013-45769-P and FFI2019-108783RB-100) for their generous financial support for both the conference and this publication. My sincere thanks also to Natalia Filatkina, Kathrin Steyer and Sören Stumpf for including the volume in the series *Formelhafte Sprache / Formulaic Language* at De Gruyter.

I would also like to express my deepest gratitude to all the authors whose valuable work is represented here, as well as to the anonymous reviewers for their constructive collaboration. Special thanks also to my fabulous research group FRASESPAL for the organization of the Conference, to Álex Alonso Santos for his very careful, and sometimes laborious, preparation of the artwork, and to Elisabeth Stanciu for all her editorial and technical support. Without their commendable efforts, the volume would not have been possible.

Carmen Mellado Blanco

<https://doi.org/10.1515/9783110520569-202>

Content

Acknowledgements — V

Carmen Mellado Blanco

Phraseology, patterns and Construction Grammar — 1

An introduction

Part I: Productive patterns: monolingual

Pedro Ivorra Ordines

Comparative constructional idioms — 29

A corpus-based study of the creativity of the [*más feo que X*] construction

Marie Kopřivová

Between phraseology and conversational routines — 53

Using spoken Czech corpora

Fabio Mollica, Sören Stumpf

Families of constructions in German — 79

A corpus-based study of constructional phrasemes with the pattern [X_{NP} attribute]

Elizaveta Onufrieva

Constructional phrasemes in Modern Greek: structure and meaning — 107

Valentina Piunno

Coordinated constructional intensifiers: patterns, function and productivity — 133

Kathrin Steyer

Preposition-noun combinations of TIME in German — 165

A pattern-based approach to minimal phraseological units

Part II: Productive patterns: bi- and multilingual

Dmitrij Dobrovol'skij, Ludmila Pöppel

Russian constructions with *nu i* in parallel corpora — 191

Laura Giacomini

Phraseology in technical texts — 215

A frame-based approach to multiword term analysis and extraction

Martina Häcker

Cross-language transfer of formulae — 235

The case of English letters

Zuriñe Sanz-Villar

German-into-Basque translation of verbal patterns — 265

Analysing trainee translators' outputs

Lena Stutz, Rita Finkbeiner

***Veni, vidi, veggie* — 287**

A contrastive corpus linguistic analysis of the phraseological construction *Veni, vidi, X* and its German equivalent *X kam, sah und Y*

Index — 315

Carmen Mellado Blanco

Phraseology, patterns and Construction Grammar

An introduction

1 About the concept of “pattern” and its relation to “construction” and “phraseologism”

The significance of phraseology and formulaic language becomes clear if we consider that a large percentage of the sentences that we utter in our day-to-day communication do not follow the *open choice principle* (Sinclair 1991), but rather the *idiom principle*¹, that is, they are composed of more or less fixed strings of words, stored and retrieved whole from memory at the time of use (Wray 2002: 9). According to corpus studies by Erman and Warren (2000), around 55% of the sentences we create, both written and oral ones, are prefabricated.² Bearing this quantitative assessment in mind, the idiomatic phenomenon constitutes, in numerical terms, a central axis in the description of languages, and

1 Sinclair’s idiom principle states that “[a] language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments” (Sinclair 1991: 110).

2 For Erman and Warren (2000: 31), “[a] prefab is a combination of at least two words favoured by native speakers in preference to an alternative combination which could have been equivalent had there been no conventionalization”. The term “prefab” corresponds to what authors from different linguistic traditions and approaches have called “lexicalized sentence stem” (Pawley and Syder 1983), “construction” (Goldberg 1995, 2006, 2019), “recurrent word-combination” (Altenberg 1998), “phraseological unit” (Gläser 1998, Fiedler 2007), “phraseme” (Mel’čuk 1998), “set phrase” (Mel’čuk 1998, Colson 2008), “lexical item” (Sinclair 1998), “idiom” and “fixed phrase” (Moon 1998), “pattern” (Hunston & Francis 2000, Gries 2008, Ebeling and Ebeling 2013), “formulaic language” (Wray 2002, Buerki 2020), “lexical bundle” (Biber and Barbieri 2007), “n-gram” (Stubbs 2007) and “phraseologism” (Gries 2008).

Note: Study within the framework of the research project *Construction Grammar and Phraseology: German and Spanish Constructional Idioms in contrast through Corpora* (FFI2019-108783RB-I00), promoted by the Spanish Ministry of Science and Innovation.

Carmen Mellado Blanco, University of Santiago de Compostela, c.mellado@usc.es

<https://doi.org/10.1515/9783110520569-001>

cannot be considered as something marginal or anecdotal. Another notable aspect of the phraseology of languages is that recurrent patterns can be identified within it. In the words of Sinclair (1991: 534), “the realm of idiomaticity in a language includes a great deal that is productive, highly structured, and worthy of serious grammatical investigation”. Thus, within the semantic or syntactic irregularity that phraseological units can exhibit, it is possible to recognise tendencies and structures that repeat; that is, to identify patterns. These formational patterns affect both the internal structure of phraseologisms and the syntagmatic profiles in which they are integrated. On the same lines, Hanks et al. (2018: 116) provide the Corpus Pattern Analysis (CPA) in order to describe “the rigidity or diversity not only of the MWE [multiword expression] itself, but of its context on either side” and the consistency of patterning to the left and right of the multiword expression.³

The concept of “pattern” involves different aspects in linguistic research. Common to all of these is that templates are identified by corpus evidence.⁴ Among the most influential theories of patterns are the following:

(1) For Ebeling and Ebeling (2013: 50), within the framework of Corpus Linguistics, a pattern is a phraseologism, that is “a recurrent sequence of orthographic words that function as a semantic unit” (2013: 50), e.g. *back and forth*, *in so far as*⁵. In this context, these authors claim that there is a difference between n-grams and patterns, in that “many of the n-grams extracted will not qualify as patterns” (Ebeling and Ebeling 2013: 69).⁶ For his part, Gries (2008: 6)

3 Sinclair’s extended-unit-of-meaning model (EVofM), in which the node constitutes the core of the EVofM, can also be considered as a pattern of this type (e.g. *to/with naked eye*) (Sinclair 1996: 83–89). In line with Stubbs (2007: 163), an EVofM “typically consist of a partly fixed core plus other variable items occurring to the left and right of the core”. This unit has its own collocational and colligational environment, semantic preference and semantic prosody. Collocational frameworks (e.g. *a + X + of*), by contrast, do not arise from autosemantic words, but rather have functional words as nodes (e.g. the indefinite article *a* or the preposition *of*) (Renouf and Sinclair, 1991).

4 The larger the corpus, the greater the probability that the results of the study reflect the reality of the language; as Sinclair (1991) stated three decades ago: “bigger is always better”.

5 Also for Schmid (2014), from the perspective of Cognitive Linguistics, patterns are identified with phraseological units. For Schmid, lexico-grammatical patterns are recurrent sequences of lexical and grammatical elements which serve an identifiable function and include all kinds of phraseologisms of distinct degrees of fixation and compositionality.

6 These authors base their work on the concept of the “n-gram” of Lindquist and Levin (2008: 144) who use the term “n-gram” for the recurring strings (with or without linguistic integrity) in the corpora and “pattern” or “construction” for meaningful, linguistically structured recurring sequences of words.

defines a pattern as “[t]he co-occurrence of a form or a lemma of a lexical item and one or more additional elements of various kinds which functions as one semantic unit in a clause or sentence and whose frequency of co-occurrence is larger than expected on the basis of chance”.

(2) Unlike Ebeling and Ebeling (2013), who focus especially on lexical patterns, Hunston and Francis’s (2000) Pattern Grammar centres on grammar patterns⁷, of the type <*explain* + *why*-clause> or <*v* + *over* + noun group / *wh*-clause>, that refer to the grammatical environment following a particular word in a particular sense. In general, then, it concerns complementation patterns, that is, the specification of items that follow the key word. For these authors, a pattern exists if a combination of words “occurs relatively frequently, if it is dependent on a particular word choice, and if there is a clear meaning associated with it” (Hunston and Francis 2000: 37). It is important to note that grammar patterns are often used with a restricted lexis. This means that the majority of the words of a pattern occur with high frequency, whereas a small number of words appear only sporadically.

For Hunston and Francis (2000), the concept of “pattern” is conceived of as having two sides:

- a. On the one hand, each word has its patterns (“a word and its patterns”) and “[t]he patterns of a word can be defined as all the words and structures which are regularly associated with the word and which contribute to its meaning” (Hunston and Francis 2000: 37). To illustrate this, the patterns of *explain* would be: *explain* + *about*; *explain* + noun group; *explain* + *why*-clause; *explain* + *how*-clause; *explain* + *what*-clause, among others. Word meanings are thus arrived at by abstraction from their lexical-grammatical co-occurrences. Interestingly, the meanings of words depend on the configuration of each pattern, in line with Firth’s (1957: 11) well-known remark: “you shall know a word by the company it keeps”.
- b. On the other hand, a pattern can be associated with words coming from a specific semantic class (“a pattern and its words”). This means that, for example, the verbs of the pattern <*v* + *over* + noun group> belong to the semantic class: *bicker*, *disagree*, *fight*, *quarrel*, and *wrangle* (Hunston and Francis 2000: 44).

⁷ The grammar patterns of Hunston and Francis (2000) coincide to a great extent with the valency patterns of Schmid (2014), that is, they are “complementation patterns associated with verbs and other valency carriers” (cf. Herbst 2010: 191–192), e.g. *believe* + N + Adj; *believe* + *that*-clause; *believe* + N (+ *to* INF).

(3) Within the CPA of Hanks (2004, 2013), applied to the *Pattern Dictionary of English*, each pattern is linked to one or more implicatures or meanings. Through CPA, each verb has a systematic analysis of the patterns of meaning and use. Also, patterns may combine various kinds of categories such as semantic types (Human, Wind, Vapour, Dust), grammatical categories (Adverbial of direction) and lexical items. In contrast to Hunston and Francis (2000), the pattern as conceived of by Hanks (2013) has a clear semantic dimension reflected in the semantic types that include the formulation and the meaning associated with the pattern.⁸ Along the same lines as Hanks (2013), lexical patterns for Steyer (2019: 131) “consist of fixed lexical components (‘lexical anchors’) as well as slots” (e.g. *There’s no X, only Y*, with the pragmatic function ‘relativization’ or ‘encouragement’). The slot fillers are linked to specific communicative situations and can have different degrees of typicality and display similar semantic and/or pragmatic characteristics.

(4) The concepts of “pattern” as seen in recent work by Hanks et al. and Hunston (cf. Hanks et al. 2018, Hunston 2019) are both very close to that of “construction” in Construction Grammar(s) (CxGs). In all these models, each formal structure is intrinsically linked to a meaning.⁹ However, methodologically they are significantly different grammatical models, since while Pattern Grammar and CPA owe their existence to Corpus Linguistics, CxGs arose from Cognitive Linguistics (cf. Croft 2013). Despite these differences, in recent years there have been attempts to bridge the gap between patterns and constructions, as witnessed by one of Hunston’s (2019) most recent contributions, in which she proposes an alignment between aspects of Pattern Grammar (Francis, 1993, Hunston and Francis 2000) and CxGs (Goldberg 2006). Hunston (2019: 324) argues that the term “construction” “be used to refer to a sub-set of instances of a grammar pattern, that subset identified by the occurrence of a limited set of node words”.

A reconciliation between patterns and constructions was already noticed in Hunston and Francis (2000) through what they called “patterns with a meaning”, that is, patterns whose meaning did not depend only on the slot fillers, but

8 As an example of this, the sense of *blow your nose* is stored in the pattern “[Human] blow {nose}” while in the sense of “the wind blows” is represented by the pattern “[Wind | Vapour | Dust] blow [No object] [Adverbial of direction]” (Hanks et al. 2018: 97).

9 Stubbs (2009: 27) summarises thus the parallels between the two schools: “Despite their different origins, Pattern Grammar and Construction Grammar both conclude that the primary unit of meaning is not the individual word, but a phrasal construction which consists of lexical, grammatical, semantic and pragmatic relations, and that the traditional part of [the] speech system is suspect: syntactic relations can be stated only within larger constructions”.

also on the structure, which is itself a carrier of meaning.¹⁰ According to these authors, in such cases, “it is possible to go further and say that it is the pattern itself, not the words with which it occurs, that has the meaning” (Hunston and Francis 2000: 100). We consider this point of view to be crucial because it implies a clear connection between Pattern Grammar and CxGs in terms of the close link between form and meaning of patterns and constructions, conceived of in both linguistic currents as two sides of the same coin.

CxGs emerge in the final years of the twentieth century from the premises of Cognitive Linguistics (Fillmore et al. 1988; Lakoff 1987) and currently embrace a healthy number of usage-based approaches”, such as Berkeley CxG, Sign Based CxG, Cognitive CxG, Goldbergian CxG, Radical CxG, Embodied CxG and Fluid CxG. The constructions, defined as “conventionalized clusters of features (syntactic, prosodic, pragmatic, semantic, textual, etc.) that recur as further indivisible associations between form and meaning (meaning is broadly understood)” (Fried 2015: 974) conform, when taken together, to the construction of a language. As in Functional Grammar (Halliday 2013), CxGs argue for the interdependence of vocabulary and syntax and rejects the idea of a sharp dichotomy between lexical items and grammatical rules. Instead, constructions are linguistic patterns at every level of generality (schematicity) and lexical specificity on the lexicon-grammar continuum. They include items from individual words, morphemes and idioms,¹¹ to partially lexically filled constructions and fully abstract grammatical constructions (Goldberg 2006: 5).

(5) Worth mentioning here is the concept of “formulaic pattern” coined by Filatkina (2018), very close to that of construction in that it embraces various levels of schematisation (e.g. *brush one’s teeth, one X after another*). Although Filatkina’s notion does not include morphemes as an object of study, unlike the proposals of CxG, yet does include formulaic patterns, extending the scope to formulaic texts and discourse. A central and innovative aspect of formulaic patterns is their syntactic and semantic flexibility, given that they “can show varying degrees of conventionalisation (ranging from high to low) with regard

10 This can be observed for example in the resultative pattern <V way prep/adv> (V which are concerned with talking) which has the meaning ‘someone uses clever, devious, or forceful language to achieve a goal, usually extricating themselves from a difficult situation, or getting into a desirable situation’. Hunston and Francis (2000: 101) illustrate this pattern with the sentence *He was more able than anyone else to argue this way out of tough situations*.

11 Idioms in the sense of substantive idioms or fully filled and fixed constructions. We ought not to forget the interest that CxGs have had in idiomatic structures from its first inception, seeking an explanation for irregular syntactic structures and semantic idiomaticity that constitute a major part of a language.

to their form, meaning and functions, but have a stable underlying syntactic and / or cognitive structure” (Filatkina 2018: 20). From this interesting point of view, modifications (e.g. **Diamanten vor die Säue werfen*, instead of *Perlen vor die Säue werfen*), defined as “an irregular, intentional and conscious intervention of a speaker into the form and / or meaning of a pattern directed at the violation of the existing norms” (2018: 26–27) are considered as a natural manifestation and a motor of diachronic change.

2 Constructional idioms, productivity and the continuum between true constructions and patterns of coining

For Phraseology, constructions with different levels of schematicity, that is, partially lexically filled constructions, have thus far attracted most of research interest from a constructional perspective (Dobrovol’skij 2011, Mollica and Schafroth 2018, Ziem 2018, López Meirama and Iglesias Iglesias in press, Mellado Blanco 2020 ed., Mellado Blanco, Mollica and Schafroth in press). Until very recently, these units had been seen as a marginal phenomenon within the phraseological research. They show an atypical nature as phraseologisms because of the fact that some of their lexical content is variable, they are productive, and in many cases also semantically transparent. Partially filled constructions or constructional idioms¹² are characterized by their pragmatic functions (Cappelle 2017), as we can confirm in the well-known examples of the *what’s X doing Y* construction (*What’s that fly doing in my soup?*), which expresses an illocutionary act of request or demand for an explanation (Kay and Fillmore 1999: 4, Schafroth 2020: 190–192) and in the Incredulity Response Construction *Him be a doctor!?* (Lambrecht 1990: 225), which expresses ‘incredulity’ (Kay 2004: 677).

According to Taylor (2016: 11), constructional idioms are “patterns (of varying degrees of productivity and schematicity) for the formation of expressions,

¹² Schmid (2014) uses the terminology of “partially filled periphery constructions” as a subgroup of lexico-grammatical patterns. Other terms commonly used for this type of construction are “lexically open idioms”, “formal idioms” (Fillmore, Kay and O’Connor 1988: 505), “linguistic schemata” (Van Lancker et al. 2012: 87), “schematic idioms” (Croft and Cruse 2004: 234), “phraseotemplates” (Schafroth 2015: 327), “Phrasem-Konstruktionen” (Dobrovol’skij 2011), “constructional idioms” (Booij 2002, Taylor, 2016: 11) and “phrasal patterns” (Michaelis, 2019: 194).

but whose syntactic, semantic, pragmatic, and even phonological properties cannot be derived from general principles, whether universal or language-specific". In phraseological research, Moon (1998: 95) refers to constructions of this type as "syntactic frames", which, she argues, constitute the most extreme case of variation within phraseologisms. In these she recognises restrictions on the kinds of lexis realizing the frame, which, as we noted above, has also been claimed to be an essential feature of patterns (Hunston and Francis 2000) and of constructional idioms (Taylor 2016). Some examples of syntactic frames provided by Moon (1998) are:

- (1) INTERROGATIVE + *or* + INTERROGATIVE (to indicate certainty):
Am I right or am I right?
- (2) *If*-CLAUSE, CLAUSE (to indicate acceptance of a situation):
If it rains, it rains.
- (3) INTERROGATIVE (as a response to an unnecessary question):
Is the Pope Catholic? Do bears shit in the woods? Do ducks swim? Does Dolly Parton sleep on her back?
- (4) QUANTIFIER + NOUN GROUP 1 + *short of* + NOUN GROUP 2 (to indicate inadequacy, especially mental inadequacy):
two sandwiches short of a picnic, a few bricks shy of a full load, a few pickles short of a jar, one shingle shy of a roof, a flying buttress short of a cathedral, one hot pepper short of an enchilada.

Considering that they are semi-schematic in nature, constructional idioms have different degrees of productivity, and thus their location within the lexis-grammar continuum is variable. For Barðdal (2008: 24), syntactic productivity, which itself relates to type and token frequency¹³, revolves around three axes, namely those of generality, regularity and extensibility. Productivity understood as regularity (in the sense of regular; rule-based; easily combinable; transparent or compositional; lexically and semantically modifiable) is closely

¹³ Type frequency refers to the "total number of types which can instantiate a construction", whereas token frequency means the "total occurrences of either one or all the types of a construction in a text or corpus" (Barðdal 2008: 27).

linked to syntactic creativity on Chomskyan lines¹⁴, that is to say, to the generation of sentences never encountered before in the sense of generating new tokens of already existing types (Barðdal2008: 29). In addition, productivity as regularity is related to the “true constructions” of Fillmore (1997, see also Kay 2013), as the only ones able to generate endless sentences without restrictions (e.g. *the X-er the Y-er* construction, Fillmore, Kay and O’Connor 1988: 507). This type of productivity as full schematicity without restrictions is also proposed by Goldberg (1995: Ch. 5) in her concept of construction.

On the contrary, Barðdal (2008: 29) sees productivity as extensibility as a phenomenon of analogy (in the sense of developing new functions, occurring with new/novel/nonce items; attracting already existing items) and, unlike productivity as regularity, is based on the extensibility of a pattern to new types, not new tokens. It is useful here to bear in mind that analogical creations are more limited according to the number of generated expressions.¹⁵ Bybee (1995: 426) also adopts the view of productivity as extensibility, calculated by the number of different types in the slot fillers, being this productivity index very significant.

The most recent studies based on true constructions and patterns of coining by Fillmore (1997) and Kay (2013) tend to play down any clear distinction between the two types of constructions and postulate a continuum between them (Traugott and Trousdale 2013, Desagulier 2015, 2016, Sampson 2016, Bergs 2019).¹⁶ In this line of reasoning, Desagulier (2016: 27) claims that even “the most schematic construction will not generate an infinity of instances” and that intermediate productivity levels must be recognised. This occurs in the *A as NP* construction, classified by Kay (2013) as pattern of coining, yet with an undeniable degree of productivity in the creation of subschemas and analogical exten-

14 Chomsky (2003: 402) defines language as “a process of free creation; its laws and principles are fixed, but the manner in which the principles of generation are used is free and infinitely varied. Even the interpretation and use of words involves a process of free creation”.

15 For Sampson (2016: 19), this type of productivity is related to the so-called “E(xtending)-creativity” and refers to “activities which characteristically produce examples that enlarge our understanding of the range of possible products of the activity”. By contrast, productivity as regularity is related to “F(ixed)-creativity”.

16 In the words of Traugott and Trousdale (2013: 185): “[a] problem with this approach is that one cannot know without experimentation whether speakers do in fact learn each string individually and whether they are likely or not to build new expressions analogized to these patterns”.

sions.¹⁷ In this context, empirical corpus analysis allows us to confirm the almost limitless creativity of speakers in generating types of apparently unproductive constructions (Mellado Blanco 2020, in press). As noted by Desagulier (2015: 123–124), “la frontière entre «constructions» (générales et productives) et «schémas dérivés» était poreuse et que l’on ne pouvait pas la cerner uniquement par des jugements introspectifs”.

3 Emergence of semi-schematic constructions from phraseologisms

Analogical extensions indeed play a crucial role in the creation of new patterns, especially through the process of lexical substitution. This is a widespread phenomenon among fully filled constructions, e.g. idioms, quotations and proverbs, which by recurrent variation of their components and a high type or token frequency of these changes can result in semi-schematic constructions. This interesting phenomenon of constructionalisation (Traugott 2007) shows several strands within phraseological units:

(1) Through the modification of proverbs, anti-proverbs can be generated¹⁸, defined by Mieder (2004: 28) as “parodied, twisted, or fractured proverbs that reveal humorous or satirical speech play with traditional proverbial wisdom”. Some formal mechanisms to create anti-proverbs are: association, change of homonyms, combination, permutation, abridgement, substitution, supplementation and syntactic change. The semantic effects of anti-proverbs can be summarised as: mitigation, apology, conservation, break of metaphor or neogenesis (Litovkina and Mieder 2006: 17–26). An example of an anti-proverb is: *Where there’s a will, there’s a lawsuit* (change of homonym *will*, on the basis on *Where there’s a will, there’s a way*). When the number of antiproverb relative to a proverb is very high, the conditions are favourable for the emergence of a “structural formula” (Mieder 2004: 85) or a “proverb pattern” (Steyer 2019). This happens with “*A (One) picture is worth a thousand Ys*” in which it is the usual noun *words* that has been replaced.” In this and similar cases, creative modifications

¹⁷ On the productivity of intensifying comparative constructions in Spanish, see Corpus Pastor (2021), Mellado Blanco (2015, 2019), and Ivorra Ordines and Mellado Blanco (2021).

¹⁸ The term “anti-proverb” was first coined in 1982 by Wolfgang Mieder. It became more established with the publication of *Twisted Wisdom: Modern Anti-Proverbs* by Litovkina and Mieder (2006). See also Mieder (in press).

can be reduced to structural formulas. On certain occasions, although modifications to a proverb do exist, we cannot speak of a pattern because they occur only occasionally. In accordance with Stumpf's (2016: 337) proposal, we distinguish between "modifiers" ("Modifikationsmuster"), if the substitution of the proverb components is occasional (under 50% of all the instances of the construction), and true phraseological schemata ("Modellbildungen") with open slots to be filled with a variety of types on a regular basis. Although this is an arbitrary division, Stumpf's (2016) study serves as a starting point for other proposals and as recognition of the continuum between the phenomenon of creative modification and the emergence of new semi-schematic constructions.

(2) A very similar phenomenon to structural formula, but more centred on clichés in journalese, is that of snowclones¹⁹, defined by Pullum (2004) as "some-assembly-required adaptable cliché frames for lazy journalists", of the type *The only good X is a dead X* (schematization of *The only good Indian is a dead Indian*). Another example is *to X or not to X* (from Shakespeare's line *to be or not to be*), with instances such as *to play or not to play*, *to eat or not to eat meat*, in order to be witty or funny (Bergs 2019: 177). Although the term "snowclone" emerged outside of CxGs, this type of pattern can be identified as a semi-schematic construction, and thus it has not gone unnoticed by certain construction grammarians. This is the case with Traugott and Trousdale (2013: 183), who describe the phenomenon of snowclones as follows: "in a snowclone a fixed specific expression becomes less fixed by virtue of introducing a variable (a formal change), while the original meaning of the micro-construction generalizes". Diachronically, they "arise from lexical constructionalization of a schema after a number of constructional changes" (Traugott and Trousdale 2013: 184). Such is the case with the biblical quotation *My cup runneth over* (Psalm 23: 5), with the original meaning 'I have a surfeit of something', which through a process of constructionalization comes to mean 'X is beyond capacity', 'X is too much' in the snowclone *My X runneth over*.

In the case of both structural formulae and snowclones, creative formations by analogical extensions often follow patterns of creativity. This is what Philip (2008: 106) argues that non-canonical forms are unpredictable, but they seem to follow systematic tendencies in their variability. Such systematicity usually responds to some kind of motivation with respect to the lexical item that is substituted or modified, either phonologically, semantically and/or morphologically (Rasulic 2010). Moreover, it is noteworthy to note that the new constructional

¹⁹ At <https://snowclones.org/> (accessed 25 April 2021) an extensive collection of snowclones can be found.

meaning builds on the canonical meaning of the phraseological unit, giving rise to a polyphonic semantic effect. Philip comments on this phenomenon as follows:

[W]hatever element is substituted, its meaning is always read in relation to the canonical phrase. The new element forces the reader to analyse the phrase both compositionally and non-compositionally, and the overall meaning is a combination of the old phrase and the new, and not a new phrase in its own right²⁰.

Philip (2008: 104)

(3) A type of pattern which is different from snowclones and structural formulae is that underlying proverbs (Peukes 1977). As Mieder (2004: 6–7) notes, “the thousands of proverbs of any language can be reduced to certain structures or patterns”, such as “Better X than Y,” “Like X, like Y,” “If X, then Y,” calling to mind such well-known proverbs as *Better poor with honour than rich with shame*, *Like father, like son*, and *If at first you don’t succeed, try, try again*. In contrast to snowclones and structural formulae, these formative models are not “Modelle der Synthese” or “productive patterns”, but rather “Modelle der Analyse” (Fleischer 1997: 192–194 apud Apresjan 1972) and can only be ascertained a posteriori due to the similarity of certain phraseological terms. This conception of patterns as formation models broadly coincides with Fillmore’s (1997) patterns of coining. Moreover, like these, patterns analysis can sometimes serve as a model for generating new units by analogy, thus achieving a certain productivity. This is the case of *Better red than dead*, created from the analysis pattern “Better X than Y”, and its opposite *Better dead than red* from the time of the Cold War with its anticommunism propaganda (Mieder 2004: 7).

In accordance with the above, in the case of structural formulae, either as modificatory patterns or as true phraseological schemata, the challenge for linguists is to identify what mechanisms act as triggers for the creative units, that is, what motivates the non-canonical instances of phraseological units.²¹ For constructional idioms, of the type of the *WXDY* (*what’s X doing Y*) construction and the Incredulity Response Construction, the priority would consist in

²⁰ In CxGs, this phenomenon is known as “coercion” (s. Michaelis 2004, González-García 2020).

²¹ Attempts have been made to answer the question of creativity in phraseological units within Blending Theory (Omazić 2005, Langlotz 2006, Rasulic 2010). On these lines, Hoffmann (2019) goes a step further and advocates for the incorporation of Conceptual Blending in CxGs in order to explain the phenomenon of creativity in all types of constructions.

determining the lexical and morphological restrictions of the slot fillers (cf. Mellado Blanco 2015, 2019, 2020).

CxGs thus constitute an ideal framework for studying phraseological variation along the lexis-grammar continuum using the resources offered by Corpus Linguistics. As noted above, the phraseological group of syntactic frames (Moon 1998) would be considered from a constructionist perspective as partially lexically filled constructions (Goldberg 2006: 5) or constructional idioms (Taylor 2016), with different degrees of schematicity and productivity. But it is not only in the field of syntactic frames that CxGs promise to bear fruit. I advocate that lexically filled constructions and lexically substantive constructions (e.g. idioms, proverbs) also offer an interesting field of study in relation to non-canonical individual variants, what in phraseology have traditionally been called “modifications” (Burger 2015: 159–178), for example *the scream of consciousness* (The Guardian, 4 May, 1996, p. 27) instead of the canonical phraseologism *the stream of consciousness* (Naciscione 2010: 168). Particularly in the case of transparent idioms, it is not uncommon that the distinction between literal and phraseological meaning loses its contours, as Lindquist & Levin point out:

Phrases are often manipulated in various ways, so that they occur in non-canonical forms and in word play. The use of word play shows that the borderline between literal and non-literal meanings is fuzzy, and that both a literal and a nonliteral meaning can be available to speakers simultaneously, although at any given moment one is usually more salient than the other.

Lindquist & Levin (2008: 143)

Frequent modifications of substantive idioms can function as a real engine of change, as a trigger for the emergence of new semi-schematic constructions²² (cf. Mellado Blanco 2020, Mellado Blanco in press). In this way, with a constructionist focus the clear division tends to disappear, not only between canonical and non-canonical use of idioms, but also between systemic variations of phraseological units (e.g. *get/have a raw deal*; *burn your boats/bridges*; *in (full) bloom*; *to wipe the slate clean/ the slate is clean*) and creative modifications of

²² Let us recall that the study of idioms is not currently one of the priorities of CxGs, which focus more on semi-schematic constructions and on constructions of a higher level of abstraction and schematicity (e.g. grammatical constructions) (cf. Buerki 2018). From my point of view, CxGs should reinforce the study of phraseologisms, not as constructions which are static in terms of their form and content, but as flexible units with great creative potential.

phraseological units “made by the individual speaker or writer who intends a particular stylistic effect” (Gläser 1998: 130).²³

The great advantage of CxGs is that variation is approached holistically in terms of productivity, analogy, regularity and schematicity, and the creative play of idioms, quotations, proverbs etc. is not interpreted as an irregular phenomenon, but simply as infrequent instantiations of constructions (Mellado Blanco 2020). This explains why no line can be drawn between phraseology and creativity²⁴, which leads some authors to consider fixed expressions as constructional frames (Rasulic 2010). Far from being conceived as a destabilising force, creativity is understood in CxGs as a manifestation of the playful nature of human beings²⁵ and as a mechanism of linguistic change by adaptation to the context of use.

4 About this volume

This collection brings together contributions to the section *Construction Grammar* of the EUROPHRAS conference 2019 *Productive Patterns in Phraseology*, held at the University of Santiago de Compostela (24–25 January 2019), and are published here in revised form after a peer review process. The volume has been expanded to include papers from colleagues whose work usefully complements the overall theme of the collection.

The book contributes to the discussion of patterns in both (corpus-based) monolingual and contrastive perspectives. Against the background of the aforementioned aspects of Pattern Grammar, Construction Grammar and Phra-

23 As I have shown in a study on the series of synonymous variants *me importa un bledo/pimiento/rábano/cuerno*, etc. (Mellado Blanco 2020), it is not unusual for the speaker to make creative modifications to a series of synonymous variants, and thus the division between systemic and occasional variants would disappear. On the theoretical and empirical impossibility of realising “a strict demarcation line between systematic variation and idiomatic word-play”, see Langlotz (2006), Dobrovol’skij (2008: 304), Petrova (2011: 201–202), Pfeiffer (2018), Filatkina et al. (2020: 5).

24 The distinction between formula and creativity as continuum, not as dichotomy is also argued for within Pattern Grammar (Hunston and Francis 2000). Weinert (1995: 198) talks on these same lines of language as a formulaic-creative continuum.

25 Early proponents of this conception include Gibbs et al. (1989), who already emphasised not only the lexical flexibility of phraseologisms, but also the interpretative ability of speakers to decode phraseological sentences that deviate from the canonical form, either in form or content (cf. Moon 1998).

seology, the volume takes up current trends in pattern research and at the same time addresses research desiderata. Most of the contributions focus on theoretical, methodological and/or empirical questions concerning semi-schematic constructions and prefabs:

- The studies of Kopřivová, Steyer and Sanz-Villar rely on the concept of lexical and/or grammatical pattern developed in Corpus Linguistics (Hunston and Francis 2000, Ebeling and Ebeling 2013). On similar lines, Giacomini's corpus-based contribution also draws on the concept of pattern, in this case from Fillmore's Frame Semantics (2008).
- The contributions by Ivorra Ordines, Mollica and Stumpf, Onufrieva, Piunno, and Dobrovol'skij and Pöppel have CxG as a backdrop, so that patterns are understood as pairings of form and meaning. These authors focus on semi-schematic constructions.
- Häcker's diachronic study of Middle English formulae is mainly based on Wray's (2002) idea of formulaic language, within the framework of Phraseology and prefabs.

The papers in this volume, as well as showing the relationships between Construction Grammar, Pattern Grammar and Phraseology, highlight some of the challenges of research on constructional idioms and prefabs in general. Some of these may be considered as new lines of research:

- If, as seems natural, we consider phraseological units and free chains to be gradual manifestations of the lexis-grammar continuum (Dobrovol'skij and Pöppel; Kopřivová), what would be the guidelines for delimiting the object of study in phraseology?
- In the case of semi-schematic constructions, how can the instances licensed by a construction be recognised among all the n-grams resulting from a corpus search? This difficulty is even greater when the fixed constituents of the semi-schematic construction are functional words (Onufrieva).
- In constructional idioms, how is it possible to accurately describe the prototypical illocutionary and pragmatic functions if this information is linked to the particular contexts of use that license each construction (Kopřivová, Piunno, Stutz and Finkbeiner, Steyer, Onufrieva, Ivorra Ordines, Mollica and Stumpf).
- What degree of exhaustivity is required in the description of the vertical and horizontal links of constructional idioms within the construction of a language? (Mollica and Stumpf).

As is evident from most of the studies in this volume, both monolingual and contrastive, semi-schematic constructions can have various contextual meanings, as is also the case with words. The problem is that since not all constituents are lexically fixed, the semantic potential of such constructions is highly complex. On the basis of this potential polysemy, the common definition of construction as pairing of form and meaning needs to be revised, since it is common for a construction to have several meanings, depending on:

- the semantic class and the (morpho-)syntactic type of fillers,
- the specific filler slot, given that some constructs can be constructionalised and experience a reanalysis,
- the extraconstructional meaning of the slot fillers when these are polysemic,
- the collocational partners and the syntagmatic profile of each construction.

The phenomenon of productivity is present in all these contributions to a greater or lesser extent, especially in relation to partially lexically filled patterns. Such patterns, with a fixed and a variable part in the form of lexical slots, are described in the lexis-grammar continuum (Onufrieva), ranging from conversational routines of oral language and written language (Kopřivová; Dobrovol'skij and Pöppel), to intensifying comparisons (Ivorra Ordines), evaluative constructional phrasemes (Mollica and Stumpf), coordinated constructional intensifiers (Piunno), lexical patterns that express 'time' (Steyer), verb preposition collocations (Sanz-Villar), and patterns based on modified quotations (Stutz and Finkbeiner). Productivity can also be observed in terminological variants of the multiword terms (Giacomini), as well as in certain formulas of Middle English that allow different kinds of variation (Häcker).

From a contrastive point of view, the studies in the volume demonstrate the usefulness of parallel corpora for the identification of equivalent units and constructions in language B, which has a very positive impact on the advancement of lexicography and foreign language learning (Dobrovol'skij and Pöppel; Giacomini, Sanz-Zuriñe). Finally, the analysis of calques and borrowings of constructions are of interest in describing the degree of influence of a foreign language at a given time (Häcker), as well as in comparing the productivity between native and borrowed constructions (Stutz and Finkbeiner).

The volume is divided into two sections, with six and five papers respectively. From a monolingual perspective, the first part includes contributions on patterns in Spanish (Ivorra Ordines), Czech (Kopřivová), German (Mollica and Stumpf, and Steyer), Modern Greek (Onufrieva), and Italian (Piunno). The second block focuses on the interlingual analysis of patterns in Russian, Swedish, English and German (Dobrovol'skij and Pöppel), German, English and Italian

(Giacomini), Middle English and its influence on medieval French (Häcker), German and Basque (Sanz-Villar), German and a construction borrowed from Latin (Stutz and Finkbeiner).

We will now comment on each of the contributions separately.

The study “Comparative constructional idioms. A corpus-based study of the creativity of the [*más feo que X*] construction”, by **Pedro Ivorra Ordines**, focuses on the semi-schematic construction [*más feo que X*]. Through a rigorous analysis of the *Sketch Engine* corpus *esTenTen18*, Ivorra Ordines considers the productivity of the construction based on the type and token of slot *X* and asks to what extent the speaker licenses the construction through lexicalised slot fillers (such as in the micro-constructions *más que feo que Picio*, *más feo que un dolor*, for example), or whether the speaker does this in a creative way. This study highlights the importance of “E-productivity” or productivity as extensibility (Barđal 2008: 29), i.e. the speaker’s ability to create new types by means of analogical extensions. The author points at an interesting continuum between lexicalised constructs and creative instances, with a clear predominance of the latter (63.63%). Moreover, in view of the results obtained from the corpora, this contribution reveals the shortcomings of many dictionaries in reflecting frequently used expressions.

In the paper “Between phraseology and conversational routines. Using spoken Czech corpora”, **Marie Kopřivová** extracts possible patterns from the conversational routines involving the diagram *to je* (‘that is’) in informal dialogues of native Czech speakers in the spoken language (oral corpus *ORTOFON*). Particularly, she looks at the trigrams (1) *to je* + Adj with values of agreement/backchannel: negative agreement, positive surprise or satisfaction, doubts; and (2) *to je* + N, with values of agreement, disagreement, negative agreement, surprise, and doubts. She also analyses tetragrams with the preposition *na* (‘in’), *za* (‘for’) and *z* (‘from’), from which discourse markers stand out numerically. Kopřivová’s study of spontaneous speech patterns is in line with Ebeling and Ebeling’s (2013) notion of pattern and is of great interest, not only for highlighting the continuum between free sequences and phraseologisms, but above all for establishing correspondences between concrete structural patterns and the illocutionary acts that they prototypically express.

The contribution of **Fabio Mollica** and **Sören Stumpf**, “Families of constructions in German. A corpus-based study of constructional phrasemes with the pattern [*X_{NP}* attribute]”, focuses on constructional phrasemes in which a lexically open noun phrase is specified by a lexical fixed postponed attribute.

The authors come to the conclusion that intensifying constructional phrasemes [X_{NP} *pur*], [X_{NP} *par excellence*], [X_{NP} *wie er/sie/es im Buche steht*], [X_{NP}, *schlechthin*], [X_{NP} *durch und durch*], [X_{NP} *erster Güte*], [X_{NP} *in Reinkultur*] and [X_{NP} *sondergleichen*], despite certain formal-structural differences with regard to the realization of the postponed attribute, can be considered as integrated parts (mesoconstructions) of a family of constructions, in which the more common abstract construction [X_{NP} attribute] (macroconstruction) is a frequent and productive pattern of German. This study proves that constructional idioms are not isolated units of the constructicon, but are integral parts of a highly structured, hierarchical network consisting of constructions with different degrees of schematicity (mesoconstructions > macroconstructions > microconstructions > constructs) (Traugott 2007).

The main aim of **Elizaveta Onufrieva's** pioneering study, “Constructional phrasemes in Modern Greek: structure and meaning”, is to classify Modern Greek constructional idioms according to the type of words that function as fixed components. In this way, and based on the *Sketch Engine* corpus *elTen-Ten14*, she distinguishes between constructional phrasemes with functional words, with semantically bleached content words, and with content words as fixed components. All these three types of constructional phrasemes share the property of having a semantically restricted variable component and irregular syntax (to different extents). The first type of construction is, on the syntax-lexicon continuum, closest to schematic/abstract constructions and its idiomatic nature is mostly based on their syntactic non-compositionality. In contrast to these, constructional phrasemes that have content words in their fixed part behave as idioms and are closer to the lexical end of the lexis-grammar continuum. Through a process of abstraction, the author is able to identify five semantic groups in the constructional idioms of Modern Greek: evaluation, high degree of quality, multitude, permanence of a quality, and negation. Onufrieva reaches the interesting conclusion with her analysis that these semantic groups are not necessarily linked to a specific structural type.

The study “Coordinated constructional intensifiers: patterns, function and productivity”, by **Valentina Piunno** considers those constructional intensifiers in Italian which have the coordinated structure [Lexeme_A + Conjunction + Lexeme_B]. Paying attention to morpho-syntactic, semantic and functional parameters, such as: i) their syntactic pattern, ii) the semantics of the lexical words in slots, iii) semantic analysability, iv) the degree of lexicalisation, and v) their functional value, Piunno classifies coordinated constructional intensifiers in Italian into different classes which are situated on a *continuum* of lexicalisation. At the poles of this continuum there are substantive constructions (e.g. *bello e*

buono) and partially filled constructions ($[[lungo\ e + X_{ADJECTIVE}] = \text{'extremely } X_{ADJECTIVE}\text{'}$). Their status is variable, in that some constructs deriving from partially filled schemes may lexicalise and the related constructional scheme may lose its productivity. One major contribution of this analysis is that it establishes for coordinated constructional intensifiers a continuum of constructionalisation depending on the following criteria: i) productivity, ii) semantic predictability, iii) lexical variability, and iv) schematicity; being these criteria the result of a corpus-based approach.

The article by **Kathrin Steyer**, “Preposition-noun combinations of TIME in German. A pattern-based approach to minimal phraseological units”, seeks to show that lexical units (mono-lexemes and multi-word expressions) as well as lexical patterns are entrenched and interconnected in many ways and on different levels of abstraction. Following Hunston and Francis (2000), Steyer defines lexical patterns as conventionalised, partially filled lexical schemes that are frozen by its frequent use, e.g. $\langle \textit{für} N_{\text{CALENDARISCH_METRISCH}} \rangle$ ($\textit{for } N_{\text{CALENDARISCH_METRISCH}}$), with items such as *für Jahre*, *für Wochen*, etc. In her corpus study, she shows that in many patterns of TIME the expression of the duration of time has a subjective nature, in such a way that these units tend to have a modal adverbial use with connotative meanings and are highly restricted by the contexts in which they are embedded. In Steyer’s theory, lexical patterns exhibit interesting parallels with partially lexically filled constructions in CxGs, although the author argues that her conception focuses much more on structures and interrelations of lexical items, and seeks to contribute to a usage-based theory of lexis.

Dmtrij Dobrovol’skij and **Ludmila Pöppel** pursue two goals in their contribution “Russian constructions with *nu i* in parallel corpora”. Firstly, they conduct a monolingual corpus-based study of the phrasal template *nu i X* (literally: well and X) and highlight two syntactic groups (*nu i N* and *nu i +* an invariable part of speech), each with clear semantic preferences: either surprise or negative connotations for *nu i N*, and indifference for *nu i +* an invariable part of speech. The constructions found in the corpus are situated on a continuum of idiomaticity that goes from idioms to free word combinations. Secondly, Dobrovol’skij and Pöppel seek to ascertain the degree of language specificity of the Russian constructions under analysis, this measured by the number of equivalent units and their degree of scattering in language B. For their study, they use the parallel corpus *OPUS2* (Russian, English, German and Swedish) and the *Russian National Corpus (RNC)*. The investigation concludes with a remarkable observation: in the field of bilingual lexicography, teaching of foreign languages and translation studies, what really matters in relation to constructional idioms is the degree of language specificity, rather than idiomaticity.

Laura Giacomini focuses her study, “Phraseology in technical texts. A frame-based approach to multiword term analysis and extraction”, on multiword terms from the semantic field of ‘thermal insulation’. Because of the limited terminological standardisation of specialised language, there is a strong proliferation of terminological variants, especially for multiword terms. Term variation is not arbitrary, but responds to some specific patterns conceived as frames, that is, as cognitive models which cover particular word meanings and argument structures. In the case of technical artefacts such as thermal insulation products, in line with the framed-based terminology used, we can distinguish frames such as FUNCTIONALITY and frame elements like MATERIAL, MATERIAL CLASS, DELIVERY FORM, TARGET, PACKAGING, TOOL, AND APPLICATION TECHNIQUE, in the sense of Frame Semantics. The main objective of this study is to implement a methodology for a frame-based analysis and extraction of multiword terms and variants from a corpus of technical texts in German, with observations on data in English and Italian, and with a view to their subsequent presentation in a lexicographic resource supporting text production.

On a diachronic level, the purpose of the paper “Cross-language transfer of formulae. The case of English letters”, by **Martina Häcker**, is to investigate the conventions of Middle English letters, documented in a corpus consisting in the correspondence of three families between 1476-1482. Starting out from Wray’s (2002) notion of “formulaic language”, the paper focuses on the use of formulae and the input of different traditions and cultures of Middle English letter writing. In the period 1100–1400, a change was seen in the language of letter writing from Old English to a variety of French, which has clear repercussions for the formulae of letters written in English throughout the 15th century. Drawing on numerous examples from her own corpus, the author identifies and comments on French calques in regularly frequent formulae at the beginning of letters, focusing on formulae of salutation, recommendation, welfare, health, acknowledgement, and astonishment. While many of these formulae are characterised by being totally rigid, others can be considered as semi-fixed, with open slots for address terms and epithets appropriate for the addressee. In other cases, such as in recommendation formulae, the author recognises interesting hybrid formations resulting from the combination of two formulae.

Zuriñe Sanz-Villar, in the study “German-into-Basque translation of verbal patterns. Analysing trainee translators’ outputs”, addresses the translations of verbal patterns from German into Basque in a small learner translation corpus. The investigation particularly aims at analysing verb preposition collocations in German, such as *halten für*, *pochen auf*, *zählen zu*, *reagieren auf*, *anpassen an*,

ankommen auf, which offer a variety of outputs depending on how they have been combined with other words or how the slots have been filled. The author opts for multi-word verbs because they represent a challenge for non-native users of English due to their semantic non-compositionality and their often polysemous meaning. Also, phraseological units of this type tend to remain outside contrastive analyses. Within the 12 analysed verbal multiword units, Sanz-Villar observes different translations tendencies depending on the patterns' characteristics, the type of tools used by the students in the translation or the influence of Spanish as a language in contact with Basque.

The final study in this volume, by **Lena Stutz** and **Rita Finkbeiner**, is “*Veni, vidi, veggio*. A contrastive corpus linguistic analysis of the phraseological construction *Veni, vidi, X* and its German equivalent *X kam, sah und Y*”. The paper offers a contrastive corpus-based study of the semi-schematic construction [*veni, vidi, X*], developed through the recurrent creative modification of the Latin quote *Veni, vidi, vinci*, and its German equivalent [*X kam, sah und Y*] in present-day German. Analysis of slot fillers using the tool *Lexpan* indicates that there are clear differences in the productive use of the two patterns, which is explained by specific structural, phonological and semantic properties of each construction. According to Stumpf's (2016) distinction between modificatory patterns and true phraseological constructions, the German construction belongs to the second group, while the Latin one can be classified as modificatory pattern taking into account the strong phonological constraints limiting the instantiation of the slot *X*. Contributions of this type are particularly useful for the precise location of the semi-schematic construction within the continuum between occasional modification and systematic schematicity.

References

- Altenberg, Bengt. 1998. On the Phraseology of Spoken English: The Evidence of Recurrent Word-Combinations. In Anthony Paul Cowie (ed.), *Phraseology: Theory, Analysis, and Applications*, 101–122. Oxford: Oxford University Press.
- Barðdal, Jóhanna. 2008. *Productivity. Evidence from Case and Argument Structure in Icelandic*. Amsterdam & Philadelphia: John Benjamins.
- Bergs, Alexander. 2019. What, If Anything, Is Linguistic Creativity? *Gestalt Theorie* 41 (2). 173–184.
- Biber, Douglas & Federica Barbieri. 2007. Lexical bundles in university spoken and written registers. *English for Specific Purposes* 26. 263–286.
- Buerki, Andreas. 2018. Formulaic sequences: a drop in the ocean of constructions or something more significant? In Ian MacKenzie & Martin A. Kayman (eds.), *Formulaicity and Creativity in Language and Literature*, 15–34. London: Routledge.

- Buerki, Andreas. 2020. *Formulaic language and linguistic change. A data-led approach*. Cambridge: Cambridge University Press.
- Booij, Geert. 2002. Constructional Idioms, Morphology and the Dutch Lexicon. *Journal of Germanic Linguistics* 14. 301–329.
- Bybee, Joan L. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10 (5). 425–455.
- Burger, Harald. 2015. *Phraseologie. Eine Einführung am Beispiel des Deutschen*. 5., neu bearbeitete Aufl. Berlin: Erich Schmidt.
- Cappelle, Bert. 2017. What's Pragmatics Doing outside Constructions? In Ilse Depraetere & Ralph Salkie (eds.), *Semantics and Pragmatics: Drawing a Line*, 115–151. Heidelberg: Springer.
- Chomsky, Noam. 2003. *For Reasons Of State*. New York: New Press.
- Colson, Jean-Pierre. 2008. Cross-linguistic phraseological studies: An overview. In Sylviane Granger & Fanny Meunier (eds.), *Phraseology. An interdisciplinary perspective*, 191–206. Amsterdam & Philadelphia: John Benjamins.
- Corpas Pastor, Gloria. 2021. Constructional idioms of 'insanity' in English and Spanish: A corpus-based study. *Lingua* 254 (3). <https://wlv.openrepository.com/handle/2436/623936>
- Croft, William. 2013. Radical Construction Grammar. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*, 211–232. Oxford: Oxford University Press.
- Croft, William & Allan Cruse. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Desagulier, Guillaume. 2015. Le statut de la fréquence dans les grammaires de constructions: *simple comme bonjour?* *Langages* 197 (1). 99–128.
- Desagulier, Guillaume. 2016. A lesson from associative learning: Asymmetry and productivity in multiple-slot constructions. *Corpus Linguistics and Linguistics Theory* 12 (2). 173–219.
- Dobrovolskij, Dmitrij. 2008. Idiom-Modifikationen aus kognitiver Perspektive. In Heidrun Kämper & Ludwig M. Eichinger (eds.), *Sprache – Kognition – Kultur. Sprache zwischen mentaler Struktur und kultureller Prägung*, 302–322. Berlin & New York: de Gruyter.
- Dobrovolskij, Dmitrij. 2011. Phraseologie und Konstruktionsgrammatik. In Alexander Lasch & Alexander Ziem (eds.), *Konstruktionsgrammatik III. Aktuelle Fragen und Lösungsansätze*, 111–130. Tübingen: Stauffenburg.
- Ebeling, Jarle & Signe Oksefjell Ebeling. 2013. *Patterns in Contrast*. Amsterdam & Philadelphia: John Benjamins.
- Erman, Britt & Beatrice Warren. 2000. The idiom principle and the open choice principle. *Text* 20 (1). 29–62.
- Fiedler, Sabine. 2007. *English Phraseology. A Coursebook*. Tübingen: Gunter Narr Verlag.
- Filatkina, Natalia. 2018. Expanding the lexicon through formulaic patterns The emergence of formulaicity in language history and modern language use. In Sabine Arndt-Lappe, Angelika Braun, Claudine Moulin & Esme Winter-Froemel (eds.), *Expanding the Lexicon. Linguistic Innovation, Morphological Productivity and Ludicity*, 15–42. Berlin & Boston: De Gruyter Mouton.
- Filatkina, Natalia, Sören Stumpf & Christian Pfeiffer. 2020. Introduction: Formulaic Language and New Data. In Elisabeth Piirainen, Natalia Filatkina, Sören Stumpf & Christian Pfeiffer (eds.), *Formulaic Language and New Data*, 1–16. Berlin & Boston: De Gruyter Mouton.

- Fillmore, Charles J. 1997. *Construction Grammar Lecture Notes*. <http://www1.icsi.berkeley.edu/~kay/bcg/lec02.html> (accessed 25 January 2021).
- Fillmore, Charles J., Paul Kay & Mary Catherine O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of *Let Alone*. *Language* 64. 501–538.
- Firth, John Rupert. 1957. A synopsis of linguistic theory 1930–1955. In John Rupert Firth (ed.), *Studies in Linguistic Analysis*, 1–32. Oxford: Philological Society.
- Francis, Gill. 1993. A Corpus-Driven Approach to Grammar: Principles, Methods and Examples. In Mona Baker, Gill Francis & Elena Tognini-Bonelli (eds.), *Text and Technology: In Honour of John Sinclair*, 137–156. Amsterdam: John Benjamins.
- Fried, Mirjam. 2015. Construction Grammar. In Artemis Alexiadou & Tibor Kiss (eds.), *Syntax – Theory and Analysis. An International Handbook*, 974–1003. Berlin: Mouton de Gruyter.
- Gibbs, Raymond W., Nandini P. Nayak, John L. Bolton & Melissa E. Keppel. 1989. Speakers' assumptions about the lexical flexibility of idioms. *Memory & Cognition* 17 (1). 58–68.
- Gläser, Rosemarie. 1998. The Stylistic Potential of Phraseological Units in the Light of Genre Analysis. In Anthony Paul Cowie (ed.), *Phraseology: Theory, Analysis, and Applications*, 125–143. Oxford: Oxford University Press.
- Goldberg, Adele E. 1995. *Constructions. A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldberg, Adele E. 2006. *Constructions at Work: the nature of generalization in language*. Oxford: Oxford University Press.
- Goldberg, Adele E. 2019. *Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions*. Princeton, NJ: Princeton University Press.
- González-García, Francisco. 2020. Metonymy meets coercion. The case of the intensification of nouns in attributive and predicative constructions in Spanish. In Annalisa Baicchi (ed.), *Figurative Meaning Construction in Thought and Language*, 152–184. Amsterdam & Philadelphia: John Benjamins.
- Gries, Stefan. 2008. Phraseology and linguistic theory: A brief survey. In Sylviane Granger & Fanny Meunier (eds.), *Phraseology: An interdisciplinary perspective*, 3–25. Amsterdam & Philadelphia: John Benjamins.
- Halliday, Michael A. K. 2013. *Halliday's Introduction to Functional Grammar*. 4th ed. London: Routledge.
- Hanks, Patrick. 2004. Corpus Pattern Analysis. In Geoffrey Williams & Sandra Vessier (eds.), *Proceedings of the XI EURALEX*. Vol. 1, 87–97. Lorient: Université de Bretagne-Sud.
- Hanks, Patrick. 2013. *Lexical Analysis. Norms and Exploitations*. MIT Press: Cambridge.
- Hanks, Patrick, Ismail El Maarouf & Michael Oakes. 2018. Flexibility of multiword expressions and Corpus Pattern Analysis. In Manfred Sailer & Stella Markantonatou (eds.), *Multiword expressions: Insights from a multi-lingual perspective*, 93–119. Berlin: Language Science Press.
- Herbst, Thomas. 2010. *English Linguistics: A Coursebook for Students of English*. Berlin: De Gruyter Mouton.
- Hoffmann, Thomas. 2019. Language and Creativity: A Construction Grammar approach to linguistic creativity. *Linguistics Vanguard* 5 (1). 1–8.
- Hunston, Susan & Gill Francis. 2000. *Pattern Grammar. A corpus-driven approach to the lexical grammar of English*. Amsterdam & Philadelphia: John Benjamins.
- Hunston, Susan. 2019. Patterns, constructions, and applied linguistics. *International Journal of Corpus Linguistics* 24 (3). 324–353.

- Ivorra Ordines, Pedro & Carmen Mellado Blanco. 2021. *Más tontos que el novio de la Chelo*. La intensificación de la estulticia en foros y chats por medio de comparaciones creativas: una aproximación desde la Gramática de Construcciones. *Estudios Románicos* 30. 39–58. <https://revistas.um.es/estudiosromanicos/article/view/471241/308431> (accessed 25 July 2021).
- Kay, Paul. 2004. Pragmatic aspects of grammatical constructions. In Larry Horn and Gregory Ward (eds.), *Handbook of Pragmatics*, 675–700. Oxford: Blackwell.
- Kay, Paul. 2013. The limits of (Construction) Grammar. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*, 32–48. Oxford University Press.
- Kay, Paul & Charles J. Fillmore. 1999. Grammatical Constructions and Linguistic Generalizations: The *What's X Doing Y?* Construction. *Language* 75 (1). 1–33.
- Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: CSLI.
- Lambrecht, Knud 1990. *What, me worry?* Mad magazine sentences revisited. *Proceedings of the 16th Annual Meeting of the Berkeley Linguistics Society*. 215–28.
- Langlotz, Andreas. 2006. *Idiomatic creativity. A cognitive-linguistic model of idiom-representation and idiom-variation in English*. Amsterdam & Philadelphia: John Benjamins.
- Lindquist, Hans & Magnus Levin. 2008. *Foot and Mouth*: The phrasal patterns of frequent nouns. In Sylviane Granger, Fanny Meunier (eds.), *Phraseology: An interdisciplinary perspective*, 143–158. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Litovkina, Anna Tóthné & Wolfgang Mieder. 2006. *Old proverbs never die, they just diversify: a collection of anti-proverbs*. Burlington: University of Vermont and Veszprém: Pannonian University of Veszprém.
- López Meirama, Belén & Nely M. Iglesias Iglesias. In press. The construction [*a todo* NOUN_{sing}]. In Evelyn Wiesinger & Inga Hennecke (eds.), *Constructions in Spanish*. Amsterdam & Philadelphia: John Benjamins.
- Mel'čuk, Igor A. 1998. Collocations and lexical functions. In Anthony Paul Cowie (ed.), *Phraseology: Theory, analysis and applications*, 23–53. Oxford: Oxford University Press.
- Mellado Blanco, Carmen. 2015. Antiphrasis-Based Comparative Constructional Idioms in Spanish. *Phraseology, Phraseodidactics and Construction Grammar(s)*. Special Issue *Journal of Social Sciences*, vol. 11 (3). 111–127. <http://www.nih.thescipub.com/abstract/jssp.2015.11.127> (accessed 25 January 2021).
- Mellado Blanco, Carmen. 2019. Phrasem-Konstruktionen kontrastiv Deutsch-Spanisch anhand von ironischen Vergleichen. *Yearbook of Phraseology* 10. 65–88.
- Mellado Blanco, Carmen. 2020. *(No) me importa un comino* y sus variantes diatópicas. Estudio de corpus desde la Gramática de Construcciones. *ELUA Anexo VII*. 89–111.
- Mellado Blanco, Carmen (ed.) 2020. *Nuevas aportaciones de la Gramática de Construcciones a los estudios de fraseología en las lenguas románicas*. *Romanica Olomucensia* 32 (1). Special Issue.
- Mellado Blanco, Carmen. In press. From idioms to semi-schematic constructions and vice versa: the case of [*a un paso de X*]. In Evelyn Wiesinger & Inga Hennecke (eds.), *Constructions in Spanish*. Amsterdam & Philadelphia: John Benjamins.
- Mellado Blanco, Carmen, Fabio Mollica & Elmar Schafroth (eds.) (in press). *Konstruktionen zwischen Lexikon und Grammatik: Phrasem-Konstruktionen monolingual, zweisprachig, multilingual*. Berlin: Mouton de Gruyter.

- Michaelis Laura. 2004. Type shifting in Construction Grammar: an integrated approach to aspectual coercion. *Cognitive Linguistics* 15 (1). 1–67.
- Michaelis, Laura. 2019. Constructions are patterns and so are fixed expressions. In Beatrix Busse & Ruth Möhlig-Falke (eds.), *Patterns in Language and Linguistics*, 193–220. Berlin: Mouton de Gruyter.
- Mieder, Wolfgang. 2004. *Proverbs: A Handbook*. Greenwood Press.
- Mieder, Wolfgang. In press. *There is no such thing as a free lunch*. New structures for modern proverbs. In Rui J.B. Soares & Outi Lauhakangas (eds.), *Proceedings of the Fifteenth Interdisciplinary Colloquium on Proverbs, November 2021, at Tavira, Portugal*. Tavira: Tipografia Tavirense.
- Mollica, Fabio & Elmar Schafroth. 2018. Der Ausdruck der Intensivierung in komparativen Phrasem-Konstruktionen im Deutschen und im Italienischen: eine konstruktionsgrammatische Untersuchung. In Kathrin Steyer (ed.), *Sprachliche Verfestigung. Chunks, Muster, Phrasem-Konstruktionen*, 103–136. Tübingen: Narr Francke Attempto Verlag.
- Moon, Rosamund. 1998. Frequencies and Forms of Phrasal Lexemes in English. In Anthony Paul Cowie (ed.), *Phraseology: Theory, Analysis, and Applications*, 79–100. Oxford: Oxford University Press.
- Naciscione, Anita. 2010. *Stylistic Use of Phraseological Units in Discourse*. Amsterdam & Philadelphia: John Benjamins.
- Omazić, Marija. 2005. Cognitive Linguistic Theories in Phraseology. *Jezikoslovlje* 6 (1). 37–56.
- Pawley, Andrew & Frances Hodgetts Syder. 1983. Two Puzzles for Linguistic and native-like influence. In Jack Richards & Richard Schmidt (eds.), *Language and Communication*, 191–226. London: Longman.
- Peukes, Gerhard. 1977. *Untersuchungen zum Sprichwort im Deutschen*. Berlin: Erich Schmidt.
- Pfeiffer, Christian. 2018. Zur Identifikation modifizierter Phraseme in Texten: ein Vorschlag für die analytische Praxis. In Sören Stumpf & Natalia Filatkina (eds.), *Formelhafte Sprache in Text und Diskurs*, 49–84. Berlin & Boston: De Gruyter.
- Philip, Gil. 2008. Reassessing the Canon: “Fixed” phrases in general reference. In Sylviane Granger & Fanny Meunier (eds), *Phraseology. An interdisciplinary perspective*, 95–108. Amsterdam: John Benjamins.
- Pullum, Geoffrey K. 2004. Snowclones: lexicographical dating to the second. *Language Log*. http://itre.cis.upenn.edu/~myl/languageelog/archives/2004_01.html. (accessed 25 January 2021).
- Rasulic, Katerina. 2010. *Long Time, No Buzz: Fixed Expressions as Constructions Frames*. *CogniTExtes* 5. <http://journals.openedition.org/cognitextes/356> (accessed 25 January 2021).
- Renouf, Antoinette J. & John M. Sinclair. 1991. Collocational frameworks in English. In K. Ajimer and B. Altenberg (Eds.), *English corpus linguistics. Studies in honour of Jan Svartvik*, 128–143. Harlow: Longman.
- Sampson, G. 2016. Two ideas of creativity. In Martin Hinton (ed.), *Evidence. Experiment and argument in linguistics and philosophy of language*, 15–26. Bern & Berlin: Peter Lang.
- Schafroth, Elmar. 2015. Italian Phrasemes as Constructions: How to Understand and Use Them. *Phraseology, Phraseodidactics and Construction Grammar(s)*. Special Issue *Journal of Social Sciences* 11 (3). 317–337.
- Schafroth, Elmar. 2020. Fraseologismi a schema fisso – basi teoriche e confronto linguistico. In Carmen Mellado Blanco (ed.), *Nuevas aportaciones de la Gramática de Construcciones a los*

- estudios de fraseología en las lenguas románicas*. Special Issue *Romanica Olomucensia* 32 (1). 173–200.
- Schmid, Hans-Jörg. 2014. Lexico-Grammatical Patterns, Pragmatic Associations and Discourse Frequency. In Thomas Herbst, Hans-Jörg Schmid & Susen Faulhaber (eds.), *Constructions – Collocations – Patterns*, 239–293. Berlin: Mouton de Gruyter.
- Sinclair, John. M. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, John. M. 1996. The search for units of meaning. *Textus* 9 (1). 75–106.
- Sinclair, John. M. 1998. The lexical item. In Edda Weigand (ed.), *Contrastive Lexical Semantics*, 1–24. Amsterdam: John Benjamins.
- Steyer, Kathrin (2019): *There's no X, only Y*. A Corpus-based Study of German and English Proverb Patterns. In Andreas Nolte & Dennis F. Mahoney (eds.), *Living by the Golden Rule. Mentor - Scholar - World Citizen. A Festschrift for Wolfgang Mieder's 75th Birthday*, 125–140. Berlin & Bern: Lang.
- Stubbs, Michael 2007. An example of frequent English: phraseology: Distributions, structures and functions. In Roberta Facchinetti (ed.), *Corpus Linguistics 25 Years On*, 89–105. Amsterdam: Rodopi.
- Stubbs, Michael. 2009. Technology and phraseology: With notes on the history of corpus linguistics. In Ute Römer & Rainer Schulze (eds.), *Exploring the Lexis-Grammar Interface*, 15–32. Amsterdam: John Benjamins.
- Stumpf, Sören. 2016. Modifikation oder Modellbildung? Das ist hier die Frage – Abgrenzungsschwierigkeiten zwischen modifizierten und modellartigen Phrasemen am Beispiel formelhafter (Ir-)Regularitäten. *Linguistische Berichte* 247. 317–342.
- Taylor, John R. 2016. Cognitive linguistics. In Keith Allan (ed.), *The Routledge Handbook of Linguistics*, 455–469. New York: Routledge.
- Traugott, Elizabeth Closs. 2007. The concept of constructional mismatch and type-shifting from the perspective of grammaticalization. *Cognitive Linguistics* 18 (4). 523–557.
- Traugott, Elizabeth Closs & Graeme Trousdale. 2013. *Constructionalization and Constructional Change*. Oxford: Oxford University Press.
- Van Lancker, Diana, Katherine Kougentakis, Krista Cameron, Carolyn Falconer & John Sidtis. 2012. *Down with ___*: The linguistic schema as intermediary between formulaic and novel expressions. *Yearbook of Phraseology* 3 (1). 87–107.
- Wray, Alison. 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Ziem, Alexander. 2018. Construction grammar meets phraseology: eine Standortbestimmung. *Muster im Sprachgebrauch. Construction grammar meets phraseology*. Special Issue *Linguistik Online* 90 (3/18). 3–19.

Part I: Productive patterns: monolingual

Pedro Ivorra Ordines

Comparative constructional idioms

A corpus-based study of the creativity of the [*más feo que X*] construction

Abstract: This study¹ adopts the view that phraseological comparisons underlie constructional idioms, given that they can be considered partially lexically filled patterns, that is, form-meaning pairings endowed with a pragmatic meaning. Against this background, the [*más feo que X*] ('very ugly') construction is analysed on the basis of the *esTenTen18* corpus, which yields a total of 1,108 occurrences with the intensifying meaning. Under the assumption that Construction Grammar is a much more suitable approach for the study of semi-schematic constructions than traditional Phraseology, this contribution aims to describe the five different senses that the construction can adopt, as well as to identify the conceptual fields of the slots that motivate its instances. Considering productivity and creativity are two sides of the same coin, the study determines whether Spanish speakers use more creative instances of the construction or lexicalised idioms, among which we focus on the highly entrenched [*más feo que pegarle a un padre*] micro-construction in order to analyse its productivity on the basis of its modifications and creative extensions attested in the corpus.

Keywords: constructional idiom, phraseological comparison, creativity, productivity.

¹ This research is being carried out within the framework of the research project PID2019-108783RB-I00: *Construction Grammar and Phraseology. German and Spanish Constructional Idioms through Corpora*. This project, funded by the Spanish Ministry of Science and Innovation, has been developed by the research group FRASESPAL under the direction of Carmen Mellado Blanco.

Pedro Ivorra Ordines, University Pompeu Fabra, Barcelona, pedro.ivorra@upf.edu

<https://doi.org/10.1515/9783110520569-002>

1 Introduction

This paper contributes to one of the most fruitful and innovative approaches to the development of phraseological research together with Construction Grammar. Even if interest in semi-schematic constructions was already present in the eighties in two of the cornerstones of constructionist approaches by Fillmore, Kay, and O'Connor (1988) and Fillmore and Kay (1999), it was not until recently that we have experienced an increasing interest from both language-specific and contrastive perspectives. In the context of the present study, the notion of “constructional idiom” gains importance – rather than idiom with modifications – given that a detailed corpus-based analysis of the [*más feo que X*] construction shows the difficulties in drawing a line between lexicalised idioms and creative instances of the construction because dictionaries may not include the most frequent idioms.² This study therefore highlights the creativity and variation that the [*más feo que X*] construction and the [*más feo que pegarle a un padre*] micro-construction³ undergo, given that speakers’ creativity is so common that it even allows variation in idioms that are fully specified at first sight (cf. Mellado Blanco, Corpas Pastor, and Berty 2021). Despite speakers’ ability in licensing new creative instances, the regularity in the irregularity can be captured, which means that creativity follows certain patterns.

The present study brings together the existing research on constructional idioms in general and conducts an analysis of the [*más feo que X*] construction in order to account for its characteristic traits and its creativity. To do so, a state of the art of the intersection between Phraseology and Construction Grammar is presented to determine whether the object of study might benefit from a construction-based approach (Section 2). Under the assumption that phraseological comparisons underlie constructional idioms (Section 3 and 4), a description of the [*más feo que X*] construction is given and, on the basis of the *esTenTen18* corpus, constructional polysemy is explained, and usage restrictions and preferred fillers are identified. Moreover, the productivity of the construction is tested by means of its type frequency and a small study of the creativity of the

2 This difficulty is suggested in the title of the study by Martí Solano (2013) “from idiom variants to open-slot idioms”, which highlights the problems in delimiting idioms with variants and idioms with empty slots.

3 Following the terminology by Traugott (2008), [*más feo que pegarle a un padre*] could be considered a micro-construction, that is, an individual construction-type.

[*más feo que pegarle a un padre*] micro-construction is carried out (Section 5). Conclusions drawn from this research are summarised in Section 6.

2 The interplay between Phraseology and Construction Grammar

Phraseology and Construction Grammar have many interests in common. This is not something new since the first construction-based contributions focused mainly on those idiomatic expressions that had long been neglected in the phraseological tradition: “it was prime examples of phraseologisms whose analysis “gave rise” to Construction Grammar in the first place” (Gries 2008: 14; cf. Croft and Cruse 2004: 225). Despite the fact that Phraseology and Construction Grammar have developed independently, ignoring how much they could benefit from one another, they have always agreed upon the fact that idiomatic expressions do not have a marginal status but are at the core of what they consider to be fundamental entities.

In rough outline, Construction Grammar can profit from a descriptive classificatory apparatus in the study of phraseological units, whereas Phraseology can make use of elaborate theoretical mechanisms and a usage-based approach to analyse structural, semantic, and pragmatic restrictions. This convergence has led to an increased number of contributions both from language-specific and contrastive points of view with the aim of capturing the regularity in the irregularity (Ziem 2018: 3; Stutz and Finkbeiner in this volume).

All phraseological units may, at first sight, be studied under a construction-based approach, especially if taking into account a wide conception of Phraseology (cf. Corpas Pastor 1996). This high degree of compatibility between phraseological research and Construction Grammar is particularly noticeable if we consider the definition that Goldberg (1995: 4) proposed: “C is a construction iff_{def} C is a form-meaning pair $\langle F_i, S_i \rangle$ such that some aspect of F_i or some aspect of S_i is not strictly predictable from C’s components parts or from other previously established construction”. According to this definition, all phraseological units can be understood as constructions, and hence, Phraseology would belong to the object of study of Construction Grammar, although claiming that all phraseological units can be analysed from a construction-based point of view is

to some extent debatable.⁴ Besides, Construction Grammar is not only devoted to the study of phraseological units (cf. Goldberg 1995, 2006) and Phraseology has its own goals and specific concepts.

Surprisingly enough, not until Goldberg's (2006: 5) definition of construction did linguists start to apply constructionist mechanisms to the study of phraseological units, taking into account that the new definition also considered as constructions those compositional expressions, as long as they occurred with sufficient frequency. This new definition was adopted under the assumption that "[n]ewer research has emphasized the extent to which speech consists of formulaic or prefabricated language" (Goldberg 2013: 26).

As noted, not all phraseological units can benefit from a construction-based analysis. Following the distinction that Fillmore, Kay, and O'Connor (1988) put forward, this would be the case of "substantive idioms", i.e. phraseological units that cannot be grammatically modified and whose components are lexically specified. Given that they can be considered lexical items, they would not require radical methodological changes and, therefore, they can still be studied under the traditional approaches based on morphology and syntax. There are, however, certain units that may profit from the mechanisms and the approach that Construction Grammar⁵ offers, since these units cannot be satisfactorily explained from the traditional phraseological point of view.

The phraseological units that can be fruitfully explained under a constructionist approach are those units that have some fixed constituents in their structure while others, subject to semantic restrictions, must be lexically filled in the discourse, i.e. structures with a global meaning with open slots in their lexical structures that must be filled with free lexical components that have certain semantic constraints. Therefore: on the one hand, they are not fully specified because the slots cannot be freely filled, so they cannot be considered lexemes and, at the same time, they do not belong to the syntax; on the other hand, they are not syntactic structures because they are partially lexically specified, so they do not belong to the grammar either.⁶ The kind of unit that is being described

⁴ The notion of "construction" is wider than the notion of "phraseological unit", which may explain the current situation of Phraseology: there is a clear extension of its limits, especially from the periphery (Feilke 2007: 64).

⁵ Cognitive Construction Grammar is the most suitable approach for the study of phraseological units. This approach may provide the answer to those problems that have arisen within the phraseological tradition concerning the delimitation of concepts and the taxonomy of phraseological units (Hennemann 2016: 68).

⁶ Construction Grammar is mainly useful for semi-schematic constructions and fully schematic constructions (grammatical structures). Since the latter does not belong to the field of Phra-

here perfectly fits one of the tenets of the constructionist approaches to grammar: the lexicon-grammar continuum. The generalization of the notion of “grammatical construction” to account for the language of a speaker implies that the grammar of a language is rather understood as a continuum of symbolic units between the lexicon and the grammar. Between these two poles there is a continuum of constructions that are partially schematic from the grammatical point of view or partially specified from the lexicon point of view (cf. Langacker 1987: 25–27, 35–36).

Despite the apparent commonalities between these two approaches, only in recent years have we witnessed an increased interest in the study of phraseological units from a construction-based approach (cf. Mellado Blanco 2015; Dobrovol'skij 2016; Mollica and Schafroth 2018). Particularly, in the Spanish phraseological tradition, these units have received little attention due to two main factors (Mellado Blanco 2020):

1. Contrary to the common belief within the phraseological tradition claiming that fixedness is one of the traits that characterises phraseological units, *esquemas fraseológicos* ‘phraseological schemas’ do not comply with this trait, given that they lack in fixedness. Including these particular units within the object of study of Phraseology would consequently go against the tenets of the discipline. Another trait that would go against the main traits is the productivity that defines phraseological units of this kind.
2. In accordance with the previous factor, since the very earliest discussions fixedness has been related to defectiveness because in its structure there is some kind of syntactic or morphological irregularity. As these units are characterised by a productive structure, they would be placed in the periphery of the morphological system.

3 What does the term constructional idiom mean?

The concept of constructional idiom gains importance in the context of this study in referring to those patterns or syntactic schemas with empty slots endowed with a specific pragmatic meaning. Other approximately equivalent concepts are “formal or lexically open idioms” (Fillmore, Kay, and O'Connor

seology, the intersection between Phraseology and Construction Grammar is by means of semi-schematic constructions (Dobrovol'skij 2018: 11).

1988), “schematic idioms” (Croft and Cruse 2004), “partially lexically filled phrasal pattern” (Goldberg 2006), and “constructional idioms”, following Booi’s terminology (2002: 301–302), who defines them as “syntactic constructions with a (partially or fully) non-compositional meaning contributed by the construction, in which – unlike idioms in the traditional sense – only a subset (possibly empty) of the terminal elements is fixed”.

Constructional idioms are mainly characterised by two principal traits that are applicable to constructions in general: pairings of form and function, and the idea of graduality and continuum. This notion is based on the degree of specificity and lexical saturation of construction’s constituents, which perfectly fits with what Goldberg (2006: 5) considers as constructions: “including morphemes or words, idioms, partially lexically filled and fully general phrasal patterns”. Therefore, even the most schematic constructions fulfill linguistic functions, so hence certain phraseological units are interesting from a construction-based approach because constructionist mechanisms allow for a systematic description of the changes that take place in its constructional components in order to see the structural and semantic properties that constrain the potential lexical fillers that may be inserted into the open slots of the construction.

Assuming that lexicon and grammar are not two separate modules but rather correspond to a continuum, constructional idioms are situated along this lexicon-grammar continuum. Taking this into account, the more abstract and productive the construction is, the closer it is to the grammatical pole; on the contrary, the more “phraseological”, the closer to the lexical pole. This means that constructions are licensed in the Goldberg’s sense and not generated, so that “if the conventionalized pattern of such unit is violated (e.g. including lexemes which are semantically incoherent with the whole construction), the unit loses its constructional meaning” (Piunno 2018: 143).

As a way of summarizing the preceding description, constructional idioms are characterised by the following traits:

- They make up schematic patterns, endowed with a global pragmatic meaning (intensification, attenuation, speaker’s intentionality).
- Some constituents of the syntactic structure are lexically specified, while others are free slots that must be renewed in the discourse.
- They have gradation in their schema productivity.
- The semantic, pragmatic, and syntactic traits are not entirely deductible from the grammatical and lexical principles of a language.
- The lexical renewal of the free slots is more or less free, having some semantic restrictions.

- They are placed between the lexical pole, made up of words, idioms or sayings, and the grammatical pole, which comprises fully schematic constructions.

4 Object of study: phraseological comparisons as constructional idioms

Phraseological comparisons are the most prototypical phraseological units to express a quality in the highest degree in Spanish. The intensifying value of this phenomenon is an important source of support for the appellative function with regard to the interlocutor, so that by means of this phraseological unit the speaker is trying to influence the hearer in a certain direction. Phraseological comparisons are made up of a parameter of comparison that can be an adjective, a noun or a verb, a standard marker (in the case of our study *más...que* ‘more...than’), and a standard of comparison. The comparee does not belong to the fixed constituents of the phraseological comparison since it is usually the subject of the sentence.

This study stems from the assumption that phraseological comparisons underlie constructional idioms, so they can be understood as pairings of form and meaning in the constructionist sense (cf. Mellado Blanco 2012; Mollica and Schafroth 2018). Phraseological comparisons behave in a similar way to constructional idioms since at least one constituent must be lexically filled, which explains why they are situated within the continuum of schematicity and are productive. This construction is hence placed on the continuum that exists between the two poles of lexical coherence (specificity) and openness (schematicity).

Intensification is precisely what differentiates comparative constructions (cf. Hilpert 2010) and comparative constructional idioms. Another difference lies in the standard of comparison, because in the former an extralinguistic referent is alluded, while in the latter it is the result of a linguistic convention. This explains why they cannot be considered the same construction, given that “it would be impossible to arrange them into a hierarchical tree: first, selecting the root node would be a wholly arbitrary decision and second, the variants only display partial overlaps, not the kind of heritage [...] where the same syntactic configuration is preserved throughout the network” (Fried 2015: 986). Despite being considered two different constructions they can be placed on the idiomaticity continuum, which ranges “from structures that are completely

idiosyncratic and lexically particular to structures that share most of their semantic and syntactic properties with other grammatical patterns” (Diessel 2019: 56). Following the form and meaning description by Hoffmann (2017), both constructions are given, with a bidirectional arrow to refer to the arbitrary pairing of form and meaning.

– Comparative construction

FORM: [X SER *más* ADJ *que* Y] ‘X BE more ADJ than Y’ ↔ MEANING: ‘X es *más* ADJ *que* Y’ ‘X is more ADJ than Y’

- (1) *Hombre, peso un poco menos y soy **más feo que José**, obviamente, como dices [...].*⁷

‘Man, I weigh a bit less and I am uglier than Jose, obviously, as you say [...].’

(esTenTen18, 382306380)

– Comparative constructional idiom

FORM: [X SER *más feo que* Y] ‘[X BE uglier than Y]’ ↔ MEANING: ‘muy feo’ ‘very ugly’ {intensification}

- (2) *Se le recuerda como el mejor científico del siglo XX por haber ganado dos veces el Premio Nobel al hombre **más sexy del mundo**, a pesar de ser **más feo que <inserta tu nombre aquí>**.*⁸

‘He is mainly remembered as the best scientist of the 20th century for having won twice the Nobel Prize for the sexiest man in the world, despite being uglier than <insert your name here>.’

(esTenTen18, 16012132830)

⁷ The examples have been subjected to a spelling and punctuation review.

⁸ Examples are translated literally so that the image of the comparisons can be appreciated by the reader. All of them adhere to the meaning of ugliness in its highest degree. As is explained in detail in the section 5.2., the [*más feo que* X] ‘uglier than X’ construction has one prototypical meaning from which four extensions are posited. In this particular example, the central meaning is alluded, which refers to physical ugliness.

5 The [más feo que X] construction

5.1 Corpus and methodology

For the analysis of the present study, the *esTenTen18* corpus was used, which comprises the *European Spanish Web* and the *American Web* subcorpora. While European Spanish represents almost half of the corpus (49.32%), American Spanish represents 46.46%⁹, and 4.21% is from an undefined origin. The corpus consists of 17,553,075,259 words, which makes it a large corpus, available in *Sketch Engine*. Even though idioms tend to be low-frequency items, the fact that most of the texts in the corpus are from webpages including forums and chats makes it suitable for their study because idioms usually appear in conversations (especially informal, colloquial registers or slang). It is precisely in these registers that productivity and creativity are a predominant tendency, as will be seen in the great variety of examples given.

A search of the pattern <más feo que> in the *esTenTen18* corpus returns 3,238 occurrences. However, not all occurrences comply with the quality of ugliness in its highest degree that particularises the type of constructional idiom that we are dealing with in this study. This means that in this first search other grammatical phenomena can be pointed out, such as grammatical comparisons (examples 3–4) or constructions where someone's beauty is highlighted by means of the rhetorical device of irony (example 5).

- (3) *Los propietarios de coches Dacia franceses han demostrado que no es necesario devanarse los sesos pensando en prestaciones, valor de reventa y ni siquiera en si el coche será más bonito o **más feo que el del vecino**, porque no les importa en absoluto.*

'French Dacia car owners have shown that there is no need to rack one's brain over performance, resale value, and even whether the car will be more beautiful or uglier than your neighbour's, because they do not care at all.'

(*esTenTen18*, 524886689)

- (4) *Si no haces cosas estúpidas cuando eres joven, no tendrás nada de que reírte cuando llegues a viejo. Tips de Belleza: "Camina siempre con alguien **más***

⁹ The similarity in size may yield interesting results concerning diatopic variation. Due to limited space, however, we cannot devote more space to this interesting point.

feo que tú”.

‘If you don’t do stupid things when you are young, you will have nothing to laugh about when you get old. Beauty tips: “Always walk with someone uglier than you”.’

(esTenTen18, 2118038306)

- (5) *Tendré que pensármelo un ratito para ponerlo en mi blog :) Recogido queda el guante, compañero (aunque tú eres un cacho más feo que Rita Hayworth, por lo del símil del guante, vamos).*

‘I will have to think over it a bit before posting it in my blog :) I mean, the challenge is taken up, mate (although you are uglier than Rita Hayworth, because of the glove simile).’

(esTenTen18, 9270323370)

Considering that the 3,238 occurrences resulting from the extraction represent other linguistic phenomena, these occurrences have to be discarded so that the ones to be considered in this analysis need to have the meaning of ugliness in its highest degree. After the filtering, there are 1,108 constructional idioms intensifying the quality of ugliness, which roughly corresponds to one third of the total number of occurrences resulting from the non-filtered extraction (34.22% to be more precise).

5.2 General traits of the construction

The existence of polysemy perfectly fits with the notion of “construction”, given that, understood as a form-meaning pairing, it is subject to polysemy effects, in the same way as words are. The [*más feo que X*] construction is, therefore, associated with a family of closely related senses rather than a single, fixed abstract sense (Goldberg 1995: 31; cf. González-García 2013: 190). Considering the differences in meaning that may arise within this construction, the semantics involved can be best clustered as groups of exemplar categories by semantic similarity, and these, in turn, can exhibit prototype effects (cf. Bybee 2013). A polysemous analysis allows for the recognition of a prototypical sense and extensions from this sense that can be related via polysemy links.

The entries that the *Diccionario de uso del español* (DUE) provides are taken as a starting point in order to determine the different senses that can be found in the occurrences of the corpus that we are working with. In the following paragraphs, we will indicate the five senses that characterise the [*más feo que X*]

construction together with an example in order to see the small nuances that differentiate and particularise these senses. Some definitions will be used in the same way as the dictionary indicates, while others will be modified so as to adapt them to the needs of this study and one final sense will be posited.

The central sense can be argued to be the one in which physical ugliness is alluded to. This is the most prototypical sense, and from this, extensions of meaning can be posited by means of polysemy links. Out of 1,108 occurrences, the prototypical sense is found in 1,048 occurrences, which represents the vast majority of the corpus (94.58%). According to the DUE (2007: 1348), it is defined as follows “1 Se aplica a lo que impresiona desagradablemente el sentido de la vista” ‘1 Applied to what makes an unpleasant impression to the sense of sight’.

- (6) *La gente que lleva Mac se cree la élite. Por el simple hecho del diseño. Cuando las mejores computadoras de los años 60 eran **más feas que pegar a un padre**. No nos equivoquemos y os lo digo yo, que soy Mac-ero convencido.*
 ‘People with a Mac think they belong to the elite. Just because of its design. When the best computers of the sixties were uglier than hitting a father. Let’s not be mistaken and it’s me who’s telling you, I’m a committed Mac supporter.’
 (esTenTen18, 5711940005)

The first extension of meaning has the sense that conforms to the following definition: “2 Aplicado a acciones, *indelicado” ‘2 Applied to actions, *indelicate’ (DUE 2007: 1348). However, we should add the nuance of ‘blamable’, ‘denigrating’ to refer to any action which may be offensive to someone. Contrary to the prototypical sense, which is clearly the most frequent, this extension, as well as the following ones, is far less frequent. Out of 1,108 occurrences, this sense only has 30 occurrences¹⁰ in the corpus, which represents 2.71%.

10 Particularly interesting is the occurrence that can both fit in the central sense and the first extension, given that it is both visually unappealing and denigrating at the same time: (A veces puede significar “un tiempo indefinido o también la no realización de algo”, sin decirlo directamente). Esto es **más feo que pegarle a la mamá** (“algo es feísimo, asqueroso”). ‘(Sometimes it means “an undefined time or not carrying something out” without saying it directly). This is uglier than hitting the mother (“something is very ugly, disgusting”). (esTenTen18, 11123199294).

- (7) *¿Alguien sabe dónde se puede descargar esta serie? Pero en V.O., porque ver una serie inglesa con doblaje español es **más feo que pegarle a una madre**.*
 ‘Anyone knows where I can download this TV series in V.O.?, because watching an English TV series dubbed in Spanish is uglier than hitting a mother.’
 (esTenTen18, 10157645207)

The second extension of meaning adheres to the following definition: “3 De mal aspecto; de manera que *amenaza con algo desagradable” ‘3 Bad looking, so that it *threatens with something unpleasant’ (DUE 2007: 1348). This sense also exhibits a low frequency in the corpus: 14 appearances out of 1,108 occurrences, which only represents 1.26%.

- (8) – *¿Qué te pasa, Juan? Te veo **más feo que al gobierno de Maduro**, te veo muy afligido. – Es que mi mujer tiene un amante, pero yo no lloro por esa vaina, [...].*
 ‘–What is wrong with you, Juan? You look uglier than Maduro’s government, you look very afflicted. –My wife has a lover, but I don’t cry for such a crap, [...].’
 (esTenTen18, 6601659060)

The third extension of meaning makes reference to the definition as follows: “4 Col. Desagradable al gusto o al olfato.” ‘4 Col. Disagreeable to taste or smell’ (DUE 2007: 1348). Considering the fact that some of the examples allude to the sense of hearing, it should be included in the definition, usually referring to a noise that is unpleasant to hear. This sense also has low frequency, given that there are 12 occurrences reflecting this specific sense. This means that only 1.08% of the constructs have this meaning.

- (9) *Esta radio hace cinco años era infumable y ahora, exceptuando la música (canciones sesenteras **más feas que Picio**) y las oraciones comunitarias [...], el resto es una radio católica de calidad media [...].*
 ‘This radio was awful five years ago and now, except for the music (songs from the sixties uglier than Picio) and the community prayers [...], the rest is a Catholic radio of medium quality [...].’
 (esTenTen18, 10652066356)

The last extension of meaning has no corresponding definition in the dictionary that serves as a basis for this case of constructional polysemy. Following the

examples that comply with this meaning, it can be stated that it refers to when someone does not feel like doing anything, when doing something is unappealing. There are three occurrences that stick to this specific meaning, which means that only 0.27% adhere to this sense.

(10) *Cuando la salud se le fue y la muerte se lo llevó, sus amigos compungidos pusieron un cartelito a los pies del ataúd como preciado homenaje: “Morirse es más feo que Lanata en portaligas”.*

‘When he lost his health and death took him away, his remorseful friends placed a little sign at the foot of the coffin as a cherished tribute: “Dying is uglier than Lanata in garter belts”.’

(esTenTen18, 9063348459)

This study stemmed from the assumption that phraseological comparisons underlie constructional idioms, which allows us to make an accurate description of the changes that take place in its constructional components so as to see the structural and semantic properties that constrain the lexical fillers that may be inserted into the open slot of the construction. Generally speaking, the slot is filled by means of three different referents: ugly referents (both people or things), disgraceful, blamable actions, and negative situations.

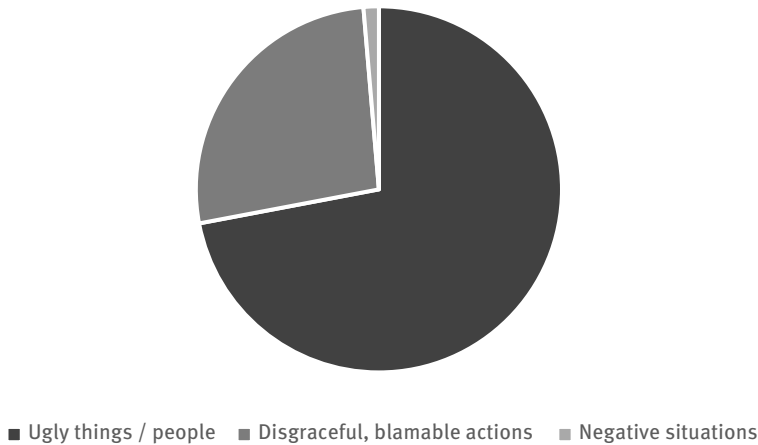


Fig. 1: Conceptual field of the slot X of the [*más feo que X*] ‘uglier than X’ construction (esTenTen18)

As Figure 1 shows, the most predominant conceptual field of the slot X in the [*más feo que X*] construction is represented by ugly objects or people (example 11), which amounts to 798 instances, that is 72% of the occurrences, followed by disgraceful, blamable actions (example 12). This second group comprises 295 instances, which represents 27%. Finally, only 1% (15 instances) is represented by negative situations (example 13). Despite using only three conceptual fields as referents to intensify the quality of ugliness, it is worth noting some recurrent lexical fields: kinship and human naming, household appliances, means of transport, religious-related items, famous characters, nationalities, animals, scatology, and food.

- (11) *Me gustan los insectos. ¿Por qué todo el mundo ama las mariposas, Manu? O sea, son horrendas, son **más feas que mi madre sin maquillaje**. Los caracoles son adorables, por ejemplo... – Los caracoles son hermafroditas.*
 ‘I like insects. Why does everyone in the world love butterflies, Manu? I mean, they are awful, uglier than my mother with no makeup. Snails are adorable, for instance... –Snails are hermaphrodites.’
 (esTenTen18, 1432482827)
- (12) *En cuanto aspecto, se sigue comentando que será el de unas gafas deportivas, pero también aseguran que no estarán pensadas para ser usadas continuamente, por lo que eso quiere decir que muy probablemente serán **más feas que pegarle a un padre** o tan incómodas como unas gafas de alambre de espino.*
 ‘As far as appearance is concerned, they keep saying that it will be that of a pair of sports glasses, but they also assure that they will not be designed to be used continuously, so that means they will most likely be uglier than hitting a father or as uncomfortable as a pair of barbed wire glasses.’
 (esTenTen18, 1336124451)
- (13) *Keky: No seas extremista, nene. El ejemplo que me pones es el de “el tuerto en el país de los ciegos”. Cierto, el tío sería **más feo que un susto** e iría extravagante, pero... era “diferente” y eso atrae, aunque sea por curiosidad.*
 ‘Keky: Don’t be an extremist, baby. The example you give me is that of “the one-eyed man in the land of the blind”. True, the guy would be uglier than a fright and would be extravagant but... he was “different” and that attracts attention, albeit out of curiosity.’
 (esTenTen18, 1160991853)

Finally, it is also worth indicating that the frequent use of a specific standard of comparison may lead to its automatism, which provokes its bleaching. This type of standard of comparison is known as “passepartout”, which is mainly used when the specific standard of comparison does not come to the speaker’s mind, or when a standard of comparison does not exist for a specific construction. Examples 14–16 contain “passepartout” standards of comparison¹¹ that are found in the corpus that we are working with:

(14) *Esa misma técnica la hace un tío que no tenga ninguna confianza y otro que sea **más feo que un demonio**, y claramente no funcionaría.*

‘If this technique is used by another guy with no confidence or another who is uglier than a devil, and surely it would not work.’

(esTenTen18, 13441743586)

(15) *La Diablesa tiene cuernos, [...] una manzana en una mano, para indicar su conexión con Eva con toda la mala leche del mundo, y un montón de signos que no cabrían en este artículo... Y, naturalmente, **más fea que la leche**...*

‘The She-Devil has horns, [...] an apple in her hand, to show her connection to Eve with all the bad blood in the world, and a lot of signs that would not fit in this article... And obviously, uglier than milk.’

(esTenTen18, 20043444787)

(16) *Tenías que haber ido con él y llevarle la copa esa. Por cierto, ahora que ha acabado el mundial, ya puedo decir que es **más fea que la hostia**.*

‘You should have gone with him and brought him that cup. By the way, now that the World Cup is over, I can already say that it [the cup] is uglier than the host.’

(esTenTen18, 4104500380)

¹¹ Some of these micro-constructions are: [*más malo que un demonio*] ‘worse than a devil’, [*más satánico que un demonio*] ‘more diabolical than a devil’, [*más fuerte que un demonio*] ‘stronger than a devil’, [*más rojo que un demonio*] ‘redder than a devil’, [*más loco que un demonio*] ‘crazier than a devil’, [*más bonito que la leche*] ‘more beautiful than the milk’, [*más inútil que la leche*] ‘more useless than the milk’, [*más contento que la leche*] ‘happier than the milk’, [*más majo que la leche*] ‘friendlier than the milk’, [*más rápido que la leche*] ‘faster than the milk’, [*más contento que la hostia*] ‘happier than the host’, [*más lento que la hostia*] ‘slower than the host’, [*más tonto que la hostia*] ‘sillier than the host’, [*más majo que la hostia*] ‘friendlier than the host’, [*más raro que la hostia*] ‘rarer than the host’.

5.3 Creativity: the role of analogical extensions

Constructional approaches share with mainstream generative grammar the assumption that creativity is an essential property of language and the goal of linguistics is, therefore, to give an explanation for the creative potential of language. Considering that constructions can be combined freely as long as they are not in conflict means that “a speaker is free to creatively combine constructions as long as constructions exist in the language [...] to categorize the target message” (Goldberg 2006: 22).

In linguistics approaches, the position that is commonly adopted is that productivity is intrinsically related to creativity or analogy to establish a distinction between novel utterances that are regular and systematic – i.e. productive – and those that are exceptional and non-systematic – i.e. creative – (Zeschel 2012: 122). The process of creativity, or analogy, perfectly fits on the lexicon-grammar continuum given that it allows us to empirically study constructions and, at the same time, to reach generalizations and abstractions. This explains why productivity and analogy are two sides of the same coin, so the difference between “full” productivity and analogy is a difference in degree, not in kind (Barðdal 2008: 2).

The phenomenon of productivity refers to speakers’ ability to license new constructions that have never been encountered before. There is a considerable consensus among linguists (Bybee 1995, 2013) in regard to the claim that type frequency determines the productivity of a construction, given that it “counts how many different lexical items a certain pattern or construction is applicable to” (Bybee and Thompson 1997: 378). Following this reasoning, a distinction is established in order to determine the type frequency of the [*más feo que X*] construction: on the one hand, those lexicalised idioms that appear in the dictionary together with their analogical extensions, which do not count as different items since analogical extensions are posited from lexicalised idioms; on the other hand, creative instances of the construction under study, which may be related to each other via cognitive links in its formation, but are not extensions of a lexicalised form.

The first group is made up of 405 occurrences, which represents 36.55% of the total number of occurrences in the corpus. Out of these 405 occurrences, 329 appear in the lexicographical resources¹² (81.23%), while 76 are considered ana-

¹² *Diccionario fraseológico documentado del español actual: locuciones y modismos*, *Diccionario fraseológico del español moderno* and *Gran Diccionario de frases hechas* are used as lexicographic resources to determine those idioms that are lexicalised.

logical extensions of a lexicalised form. This means that there is little variation or creativity in those idioms that appear in lexicographic sources, given that only 18.77% of the occurrences are considered variants of a lexicalised form. In this group, only eleven different lexical items are counted¹³, since analogical extensions belong to the count of the lexicalised form from which they are posited.

The second group is made up of 703 occurrences, which represents 63.45% of the total number of the occurrences in the corpus. These instances are not modifications from a lexicalised form, but creative instances from the [*más feo que X*] construction. In this group, 474 different lexical items are counted, which means that the productivity of the construction is very high considering the high number of types (42.78% of the creative occurrences are different from each other). The wide variety of types makes the [*más feo que X*] construction a highly productive pattern, since “the more items the category must cover, the more general will be its criterial features and the more likely it will be to extend to new items” (Bybee and Thompson 1997: 384). This is intrinsically related to the fact that many different lexical items (famous characters, household appliances, means of transport, religious-related items, scatology, among many others) are covered by the [*más feo que X*] construction, which ensures that the construction is highly entrenched, making it more accessible for further use with new items.

Under the assumption that productivity and creativity or analogy are two sides of the same coin, token frequency is no less important than type frequency in determining the productivity of a construction, in contrast to the common belief that claims that token frequency detracts from productivity (cf. Bybee 1995, 2010). According to Barðdal (2008: 98), therefore, token frequency has an important role when positing analogical extensions in low-level constructions, given that “highly entrenched tokens are more easily activated than non-entrenched tokens, and are thus more available as model items than infrequently occurring tokens”. In other words, creativity is considered as a productivity index of a pattern because creative uses of language indicate language user’s extension strategies. Against this background, the lexicalised, highly frequent

13 [*más feo que un dolor*] ‘uglier than a pain’, [*más feo que un pecado*] ‘uglier than a sin’, [*más feo que pegarle a un padre*] ‘uglier than hitting a father’, [*más feo que Picio*] ‘uglier than Picio’, [*más feo que Carracuca*] ‘uglier than Carracuca’, [*más feo que Dios*] ‘uglier than God’, [*más feo que un demonio*] ‘uglier than a demon’, [*más feo que un diablo*] ‘uglier than a devil’, [*más feo que la leche*] ‘uglier than milk’, [*más feo que la hostia*] ‘uglier than hell’, [*más feo que una bruja*] ‘uglier than a witch’.

[*más feo que pegarle a un padre*] micro-construction is analysed so as to discuss whether it can be considered a productive constructional idiom on the basis of the modifications and creative extensions attested in the corpus. In order to test it, the 50% benchmark suggested by Stumpf (2016) is adopted to see whether [*más feo que pegarle a un padre*] is a modificatory pattern that allows for occasional modifications or a true phraseological schema. Considering that only 52 occurrences are modifications or creative puns out of the 149 total occurrences (34.9%), we are not dealing with a phraseological schema, but rather a modificatory pattern with its analogical extensions and modifications. Besides, the number of hapaxes is very low: only 12 occurrences appear once in the corpus, which means that the micro-construction is not highly entrenched in this sense (examples 17–19).

- (17) *Basta con que mires los perfiles de las tías, siempre con filtros, con sombras y con poca luz para ocultar sus estrías, sus arrugas, sus granos. Son **más feas que pegarle a un padre con un calcetín lleno de arroz**, directamente.*

‘Just look at the chicks’ profiles, always with filters, shadows and low light to hide their stretch marks, wrinkles, pimples. They are basically uglier than hitting a father with a sock full of rice.’

(esTenTen18, 15183977707)

- (18) *Tenéis serios problemas psiquiátricos, tronkos, pero que muy serios. ¡Vergüenza de género! Coño, pero si eres **más fea que pegarle a un padre en la cara con una alpargata vieja**.*

‘You have serious psychiatric problems, mates, and big ones. Gender shame! Fuck, you are uglier than hitting a father’s face with an old espadrille.’

(esTenTen18, 11609351447)

- (19) *Ferro y Sandman se lo encuentran con una **más fea que pegarle a un padre con una foto de una nevera por detrás**¹⁴. Después, van Mel y Sandman y se encuentran a Ferro con una... Bueno, mejor no hablar, que puede herir sensibilidades.*

‘Ferro and Sandman find him with one uglier than hitting a father with a

¹⁴ Given the coincidental resemblance between the strings of instances of the target construction and the contaminating construction, it can be stated that this instance arises from the interaction of [*más feo que pegarle a un padre*] and [*más feo que una nevera por detrás*]. Since this is not the goal of the current paper, we cannot devote more space to this interesting point.

picture of a back's fridge. Then, Mel and Sandman find Ferro with a... Well, better not to talk, that can hurt feelings.'

(esTenTen18, 12968603433)

Cognitively speaking, extensions make reference to the instrument that is being used to hit the father, where socks are the most predominant ones, but we can also find slippers. What calls our attention is the fact that these instruments are accompanied by negative adjectives that highlight the denigrating, blamable character of the action. Extensions are considered to be a frequent mechanism in phraseological comparisons given that the main aim of the construction is to intensify the idiomatic meaning of the comparison. Examples 20–21 show analogical extensions with a humorous effect and can be seen as modifications that may become lexicalised over time:

- (20) *Superficial, ¡no!, simplemente que sigo la premisa de mente sana in corpore sano. Por cierto, Hitler y Franco eran **más feos que pegar a un padre con un calcetín sucio.***

'Superficial, no, simply that I follow the premise of *mens sana in corpore sano*. By the way, Hitler and Franco were uglier than hitting a father with a dirty sock.'

(esTenTen18, 17244375234)

- (21) *Para que te hagas una idea, el HuffPost de España ha publicado una portada que es **más fea que pegar a un padre con un calcetín sudado.** ¿Qué ha conseguido con esto? Notoriedad.*

'To give you an idea, the HuffPost in Spain has published a cover that is uglier than hitting a father with a sweaty sock. What has this achieved? Notoriety.'

(esTenTen18, 3484298468)

6 Summary

In this contribution we have examined the constructional idiom [*más feo que X*]. Construction Grammar was chosen as a descriptive model because it offers many advantages when describing partially lexically filled constructions. While Phraseology has difficulties in classifying phraseological comparisons and sometimes its classification is not consistent in Spanish phraseological research (cf. Penadés Martínez 2012: 130–132), Construction Grammar allows for a uni-

form description of the structural, semantic, and pragmatic restrictions that characterise the free slot of the construction. Besides, Construction Grammar enables us to account for the potential creative use of language.

Under the assumption that semi-schematic constructions represent the intersection between Phraseology and Construction Grammar, phraseological units with empty slots are of great interest to constructionist approaches because they fit perfectly on the lexicon-grammar continuum. Our study supports the view that phraseological comparisons underlie constructional idioms, given that they can be considered patterns with empty slots that are endowed with a specific pragmatic meaning.

A construction-based approach of the [*más feo que X*] construction allows us to give full account of the phenomenon of constructional polysemy, which is an important trait in the study of the 1,108 occurrences because five different meanings can be distinguished: physical ugliness (94.58%), blamable, denigrating actions (2.71%), threatening with something unpleasant (1.26%), disagreeable to taste, smell or hearing (1.08%), something unappealing to do (0.27%). Besides, the assumption that phraseological comparisons underlie constructional idioms enables us to identify the properties of the potential fillers that may be inserted into the open slot of the construction. The analysis of the corpus has shown, therefore, that three types of referents intensify the quality of ugliness: 72% of the occurrences are represented by ugly objects or people, 27% by blamable, disgraceful actions, and 1% by negative situations.

In pursuing the description of the creativity of the [*más feo que X*] construction, the results of the study have shown that speakers are more likely to opt for creative uses of the construction (63.45%) in lieu of lexicalised forms (36.55%), which, despite this, are frequently used (for instance, *más feo que pegarle a un padre* with 97 occurrences). Adopting the view that constructions exist at various levels of schematicity implies gradability, which means that productivity is a gradable phenomenon. Since the slot can be filled with 485 lexical items (counting both groups), this means not only that the [*más feo que X*] construction is highly entrenched in this sense, but it also ensures that “[c]onstructions that have been heard used with a wide variety of types are more likely to be extended broadly than constructions that have used with a semantically circumscribed set of types” (Goldberg 2006: 99).

The study highlights the importance of token frequency in determining how productive a construction can be, and thus being considered as an index of productivity. This assumption is of great importance in low-level constructions, which accounts for analogical extensions that are posited from [*más feo que pegarle a un padre*]. Given that only 34.9% of the occurrences are modifications

from the prototypical construction, this cannot be considered a case of phraseological schema following the 50% benchmark proposed by Stumpf (2016). These analogical extensions serve as a mechanism to creatively intensify the standard of comparison, which sometimes adopts a humorous character.

Last but not least, the present study has illustrated the fine line that exists between lexicalised forms and creative instances. The assumption that lexicon and grammar are not two separate modules enables us to place idioms (lexicalised forms) and constructions (patterns with empty slots) on the lexicon-grammar continuum. Besides, the notion of “constructional idiom” gains importance in this context since it allows a unified account of productivity and analogy, which is in perfect consonance with the assumption that productivity is a gradable phenomenon, and productivity and analogy are located at different points on the productivity cline.

References

- Barðdal, Jóhanna. 2008. *Productivity. Evidence from Case and Argument Structure in Icelandic*. Amsterdam & Philadelphia: John Benjamins.
- Boas, Hans & Alexander Ziem. 2018. Approaching German syntax from a constructionist approach. In Hans Boas & Alexander Ziem (eds.), *Constructional approaches to syntactic structures in German*, 1–44. Berlin & New York: Mouton de Gruyter.
- Booij, Geert. 2002. Constructional Idioms, Morphology, and the Dutch Lexicon. *Journal of Germanic Linguistics* 14 (4). 301–327. <https://www.cambridge.org/core/journals/journal-of-germanic-linguistics/article/abs/div-classtitleconstructional-idioms-morphology-and-the-dutch-lexicondiv/949BC19778B CA818B4C44D130C133714> (accessed 17 July 2020).
- Bybee, Joan. 1995. Regular Morphology and the Lexicon. *Language and Cognitive Processes* 10. 425–455.
- Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
- Bybee, Joan. 2013. Usage-based Theory and Exemplar Representation of Constructions. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*, 49–69. Oxford: Oxford University Press.
- Bybee, Joan & Sandra Thompson. 1997. Three Frequency Effects in Syntax. *Proceedings of the Twenty-Third Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on Pragmatics and Grammatical Structure*, 378–388. <https://journals.linguisticsociety.org/proceedings/index.php/BLS/article/view/1293/1077> (accessed 17 July 2020).
- Corpas Pastor, Gloria. 1996. *Manual de fraseología española*. Madrid: Gredos.
- Croft, William & Alan Cruse. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- DFDEA: Seco, Andrés, Olimpia Andrés & Gabino Ramos. 2004. *Diccionario fraseológico documentado del español actual: locuciones y modismos*. Madrid: Aguilar.

- DFEM: Varela, Fernando & Hugo Kubarth. 1994. *Diccionario fraseológico del español moderno*. Madrid: Gredos.
- Diessel, Holger. 2019. Usage-based construction grammar. In Ewa Dąbrowska & Dagmar Divjak (eds.), *Cognitive Linguistics – A Survey of Linguistic Subfields*, 50–80. Berlin & Boston: Mouton de Gruyter.
- Dobrovol'skij, Dimitrij. 2016. Fraseología y Gramática de Construcciones. *Language Design* 18. 71–106. http://elies.rediris.es/Language_Design/LD18/LD18_03_DOBROVOLSKII.pdf (accessed 17 July 2020)
- Dobrovol'skij, Dimitrij. 2018. Sind Idiome Konstruktionen? In Kathrin Steyer (ed.), *Sprachliche Verfestigung: Wortverbindungen, Muster, Phrasem-Konstruktionen*, 11–24. Tübingen: Narr Francke Attempto.
- DUE: Moliner, María. 2007. *Diccionario de uso del español*. Madrid: Gredos.
- Feilke, Helmuth. 2007. Syntaktische Aspekte der Phraseologie III: Construction Grammar und verwandte Ansätze. In Harald Burger, Dimitrij Dobrovol'skij, Peter Kühn & Neal Norrick (eds.), *Phraseologie. Ein internationales Handbuch zeitgenössischer Forschung*, 63–76. Berlin & New York: Mouton de Gruyter.
- Fillmore, Charles J., Paul Kay & Mary Catherine O'Connor. 1988. Regularity and Idiomaticity in Grammatical Constructions: The Case of *Let Alone*. *Language* 64 (3). 501–538. <http://pdfs.semanticscholar.org/23e2/7224928a008c5be84a7ea1452e0c627fe380.pdf> (accessed 17 July 2020)
- Fried, Mirjam. 2015. Construction Grammar. In Tibor Kiss & Artemis Alexiadou (eds.), *Syntax – Theory and Analysis. An International Handbook*. Volume 2, 974–1003. Berlin: Walter de Gruyter.
- GDFH: Larousse. 2001. *Gran Diccionario de frases hechas*. Barcelona: Spes Editorial.
- Goldberg, Adele. 1995. *A Construction Grammar Approach to Argument Structure*. Chicago & London: The University of Chicago Press.
- Goldberg, Adele. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford & New York: Oxford University Press.
- Goldberg, Adele. 2013. Constructionist Approaches. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*, 15–31. Oxford & New York: Oxford University Press.
- González-García, Francisco. 2013. Breves consideraciones en torno a la interacción entre coerción y polisemia: El caso de la predicación secundaria con verbos de cognición en español. In Sabine De Knop, Fabio Mollica & Julia Kuhn (eds.), *Konstruktionsgrammatik in den romanischen Sprachen*, 187–203. Frankfurt: Peter Lang.
- Gries, Stefan. 2008. Phraseology and linguistic theory. A brief survey. In Sylviane Granger & Fanny Meunier (eds.), *Phraseology: An interdisciplinary perspective*, 3–25. Amsterdam & Philadelphia: John Benjamins.
- Hennemann, Anja. 2016. Phraseologismen aus konstruktionsgrammatischer Sicht. In Christian Discher, Benjamin Meisnitzer & Claudia Schlaak (eds.), *Komplexität von Phraseologismen in den romanischen Sprachen: Theorie und Praxis in der Linguistik und der Fremdsprachendidaktik*, 51–71. St Ingbert: Röhrig Universitätsverlag.
- Hilpert, Martin. 2010. Comparing comparatives: A corpus-based study of comparative constructions in English and Swedish. In Hans Boas (ed.), *Contrastive Studies in Construction Grammar*, 1–42. Amsterdam & Philadelphia: John Benjamins.

- Hoffmann, Thomas. 2017. The Renaissance of constructions: From constructions to Construction Grammars. In Barbara Dancygier (ed.), *The Cambridge Handbook of Cognitive Linguistics*, 284–309. Cambridge: Cambridge University Press.
- Kay, Paul & Charles J. Fillmore. 1999. Grammatical Constructions and Linguistic Generalizations: The *What's X Doing Y?* Construction. *Language* 75 (1). 1–33.
https://pdfs.semanticscholar.org/0b8f/b343f979bf1bc48eed705cd1ada4075140de.pdf?_ga=2.208618582.572245901.1612553016-61201396.1604511679 (accessed 17 July 2020)
- Langacker, Ronald. 1987. *Foundations of Cognitive Grammar: Volume I: Theoretical Prerequisites*. Stanford & California: Stanford University Press.
- Martí Solano, Ramón. 2013. From idiom variation to open-slot idioms: close-ended and open-ended variational paradigms. In Joanna Szerszunowicz, Bogusław Nowowiejski, Katsumasa Yagi & Takaaki Kanzaki (eds.), *Research on Phraseology Across Continents*, 149–162. Białystok: University of Białystok Publishing House. <https://hal-unilim.archives-ouvertes.fr/hal-00933011/document> (accessed 17 July 2020).
- Mellado Blanco, Carmen. 2012. Las comparaciones fijas en alemán y español: algunos apuntes contrastivos en torno a la imagen. *Linred: Lingüística en la Red* 10. 1–32.
http://www.linred.es/monograficos_pdf/LR_monografico10-articulo1.pdf (accessed 17 July 2020).
- Mellado Blanco, Carmen. 2015. Antiphrasis-Based Comparative Constructional Idioms in Spanish. *Journal of Social Sciences* 11 (3). 111–127.
<https://thescipub.com/pdf/jssp.2015.111.127.pdf> (accessed 17 July 2020).
- Mellado Blanco, Carmen. 2020. Esquemas fraseológicos y construcciones fraseológicas en el continuo léxico-gramática. In Carsten Sinner, Esteban Montoro del Arco & Encarnación Tabares (eds.), *Clases y categorías en la fraseología española*, 13–36. Frankfurt am Main: Peter Lang.
- Mellado Blanco, Carmen, Gloria Corpas Pastor & Katrin Berty. 2021. El hablar y el discurso repetido: la fraseología. In Óscar Loureda & Ángela Schrott (eds.), *Manual de lingüística del hablar*, 519–540. Berlin & New York: Mouton de Gruyter.
- Mollica, Fabio & Elmar Schafroth. 2018. Der Ausdruck der Intensivierung in komparativen Phrasem-Konstruktionen im Deutschen und im Italienischen: eine konstruktionsgrammatische Untersuchung. In Kathrin Steyer (ed.), *Sprachliche Verfestigung. Chunks, Muster, Phrasem-Konstruktionen*, 103–136. Tübingen: Narr Francke Attempto Verlag.
- Penadés Martínez, Inmaculada. 2012. *Gramática y semántica de las locuciones*. Alcalá de Henares: Servicio de Publicaciones de la Universidad de Alcalá.
- Piunno, Valentina. 2018. Negated multiword expressions. Types, properties and lexicalization degrees. In Natalia Filatkina & Sören Stumpf (eds.), *Conventionalization and Variation*, 125–147. Frankfurt am Main: Peter Lang.
- SkE. *Sketch Engine: Corpus query system*. Corpus: esTenTen18. <http://sketchengine.co.uk/> (accessed 17 July 2020)
- Stumpf, Sören. 2016. Modifikation oder Modellbildung? Das ist hier die Frage – Abgrenzungsschwierigkeiten zwischen modifizierten und modellartigen Phrasemen am Beispiel formelhafter (Ir-)Regularitäten. *Linguistische Berichte* 247. 317–342.
- Stutz, Lena & Rita Finkbeiner. In this volume. *Veni, vidi, veggio*. A contrastive corpus linguistic analysis of the phraseological construction *Veni, vidi, X* and its German equivalent *X kam, sah und Y*. In Carmen Mellado Blanco (ed.), *Productive Patterns in Phraseology and Construction Grammar. A multilingual Approach*, 287–314. Berlin & Boston: Mouton de Gruyter.

- Traugott, Elizabeth Closs. 2008. Grammaticalization, constructions and the incremental development of language: Suggestion from the development of Degree Modifiers in English. In Regine Eckardt, Gerhard Jäger & Tonjes Veenstra (eds.), *Variation, Selection, Development: Probing the evolutionary model of language change*, 299–250. Berlin & Boston: Mouton de Gruyter.
- Zeschel, Arne. 2012. *Incipient Productivity: A Construction-Based Approach to Linguistic Creativity*. Berlin & New York: Mouton de Gruyter.
- Ziem, Alexander. 2018. Construction Grammar meets Phraseology: eine Standortbestimmung. *Linguistik Online* 90 (3). 3–19. <https://bop.unibe.ch/linguistik-online/article/view/4316/6450> (accessed 17 July 2020).

Marie Kopřivová

Between phraseology and conversational routines

Using spoken Czech corpora

Abstract: This article focuses on the description of the frequent types of the stereotypical reactions in everyday communication. Such reactions represent certain repeating patterns typical of spontaneous speech. They are not dependent on the specific speaker and represent a strategy by which the conversation partner takes the floor. These formulas are searched for in spoken corpora of Czech using the most frequent bigram *to je* 'it is' further combined with other word classes. Thus, trigrams to tetragrams are searched for which are repeated in a conversation and whose function can be inferred from the context. Some of these structures belong to idioms and have their idiomatic meaning but they are usually open patterns using a wider alternation of lexical elements (e.g. combinations with evaluative adjectives). One of these elements tends to be central, which means that they can be detected using only this method. Another important feature is that the most common form of an n-gram is stable, and other words, such as discourse markers are rarely inserted into it. From this it can be concluded that these reactions can be used as compact units, enabling speakers to react quickly in a conversation and express their support, agreement or to prepare for opposition to their partners.

Keywords: phraseme annotation, Czech corpora, spoken language, conversational routines, idioms.

1 Introduction

Spoken language corpora can be leveraged to identify conversational routines (Coulmas 1981, Bladas 2012) in everyday conversation. In the case of Czech, almost 7 million words of recordings and transcripts of intimate discourse (Clancy 2016) are available across two spoken language corpora, *ORAL* and

Marie Kopřivová, Institute of the Czech National Corpus, Charles University;
marie.koprivova@ff.cuni.cz

<https://doi.org/10.1515/9783110520569-003>

ORTOFON (Komrsková et al. 2017, Lukeš et al. 2015).

Based on the most frequent bigrams, a structure beginning with the words *to je* ‘it is’ was identified in spoken Czech. It can be complemented by a noun (sometimes with a preposition), adjective or adverb, and many of the combinations are frequently used as stereotypical reactions in conversation (e.g. *to je dobrý / jasný / pravda* – ‘it’s all right / clear / true’, i.e. ‘that’s right’).

Some of these reactions are idiomatic in nature, and they are also captured in the *Dictionary of idioms* (Čermák et al. 2009, Kopřivová and Hnátková 2014): *to je jedno* ‘never mind that’, ‘it doesn’t matter’; *to je v háji* ‘we’re screwed’). Others are being creatively innovated by speakers (*to je tak jednoduchý* ‘it’s so simple’). Corpora even make it possible to monitor fluctuations in use, for instance an increase in the popularity of *to je v pohodě* ‘it’s fine’. Looking beyond intimate discourse, these reactions are also often found in fiction and film subtitles.

The structure neuter pronoun *to* + verb *je* is a pattern that is often used in spoken language as a form of reaction (Čermák and Kopřivová 2016); sometimes it is possible to describe it as a discourse marker (Aijmer 1996, Fraser 2006), sometimes as a purely compositional expression, and sometimes as an idiom (Bladas 2012). The present paper will endeavor to delimit the boundaries between these uses. It will also be shown how the lexical content of this structure differs in spoken and written language and in different types of spoken discourse.

2 Data in the Czech corpora

The *Czech National Corpus* contains extensive written corpora (about 4 billion tokens); on the other hand, the amount of spoken language data is small (about 8 million tokens). Nevertheless, some language phenomena are shown to be much more common in the spoken language. The spoken data collection focuses on intimate discourse, which includes the *ORAL* (Kopřivová et al. 2017a, Kopřivová et al. 2017d) and *ORTOFON* (Komrsková et al. 2017, Kopřivová et al. 2017c) corpora. The *DIALOG* corpus (Peterek et al. 2007) contains media interviews, similarly to the monologue corpus *ORATOR* (Kopřivová et al. 2019). We use the representative corpus *SYN2015* (Křen et al. 2015) for comparison with the written language.

2.1 Written corpus *SYN2015*

SYN2015 is a representative corpus of contemporary written Czech. It contains 100 million tokens. The corpus is divided into three groups of equal size: fiction, non-fiction, and newspapers and magazines. These groups are further divided into individual genres (Křen et al. 2015). Multi-word expressions (MWE) are annotated in this corpus. This annotation is based on the *Dictionary of Czech phraseology and idiomatics* (Čermák et al. 2009); therefore, idioms and phrasemes are currently identified in the corpus. Around 4% of words are marked as phraseme components. Other primarily non-idiomatic types, such as terms or time expressions, will be gradually added (Hnátková et al. 2017, Koprřivová 2017b).

The MWE annotation using the *FRANTA* tool (Hnátková et al. 2017) in *SYN2015* is implemented using two attributes. One attribute is the so-called *collocation lemma* (*col_lemma*), which represents the prototypical form of the MWE, i.e. basically a dictionary headword. The other attribute is the so-called *collocation type*, which is further subdivided into two positions. The first position specifies the collocation type proper (e.g. verbal, proverbial, etc.), and the second one identifies the primary vs. secondary element in the given MWE instance. This distinction is mostly technical, and its primary purpose is to allow for distinguishing between the number of occurrences of an MWE as a whole vs. the number of occurrences of all of its components. In each occurrence of an MWE, only one element is identified as primary, the remaining ones being tagged as secondary. An example of an MWE annotation is given in Figure 1.

Idiom: <i>To je v háji</i> . 'It is in the forest.', 'It is gone.'		
Token	collocation lemma	collocation type
<i>To</i>	<i>být v háji</i>	VZ (part of verbal idiom)
<i>je</i>	<i>být v háji</i>	VZ (part of verbal idiom)
<i>v</i>	<i>být v háji</i>	VZ (part of verbal idiom)
<i>háji</i>	<i>být v háji</i>	VH (main part of verbal idiom)

Fig. 1: MWE annotation example (corpus *SYN2015*)

Unfortunately, the annotation is based on the tokenization of the written text and assumes an accent for the written language, and it is therefore not yet possible to use it in the transcript of a spoken corpus. It is necessary to look for other ways to identify other phrasemes.

2.2 Spoken corpora

Currently the largest corpora of spoken Czech are the *ORAL* (Kopřivová et al. 2017a, Lukeš et al. 2015) and *ORTOFON* corpora (Komrsková et al. 2017). Both corpora contain recordings of spontaneous everyday communication. The conversations took place among people who know each other well – they are partners, family members or friends. This is the type of communication that Clancy calls intimate discourse (Clancy 2015). Both corpora were gathered from adult speakers throughout the Czech Republic. Both corpora have been balanced according to four sociolinguistic variables: gender, age, the highest achieved level of education, and childhood region (dialect regions). The *ORTOFON* corpus is more precisely balanced and contains two levels of transcription: orthographic and phonetic. Recordings were made in the years 2012–2019, and contains million words. The *ORAL* corpus is comprised of recordings made in the years 2002–2011, and the overall size of the corpus is 5.4 million words.

The transcriptions of neither corpora take syntactic units into account and use paused punctuation instead. For this reason, it is not possible to use the annotation procedures applied in the written corpus and it is necessary to look for another way of searching for phraseological connections.

3 Methodology

Spoken corpora data coming from an interaction between speakers are compact, with less lexical richness than written texts or lectures. It can therefore be expected that the principles of this type of communication will be reflected in certain lexical patterns. Due to the small-scale nature of much spoken data, it is also possible to analyze frequent phenomena manually in more detail.

The analysis of the most frequent structures will be based on what is known as n-grams (cf. e.g. Foroozan et al. 2018). In spoken corpus, n-grams are understood as a sequence of word forms that do not need to manifest a syntactic or morphological relationship, hence they do not constitute collocations. Certain frequently co-occurring word forms, however, may create collocations, which provide more information about a given text than the individual word forms. N-grams are also commonly used in large-scale untagged data, for example for their checking or for tracking the development of a particular n-gram in use (cf. Michel et al. 2011), in NLP and in machine translation. Various statistical

measures can be used for evaluating n-grams¹ (cf. Pecina 2009). Among the most common bigrams and trigrams in informal spoken discourse are combinations of short words that are conventionally known as particles (cf. e.g. Nekula 1996, 2017, Hoffmannová et al. 2019). For the purpose of this analysis, the word category of some word forms will also be taken into account.

We started from the most frequent words which have some special function in communication (Aijmer 1996, Čermák and Kopřivová 2016). In dialogue, they often represent response types that allow the speaker to gain time to formulate an answer. Discourse markers are often used in these cases (Fraser 2006, Hoffmannová et al. 2019). These are usually short words with the pragmatic function of agreement, affirmations such as *no*, *tak*, *jo*.

The most frequent word in spoken Czech corpora is *to*², and the most frequent bigram is *to je* ‘it is’. It is a combination of the neuter pronoun *it* and the third person of the verb *to be* in the present tense. There are 7,653 instances of this bigram in the *ORTOFON* corpus. The second most frequent bigram *já jsem* ‘I am’ which has 3,726 instances³, followed by the reverse order *je to*, with 3,256 instances. The position of this combination in dialogue is indeed central, regardless of the sociolinguistic characteristics of the speakers.

Next, we focused on trigrams containing the basis *to je*. Three quarters of the occurrences of this bigram are at the beginning of the utterance, so we will look for the continuation of the pattern on the right and search for the POS tag (so-called POS gram, combination with a part of speech). Table 1 shows that trigrams containing the selected parts of speech are several times more frequent in the spoken corpus than in the written one.

1 It is also possible to omit (skip over) a particular component in a sequence, which results in what is known as skip-grams.

2 It is interesting that in the upcoming corpus of monologues (lectures, etc.), the same lexical item as in the written corpus is the most common.

3 The high frequency of this bigram is also interesting because Czech – especially in its written form – does not typically use personal pronouns, as the person is expressed in the verb form. However, in this type of conversation, the first-person pronoun is highly frequent.

Tab. 1: Trigrams beginning with *to je* in written and spoken Czech grouped by POS of the third (SYN2015 and ORAL)

POS	Written SYN2015 (ipm)	Spoken ORAL (ipm)
noun	102	729
adjective	106	1,486
preposition	38	395
adverb	90	1,138

In contrast, written language contains other structures than in the spoken language. The trigram is part of a longer sentence or a sentence that has a relatively complex syntactic structure, as shown in the following examples. Structures similar as in the spoken language occur only in instances of direct speech.

- (1) *myslím, že to je **dobrý** nápad.*
'I think that is a good idea.'
- (2) *to je **pravda**, že je daleko.*
'That is true that he is a long way.'
- (3) *to je v podstatě jediné, o co se v reformě hraje.*
'This is basically the only thing that the reform is about.'

In spoken language, these patterns are much simpler and express conversational routines as a type of response. The most common ones will be listed in the following overview.

4 Frequent patterns with nouns

Let us start with the bigram *to je* combined with adjectives, nouns, and prepositional phrases containing a noun or a pronoun. In written Czech, these trigrams and tetragrams are often further elaborated, whereas in spoken language, they can form standalone collocations, which will be analysed here in more detail and characterised in terms of responses which they represent in dialogue.

The bigram in question combined with adjectives proves to be an open class particularly in terms of a positive or negative “affirmation”. This group includes frequently used prototypes but is at the same time open to “new” adjectives

which may appear with a new generation of speakers, such as the phrase *to je hustý* ‘it is cool’ (lit. ‘dense’).

Nouns as well as adjectives are allowed, in the given pattern with the function of positive or negative evaluation, to be replaced by other nouns. At the same time, it can be observed that the noun *fakt* ‘fact’ is transforming into an intensifying discourse marker, which is evidenced by its occurrence before nouns (14) and is reflected also in the morphological analysis of Czech, which categorises such examples as particles (see 5.3.2).

4.1 Patterns with an adjective

The most common part of speech that the bigram *to je* combines with in spoken corpora is the adjective. However, some occurrences do not form a meaningful trigram but instead continue with a noun. Syntactically, therefore, it is not a verbonominal predicate but a description of a thing (4). In spoken language, it is common for replies not to be contiguous or to be interleaved by different expressions with the function of discourse markers (5, 6; the dots in the transcript indicate pauses).

- (4) *to je nová voňavka*
‘it is a new perfume’
- (5) *vono to sice vypadá že to je veliký a todle ale ... prostě perfektně ovladatelný vidíš z toho krásně*
‘it seems that it is big and that but ... just perfectly manageable you see from it (the car)’
- (6) *to je strašný teďka ta péče*
‘it’s terrible now that care’

Adjectives with evaluative significance are often used, and we can divide them into several groups. This classification can, moreover, be supported by intonation.

4.1.1 Agreement or backchannel

This group is mainly represented by the adjectives *jasný / jasné* ‘clear’. These reactions express agreement with the content of the speech, or the partner in the dialogue notifies the speaker that he or she understood (7). Another common adjective is *možný / možné* ‘possible’. With this adjective, the partner in the dialog expresses agreement with something in which the other speaker showed uncertainty (8) – often expressed by the verb form *myslím* ‘I think’. The difference between these adjectives is also that the adjective *clear* is used alone as an ellipsis like a series of other evaluative adjectives. For the adjective *možný* this does not apply and the adverbial form *možná* is used. It serves as a vague promise with the meaning of ‘perhaps’.

- (7) A: *ted' mi to docela přide vhod*
 B: *to je jasný*
 A: ‘now I find it quite useful’
 B: ‘it’s clear’
- (8) A: *on sa mně zdá dělá tu medovinu*
 B: *jo to je možné já nevím no*
 A: ‘he seems to make this mead’
 B: ‘yeah, it’s possible, I don’t know’

4.1.2 Negative agreement

This group includes the following adjectives with negative connotations and an expressive pragmatic function: *špatný* ‘bad’, *hrozný* ‘horrible’, *blbý* ‘stupid’, *strašný* ‘terrible’, *šílený* ‘crazy’, *drsný* ‘harsh’. These adjectives appear in reactions that imply agreement with the speaker’s assessment of the situation or the negative information. It is therefore the same attitude to the situation but a negative one. It may express dismay. This structure also often appears as a summary after describing a certain negative event, so it may not be the reaction of the other speaker.

- (9) A: *to máš za pozemek úplně neskutečný prachy a ještě ten barák*
 B: *to je šílený*
 A: ‘you have completely unreal money for the land and also the house’
 B: ‘it’s crazy’

4.1.3 Positive surprise or satisfaction (including ironic usage)

The opposite group consists of positive evaluative adjectives: *dobrý* ‘good’, *hezký* ‘nice’, *pěkný* ‘pretty’, *krásný* ‘beautiful’, *super* ‘superb’, *výborný* ‘excellent’, *úžasný* ‘amazing’, *skvělý* ‘great’, *hustý* ‘cool’ (lit. ‘dense’). The most vague meaning can be found in conjunction with the adjective *dobrý* ‘good’, which can express simply that something does not matter to a person (10), or it can be praise, an expression of enthusiasm (11), or, with appropriate intonation, it can express that the described situation is perceived negatively. All the adjectives in this group can convey this ironic meaning. The adjective *hustý* ‘cool’ belongs to modern expressions and its frequency has been increasing since 2005.

(10) *A: tutá to tobě nebo mně?*

B: mně to nevadí to bude zas ten hovor to je dobrý

A: ‘is it ringing for you or me?’

B: ‘it doesn’t matter, it will be the call again, it’s good’

(11) *to je dobrý to je na ... vtip dobrý výborný*

‘it’s good, it’s ... a good joke, excellent’

(12) *to je dobrý když to řeknu já tak se ... urazí*

‘it’s good, when I say it, then ... he’s offended’

4.1.4 Doubts

When the speaker needs to express certain doubts, distance, or mild interest or to indicate a slight disagreement with the dialogue partner (13), he or she can choose structures with the following adjectives: *divný* ‘curious’, *zajímavý* ‘interesting’, *těžký* ‘hard’, *zvláštní* ‘strange’.

(13) *A: uděláme úplně hustou česnečku nemusíme nic jiného mít*

B: to je divné mít jenom polívku ne?

A: ‘we’ll make a totally thick garlic soup, we don’t need to have anything else’

B: ‘it’s strange to have soup only, no?’

4.2 Patterns with a noun

Nominal groups are not so varied. We can divide these structures into similar groups, but with fewer elements. Individual nouns are less frequent. Nominal expressions are less transparent, they are not as evaluative as adjectives but are more often idioms.

4.2.1 Agreement or backchannel

Nominal structures expressing agreement include *fakt* ‘fact’ and *pravda* ‘truth’. In contrast to adjectival structures, they convey consent and generally do not contain further evaluations.

The word *fakt* ‘fact’ can be used alone to ask for a confirmation as to whether the speaker is serious. It also appears in many contexts as an intensifying discourse marker. It thus takes on quite different functions in spoken and written contexts. It also allows for various collocability.

- (14) *to je fakt zajímavý*
‘that’s really interesting’

4.2.2 Disagreement

This group includes mainly nouns: *blbost* ‘bullshit’ and *nesmysl* ‘nonsense’. They express disagreement, which, however, is often mitigated by a supplement, a question (15).

- (15) A: *na tom vydělaly dioptrický clony*
B: *to nevím teda ... to je blbost ne?*
A: ‘they made profit on dioptric screens’
B: ‘I don’t know, really, it’s bullshit, no?’

4.2.3 Negative agreement

This primarily includes the following nouns: *škoda* ‘pity’, *hrůza* ‘horror’, *konec* ‘end’. The noun *škoda* serves to express regret (16). The repertoire of nouns is much more limited than that of adjectives.

- (16) A: *já možná v sobotu pojedu do Brna*
 B: *tak to je škoda že tam zrovna nespím*
 A: 'I may go to Brno on Saturday'
 B: 'so it's a pity I'm not sleeping there then'

4.2.4 Positive surprise or satisfaction (including ironic usage)

There are some fluctuations among nouns expressing a positive reaction. These are the nouns *síla* 'powerful' and *paráda* 'awesome', whose frequency has been increasing in recent years, and *mazec* 'blast', which is widely used by speakers from Moravia and Silesia (eastern parts of the Czech Republic).

4.2.5 Doubts

Even among nouns there are formulations expressing hesitation or deliberation, which are mainly *otázka* 'question' and *problém* 'problem'.

4.3 Patterns with a preposition and a (pro)noun

There are a lot of combinations with different nouns, but these structures are less frequent. Table 2 shows the ipm frequency in both spoken corpora. It is clear that this frequency does not change in the newer data. The most common are patterns with the preposition *v* 'in', which are associated with place and time expressions. As with nouns without prepositions, idioms are more common.

Tab. 2: Trigrams starting with *to je* in spoken Czech

Preposition	ORAL (ipm)	ORTOFON (ipm)
<i>v</i> 'in'	99	107
<i>na</i> 'for'	82	93
<i>za</i> 'for'	48	55
<i>z</i> 'from'	33	32
<i>pro</i> 'for'	23	24

Preposition	ORAL (ipm)	ORTOFON (ipm)
<i>od</i> 'from'	21	16
<i>o</i> 'about'	20	23

4.3.1 Patterns with the preposition *v* 'in'

For patterns with the preposition *v* 'in', we can recognise the following groups:

1. Positive expressions: *v pohodě* 'OK', *v pořádku* 'fine', *v klidu / klídku* 'peace'
2. Negative (very expressive, some offensive): *v prdeli* lit. 'ass', *v pytlí* lit. 'bag', *v háji* lit. 'forest', *v Prčicích* (name of a town), *v hajzlu* lit. 'toilet'
3. Expletives: *v podstatě* 'basically'

These terms are mostly typical of the spoken language, except for the expletive expression *v podstatě* 'basically'. It is very often used in journalism, from where it probably spread into the spoken language.

A very interesting development can be observed also in the expression *v pohodě* 'OK'. It has a very vague meaning, as was the case with the adjective *dobrý* 'good'. In the following examples we see that it can be used as an answer to the effect that something does not matter (17), that something is good (18) or that there is no problem with a certain person or that he or she is healthy (19).

- (17) A: *mám zavřít dveře?*
 B: *to je v pohodě*
 A: 'Should I close the door?'
 B: 'It's OK'

- (18) *parkování je v pohodě*
 'Parking is OK'

- (19) *on je v pohodě*
 'He is OK / fine / healthy'

4.3.2 Patterns with the preposition *na* 'for'

Patterns containing the preposition *na* 'for' involve mostly phrases that convey a negative response in the following forms:

1. refusal or straightforward condemnation (*to je na houby / pytel / zabítí* ‘it sucks’), some including taboo words (*to je na hovno / prd* ‘it is shit’), also including patterns with a pronominal expression (*to je na nic* ‘it is no good’);
2. confusion, evaluation of the statement as illogical (*to je na hlavu / palici* ‘it is crazy’).

Furthermore, there are instances of the collocations *to je na prodej* ‘it is for sale’ and *to je na chuť* ‘it is something to whet the appetite’. This category includes numerous examples involving pronouns, e.g. *to je na ní / mně* ‘it is up to her / me’, which conveys the meaning that a given decision or choice is within the competences of said person. This group also includes longer collocations that express a lack of understanding (*to je na mě moc složitý* ‘it is too complicated for me’) or a refusal of something offered (*to je na mě [moc] silný* ‘it is [too] strong for me’). This collocation can be further complemented by additional adjectives.

4.3.3 Patterns with the preposition *za* ‘for’

The pattern with the preposition *za* ‘for’ typically has one more constituent on the right. This type expresses a question characteristic of spoken language, e.g. *co to je za den?* ‘which day is it?’, is expressed in written language without the preposition: *který den to je?*

This idiomatic pattern can be used to convey disapproval, *co to je za blbost?* ‘what is this bullshit?’, or a degree of compassion or pity, as represented by the phrase *co to je za život* ‘what kind of life is it?’ (20).

- (20) *když už byla dlouho upoutaná jenom na lůžko tak . co to je za život že jo*
 ‘when she was bedridden for so long . what kind of life is it, right?’

4.3.4 Patterns with the prepositions *z* ‘from’ and *pro* ‘for’

Tetragrams containing the preposition *z* ‘from’ commonly serve to describe material (*to je ze dřeva* ‘it is made of wood’) or to specify time (*to je z roku 2006* ‘it is from the year 2006’). Only one phraseme was found: *to je ze života* ‘it is life’.

The preposition *pro* most frequently occurs in the n-gram *to je pro mě* ‘it is for me’, which serves to confirm whether what the speaker is offered is actually meant for him or her. However, when complemented by an adjective (*to je pro mě důležitý* ‘it is important for me’) or an adverb and an adjective (*to je pro mě úplně nepochopitelný* ‘it is totally incomprehensible for me’), this n-gram introduces the speaker’s own opinion or feeling. This response pattern is similar to the pattern *to je* + adjective (4.1) but the personal opinion is more emphasised here.

An overview of the most frequent combinations of the bigram *to je* with a preposition makes it apparent that there is a higher number of phrasemes here than there was with adjectives and nouns, which means that the patterns are not as open. Given the low frequency of occurrence though, it cannot be ruled out that a more extensive data set would produce different results.

5 Combinations of the bigram *to je* ‘it is’ with adverbs, pronouns, and particles

Adverbs and particles in Czech are classified as synsemantic word categories, whereas pronouns are on the borderline between autosemantic and synsemantic (cf. Cvrček et al. 2010: 210). Pronouns can formally stand for adjectives and nouns. Spoken language manifests a high frequency of the demonstrative pronoun *ten* ‘that’, which often refers to entities outside context (Osolsobě 2017) and is also used as a determiner with nouns (Cvrček et al. 2010: 142 and 218). The pronoun *nějaký* ‘some’ can similarly be used in the function of an indefinite article, though various combinations are possible (cf. Uhlířová 1992). As to spoken discourse, it is significant that these word categories can occur in some of the above-mentioned patterns (cf. e.g. Schneiderová 1993): *to je úplně v pohodě* ‘it is totally cool’, *to je nějaký hrozně dlouhý* ‘it is somehow too long’, *to je prostě skvělý* ‘it is just great’.

5.1 Trigrams with pronouns

Pronouns are the third most frequent word class that complements the bigram *to je*. This is mainly due to the most frequent word form *to* ‘it’. Some pronouns can be inflected as nouns; a pronoun of this type typically stands for nouns and has been mentioned in the section devoted to prepositional phrases (see 4.3).

Other pronouns have the character of adjectives; in section 5.1.3, we will focus mainly on pronouns of this type. A special case is the demonstrative pronoun *ten, ta, to* ‘that’ (see 5.1.1, cf. Kolářová 2014).

5.1.1 Trigram *to je to* ‘it is it’

The most frequent trigram is *to je to* ‘it is it’. Here is where the common feature of this type of spoken discourse is manifested, that is, repetition of words or entire parts of utterances. The reason for this repetition can be the speaker’s attempt to continue speaking or to resume speaking after an interruption by another person, even when it means repeating what has already been said. The pronoun *to* ‘it’ exercises various other functions in spoken discourse, such as expletives (*to je to je v normálním takový jako taková jako lahvička no* ‘it is it is in a regular like in like a flask’), differentiating (*to je to co jste dostali k Vánocům?* ‘it is what you got for Christmas?’), or emotional charge (*to je to lehoučký že jo?* ‘it is this light, see?’, *fuj to je to hnusný* ‘ew, it is this gross thing’; cf. Klimešová et al. 2015, Schneiderová 1992), although these are beyond the scope of this article.

The trigram *to je to* ‘this is it’ can also form a standalone collocation, which is used to express a strong agreement and a conviction that the other speaker just said something essential, something that is the point of the conversation. Without the corresponding intonation though, this trigram can serve simply as a cataphoric reference.

5.1.2 Trigram *to je něco* ‘it is something’

Another pronoun that forms collocations is *něco* ‘something’, most often in the n-gram *to je něco jinýho* ‘it is something else’, which is used to express that two situations are not comparable or, when uttered with a particular intonation, conveys an approval of the option offered (also as *to je jiná* ‘it is different’). This trigram can also function as a standalone collocation that expresses either a positive or negative surprise.

(21) *Kristina: článek v Blesku je vyložená lež*

Květa: to je něco . jedny noviny napíšíou to a v druhých se píše že to vodsuzuje vid’

Kristina: ‘the article in the Blesk is an outright lie’

Květa: ‘that is something . one newspaper writes this and another writes that it is not true’

The pronoun can also be complemented by evaluating adjectives, usually negative (*to je něco strašného / hrozného* ‘it is something terrible / horrible’) but also positive (*to je něco skvostného* ‘it is something magnificent’); the adverb *tak* ‘so’ can be used for extra emphasis (*to je něco tak hnusného* ‘it is something so nasty’).

The pronoun also functions in comparative structures: *to je něco podobného* ‘it is something similar’, *to je něco jako* ‘it is something like’.

- (22) *to je něco jako brufen jo? že utlumí bolest*
 ‘it is something like brufen, right? that it alleviates pain’

5.1.3 Trigrams *to je takový* ‘it is such’, *to je nějaký* ‘it is some’

Another relatively frequent pronoun is *takový* ‘such’, which serves to provide explanations: *střevlík piluna to je takovej jako velkej brouk černej* ‘the ground beetle it is something like a big black beetle’. A common collocation involves the word *jako* ‘like’⁴, which relativises the explanation. A similar but less frequent structure is *to je nějaký jako* ‘it is something like’, where the indefinite pronoun *nějaký / nějaký* ‘some’ weakens the statement. It serves as a weakening element inserted in the collocation: *to je takový blbý* ‘it is somewhat stupid’.

- (23) *to je takový blbý ne když zařizujete byt že ani nemůžete nikam jet . nebo ti to nechybí?*
 ‘it is somewhat stupid, right, when you’re furnishing a flat and can’t go anywhere . or you don’t miss it?’

The most common combination with the pronoun *nějaký* ‘some’ is the collocation *to je nějaký divný* ‘it is somewhat weird’, which expresses mistrust or suspicion of the speaker in response to a statement or a situation.

- (24) *tak jsme jeli dál . za chvíli tam byla šipka na zoo říkám ... to je nějaký divný ne? . to už su někde úplně jinde*

⁴ The word *jako* ‘like’ is used in a comparative function but it is a common expletive in spoken discourse (cf. Hoffmannová et al. 2019).

‘so we drove on . in a while there was an arrow to the zoo, so I say ... it is somewhat weird, no . I’m already somewhere else’

It was not confirmed in the most frequent n-grams with pronouns that pronouns could be arbitrarily inserted in n-grams with adjectives and nouns except for emphasizing or weakening the meaning with the use of the pronouns *něco*, *nějaký*, *takový* ‘something’, ‘some’, ‘such’. The pronoun *něco* is rather an intensifying variant of the phraseme *to je něco* ‘it is something’. With the pronouns *nějaký* and *takový*, a certain weakening of the meaning of the collocation can be considered.

5.2 Trigrams with adverbs

Adverbs are the second most common word class that complements the bigram *to je*. In Czech, however, adverbs typically modify verbs or adjectives, especially in the verbonominal predicate, which is suggested by the verb *být* ‘be’ in the bigram. Collocations containing a standalone adverb are rare, e.g. *to je dobře* ‘it is well’, *to je dost* ‘it is enough’, *to je moc* ‘it is too much’, *to je fajn* ‘it is nice’, *to je akorát* ‘it is just right’.

5.2.1 Trigrams with the adverb *úplně* ‘totally’

The most frequent adverb is the word form *úplně* ‘totally’⁵. The most common n-gram containing this adverb is *to je úplně jedno* ‘it does not matter’, which also exists without the emphasizing adverb and expresses that said fact does not need to be taken into account (25).

(25) *to je úplně jedno prosím tě . to je fakt úplně jedno, si toho nevšímej*
‘it doesn’t matter at all . it really doesn’t matter, don’t mind it’

The adverb *úplně* serves as an intensifier also with prepositional phrasemes, e.g. *to je úplně na houby* ‘it totally sucks’, *to je úplně v pohodě* ‘it is totally okay’, *to je úplně něco jinýho* ‘it is something totally different’, *úplně super* ‘totally awesome’. As a discourse marker, the adverb occurs in collocations with other

⁵ The reduced pronunciation of the word *úplně* supports its shift from the adverbial function to that of a discourse marker.

empty discourse markers, e.g. *úplně jako vyřešený* ‘totally like done’, *úplně jako strašný* ‘totally like awful’.

(26) *to je úplně jako automatický že jeden den v týdnu buď sobota nebo neděle se de dělat*

‘it is totally like a matter of course that one day of the week, either Saturday or Sunday, you go to work’

5.2.2 Trigrams with the adverb *taky* ‘also’

Another frequent adverb is *taky* ‘also’⁶. This adverb is most commonly found inserted in collocations with adjectives or nouns. It can express a weakened affirmation: *to je taky pravda / fakt / možný* ‘it is also true / a fact / possible’. In the collocation *to je taky dobrý* ‘it is also good’, the adverb serves as a continuator that provides a response to similar examples mentioned in the conversation, more often used ironically. As a continuator, it can combine with other discourse markers, e.g. *to je taky docela hezký* ‘it is also quite nice’.

5.2.3 Trigrams with the adverb *tak* ‘so’

The adverb *tak* ‘so’ is most commonly found in the collocation *to je tak jako* ‘it is so that’, which is complemented either by an adjective (*to je tak jako ohrazená ta zahrada* ‘it is so that like the garden is walled’) or a conjunction (*ale spíš to je tak jako . že tam byl boj vo ty turisty* ‘but it is so that like . there was a struggle for the tourists’), where the discourse marker *jako* ‘like’ serves to weaken the cleft structure with *tak* ‘so’. Much more common is the occurrence of the adverb *tak* on the left of the bigram, where it can serve as a continuator – it continues the previous utterance and can also begin a new utterance (*tak to je dobře* ‘so it is well’). It often occurs before collocations with nouns and adjectives as an emphatic element (*tak to je hustý* ‘so this is cool’).

⁶ The word *taky* takes the form of *také* in written Czech, the same as in some dialects, where it is however a form of the pronoun *takový* ‘such’: *to je také pěkně vyšívané (takové pěkně vyšívané* ‘such nice embroidery’). In spoken language it is a discourse marker used as a continuator.

5.2.4 Trigrams with the adverb *docela* ‘quite’

The adverb *docela* ‘quite’ again serves as a discourse marker with a weakening function. The most frequent collocation *to je docela dobrý* ‘it is quite good’ however typically conveys a positive evaluation of the subject of discussion. The collocation *to je dobrý* is in itself very vague and usually expresses irony, that is, a negative evaluation of the situation described in the previous utterance. It also has the meaning of ‘it is enough’ and is used idiomatically to express that the other person can keep the change. Paradoxically, the weakened construction *to je docela dobrý* ‘it is quite good’ is used rather as a positive evaluation. When used with other adjectives, the function remains weakening (*to je docela drsný* ‘it is quite harsh’, *to je docela hezký* ‘it is quite nice’).

Another collocation with the adverb *docela* is the n-gram *to je docela dost* ‘it is quite a lot’. It occurs with the following adjective (*to je docela dost hnusný* ‘it is quite gross’), adverb (*to je docela dost daleko* ‘it is quite far’), noun (*to je docela dost prachů* ‘it is quite a lot of money’) or it stands on its own (*tady za to parkoviště to je docela dost* ‘it is quite a lot for this parking’ – it is expensive), which usually refers to an amount, typically money and the like.

An analysis of the most frequent adverbs demonstrates that with a few exceptions, the adverbs in collocations serve as intensifiers or continuators, otherwise also as discourse markers (cf. Hoffmannová et al. 2019). Their function in collocations is similar to their general function in informal communication, which means that they are not a fixed component of the collocation but they can modify its meaning.

5.3 Trigrams with particles

Particles rank as the sixth most common word class that complements the bigram *to je*. It should be noted however that particles are a problematic word category and are tagged in spoken language in the same way as in written language (cf. Koprřivová et al. 2017a). The difficulty of this classification was noted in the previous section concerned with adverbs (5.2), which should be more properly classified as particles, given their function as discourse markers.

5.3.1 Trigrams with the particle *prostě* ‘just’

The particle *prostě* ‘just’ is the most common particle that complements the bigram *to je*. It is the fourth most frequent particle in the spoken corpora *ORAL* and *ORTOFON*, following the short words *no* ‘well’, *jo* ‘yeah’, *ne* ‘no’. This particle originates from an adverb which is scarcely used any more in spoken language. Context shows that it usually co-occurs with other particles in the function of a discourse marker: *jenom* ‘only’, *jako* ‘like’, *úplně* ‘totally’ (*to je prostě jenom fakt číslo tabulky* ‘it is really like just a table number’). This indicates an expletive and weakening function. A common particle with a weakening function is *asi* ‘maybe’.

5.3.2 Trigrams with the particle *fakt* ‘really’

The second most common particle is the word *fakt* ‘really’, which has appears among nouns. This word shifts towards being a particle in spoken language. It is classified as a particle when it occurs before a noun (*to je fakt pravda* ‘it is really true’) or adjective (*to je fakt neskutečný* ‘it is really impossible’). Here it serves as an intensifier.

5.3.3 Trigrams with the particle *právě* ‘exactly’

The combination of the particle *právě* ‘exactly’ with the bigram *to je* serves to convey emphasis. Furthermore, it co-occurs with the pronoun *ten* ‘that’ (*to je právě ten problém* ‘it is exactly the problem’). It is used to assure the speaker that he or she mentioned exactly what the listener considers important in the conversation. The most common response takes the form of phrasemes: *to je právě vono* ‘it is exactly that’.

A brief overview of those particles most frequently occurring with the bigram under examination makes it clear that particles function in n-grams similarly as adverbs and hence it would be convenient to set this group apart as discourse markers, as their function in spoken language is pragmatical, which is not traditionally involved in describing parts of speech.

6 Additional analysis

6.1 Inverse order

We have so far analysed the bigram *to je* in the position at the beginning of an utterance, response, or reaction, examining its collocations on the right. Moving our focus to the collocations on the left, we will find that the sentence-initial position is indeed the most frequent, followed by instances when the bigram occurs after a conjunction, particle, and adverb, respectively. These are mostly continuators and discourse markers expressing backchanneling, which often occur at the beginning of sentences, such as: *no* ‘well’, *jo* ‘yeah’, *ne* ‘no’, *vždyt’* ‘but’, *prostě* ‘just’, *ano* ‘yes’ (*prostě to je jedno* ‘it just doesn’t matter’), etc. They often involve agreement, which is expressed also by the following trigram, particularly with a noun (see 4.2.1), adjective (4.1.3), or disagreement (see 4.1.2 and 4.2.2).

The same group also includes common adverbs preceding the bigram *to je*: *tak* ‘so’ (cf. 5.2.3), *teda* ‘actually’, *taky* ‘also’, *takhle* ‘like this’, *vlastně* ‘actually’ (*jo vlastně to je národní park* ‘yeah, it is actually a national park’). Also common are interrogative adverbs: *jak* ‘how’, *kde* ‘where’, *kdy* ‘when’ (*kde to je ten Yukon?* ‘where is it, this Yukon?’); and deictic adverbs: *už* ‘already’, *tady* ‘here’, *tam* ‘there’, *ted’* ‘now’. These expressions tend to cumulate at the beginning of a sentence, and as has been shown in the previous text (e.g. 5.3.1), they occur also on the right of the collocation. For more details on some of them, see Hoffmannová et al. (2019).

6.2 Other verb tenses

Considering that the subject of the analysis here involves a verb, it is also necessary to look at its tenses. Table 3 shows a comparison of the tenses in both spoken corpora : the future tense is expressed with the bigram *to bude* ‘it will be’; in the past tense, the bigram divides into the three grammatical genders, reflecting the gender of the noun in the ending of the verb: *to byl*, *to byla*, *to bylo* ‘it was’. A comparison of the frequency of occurrences shows that the collocation in the present tense is the most common and essential in conversation. Even though some collocations take the form, for example, of a consolation oriented towards the future (*to bude dobrý* ‘it will be okay’) or the past (*to byla hrůza* ‘it was a horror’, *to byl problém* ‘it was a problem’), these are not identical

with the most frequent types discussed above and usually occur in the utterance of one person, not as a response by the listener.

Tab. 3: Verb tenses of *to je* in spoken Czech

Verb tense	Verb form	ORAL (ipm)	ORTOFON (ipm)
Present	<i>to je</i> 'it is'	6,863	6,189
Future	<i>to bude</i> 'it will'	580	634
Past	<i>to byl</i> 'it was'	339	275
	<i>to byla</i> 'it was'	295	252
	<i>to bylo</i> 'it was'	1,550	1,472

7 Conclusions

The analysis has demonstrated the presence of conversational routines involving the bigram *to je* in informal dialogues of native Czech speakers. This collocation is central when it comes to responses of the listener. Certain responses qualify as phrasemes with a fixed form, others create open patterns which can be filled with a considerably large group of components that can be modified and expanded. This is particularly the case of combinations with nouns and adjectives, which express a positive and negative affirmation or evaluation. Combinations with particles and adverbs, by contrast, manifest different characteristics in that they qualify as discourse markers, whose position and pragmatic meaning are highly changeable and must be observed in the context of all spoken utterances.

By analyzing the most common n-grams in spoken corpora, we arrived at patterns typical for spoken language. These are mainly expressions conveying reactions and attitudes to the facts mentioned in the conversation. These patterns are characterised by the occurrence of specific word categories, which manifest a significant evaluative function in the case of adjectives. Adjectives are also the most common and diverse word class in spoken corpora texts. At the same time, they have the greatest ability to spread.

The motivation of nominal expressions, as to why they were chosen for an evaluative function, is often vague. There is a large number of idioms in these groups. These patterns are less common and there are more expressive structures among them. Using corpora, we were also able to observe the changes in

the function of some expressions over time. This issue deserves a more thorough research on various types of spoken data, which, however, remain unavailable at present.

We divided the most frequent expressions into several semantic groups. These were primarily positive and negative attitudes to the facts that the conversation partner spoke about, and expressions of surprise and hesitation. There are also expressions that are used to keep the conversation going and to give the speaker a chance to prepare an answer. Furthermore, this analysis has shown that in the spoken language we encounter different types of idioms than in the written language, and that the meaning of these expressions is sometimes difficult to describe.

References

- Aijmer, Karin. 1996. *Conversational Routines in English: Convention and Creativity*. London: Routledge.
- Bladas, Óscar. 2012. Conversational routines, formulaic language and subjectification. *Journal of Pragmatics* 44. 929–957.
- Clancy, Brian. 2015. *Investigating Intimate Discourse: Exploring the spoken interaction of families, couples and friends*. London.
- Coulmas, Florian (ed.). 1981. *Conversational Routine. Explorations in Standardized Communication Situations and Prepatterned Speech*. The Hague: Mouton.
- Cvrček, Václav, Vilém Kodýtek, Marie Kopřivová, Dominika Kovářiková, Petr Sgall, Michal Šulc, Jan Táborský, Jan Volín & Martina Waclawičová. 2010. *Mluvnice současné češtiny 1. Jak se píše a jak se mluví* [Grammar of contemporary Czech 1. How to write and how to speak]. Praha: Karolinum.
- Čermák, František, Jan Holub, Renata Blatná & Marie Kopřivová. 2009. *Slovník české frazeologie a idiomatiky IV* [Dictionary of Czech phraseology]. Praha: Leda.
- Čermák, František & Marie Kopřivová. 2016. Idioms in Spoken Corpus. A Sample of Czech Data. In Gloria Corpas Pastor (ed.). *Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives*, 250–261. Geneva: Editions Tradulex.
- Forozan, Yazdani Sepideh, Zhiyuan Tan, Mohsen Kakavand & Aida Mustapha. 2018. Ngram-POS: a bigram-based linguistic and statistical feature process model for unstructured text classification. *Wireless Networks* (2018). <https://doi.org/10.1007/s11276-018-01909-0> (accessed 16 July 2020).
- Fraser, Bruce. 2006. Towards a theory of discourse markers. In Kerstin Fischer (ed.), *Approaches to discourse particles*, 1–20. Amsterdam: Elsevier.
- Hnátková, Milena, Tomáš Jelínek, Marie Kopřivová, Vladimír Petkevič, Alexandr Rosen & Hana Skoumalová. 2017. Eye of a Needle in a Haystack. In Ruslan Mitkov (ed.), *Computational and corpus-Based Phraseology. Second International conference, EuroPhras 2017 London, UK, November 13–14, Proceedings*, 160–171. Chaim: Springer international publishing.

- Hoffmanová, Jana, Jiří Homoláč & Kamila Mrázková. 2019. *Syntax mluvené češtiny* [Syntax of spoken Czech]. Praha: Academia.
- Klímešová, Petra, Zuzana Komrsková, Marie Kopřivová & David Lukeš. 2015. Slovo to v mluvených korpusech ČNK, jeho prefixace a reduplikace [The Czech word to in the spoken corpora of the CNC: prefixation and reduplication]. *Časopis pro moderní filologii* 97. 21–30.
- Kolářová, Ivana. 2014. *Význam a konkurence vybraných druhů odkazovacích zájmen v kontextu* [Meaning and concurrence of some demonstrative pronouns in context]. Brno: Masarykova univerzita.
- Komrsková, Zuzana, Marie Kopřivová, David Lukeš, Petra Poukarová, & Hana Golářová. 2017. New spoken corpora of Czech: *ORTOFON* and *DIALEKT*. *Jazykovedný časopis* [Linguistic journal] 68 (2). 219–228.
- Kopřivová, Marie & Milena Hnátková. 2014. From Dictionary to Corpus. In Vida Jesenšek & Peter Grzybek (eds.), *Phraseology in Dictionaries and Corpora*, 155–168. Maribor: Filozofska fakulteta Maribor.
- Kopřivová, Marie, David Lukeš, Zuzana Komrsková & Petra Poukarová. 2017a. Korpus *ORAL*: sestavení, lematizace a morfologické značkování [*ORAL* corpus: building, lemmatization and morphological tagging]. *Korpus – Gramatika – Axiologie* 15. 47–67.
- Kopřivová, Marie. 2017b. Contribution Towards a Corpus-Based Phraseology Minimum. In Ruslan Mitkov (ed.), *Computational and Corpus-Based Phraseology*, 220–231. Cham: Springer international publishing.
- Kopřivová, Marie, Zuzana Komrsková, David Lukeš, Petra Poukarová & Marie Škarpová. 2017c. *ORTOFON: korpus neformální mluvené češtiny s víceúrovňovým přepisem, verze 1 z 2. 6. 2017* [ORTOFON: corpus of informal spoken Czech with multi-level transcription, version 1 from 6/2/2017]. Praha: Ústav Českého národního korpusu FF UK. <http://www.korpus.cz>
- Kopřivová, Marie, David Lukeš, Zuzana Komrsková, Petra Poukarová, Martina Waclawicová, Lucie Benešová & Michal Křen. 2017d. *ORAL: korpus neformální mluvené češtiny, verze 1 z 2. 6. 2017* [ORAL: corpus of informal spoken Czech, version 1 from 6/2/2017]. Praha: Ústav Českého národního korpusu FF UK. <http://www.korpus.cz>
- Kopřivová, Marie, Zuzana Laubeová, David Lukeš, Petra Poukarová. 2019: *ORATOR: Korpus monologů, verze 1 z 30. 12. 2019* [ORATOR: corpus of monologues, version 1 from 12/30/2019]. Praha: Ústav Českého národního korpusu FF UK. <https://www.korpus.cz>
- Křen, Michal, Václav Cvrček, Tomáš Čapka, Anna Čermáková, Milena Hnátková, Lucie Chlumská, Tomáš Jelínek, Dominika Kovářiková, Václav Petkevič, Pavel Procházka, Hana Skoumalová, Michal Škrabal, Petr Truneček, Pavel Vondříčka & Adrian Zasina. 2015. *SYN2015: reprezentativní korpus psané češtiny* [SYN2015: representative corpus of written Czech]. Praha: Ústav Českého národního korpusu FF UK. <http://www.korpus.cz>
- Lukeš, David, Petra Klímešová, Zuzana Komrsková, Marie Kopřivová. 2015. Experimental Tagging of the ORAL Series Corpora: Insights on Using a Stochastic Tagger. In Pavel Král & Václav Matoušek (eds.), *Text, Speech, and Dialogue: 18th International Conference, TSD 2015, Pilsen, Czech Republic, September 14–17, 2015, Proceedings (Lecture Notes in Computer Science)*, 342–350. Chaim: Springer international publishing.
- Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew Gray, Josef Picket, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak & Erez Lieberman Aiden. 2011: Quantitative Analysis of Culture Using Millions of Digitized Books. *Science* 331. 176–182.

- Nekula, Marek. 1996. *System der Partikeln im Deutschen und Tschechischen. Unter besonderer Berücksichtigung der Abtönungspartikeln*. [Particle Systems in German and Czech. With special reference to modal particles]. Tübingen: Niemeyer.
- Nekula, Marek. 2017. Částice. [Particles] In Petr Karlík, Marek Nekula & Jana Pleskalová (eds.) *Czech Ency – Nový encyklopedický slovník češtiny*. Praha: Nakladatelství Lidové noviny.
- Osolobě, Klára. 2017. Ukazovací zájmeno [Demonstrative pronomen] In Petr Karlík, Marek Nekula & Jana Pleskalová (eds.), *Czech Ency – Nový encyklopedický slovník češtiny*. Praha: Nakladatelství Lidové noviny.
- Pecina, Pavel (2009): *Lexical association measures: Collocation extraction. Studies in computational and Theoretical Linguistics*. Praha: Institute of Formal and Applied Linguistics.
- Peterek, Nino, Petr Kaderka, Zdena Svobodová, Eva Havlová, Martin Havlík, Jana Klímová & Patricie Kubáčková. 2007. Digitisation and Automatic Alignment of the *DIALOG* Corpus: A Prosodically Annotated Corpus of Czech Television Debates. In Václav Matoušek & Pavel Mautner (eds.), *Text, Speech and Dialogue: 10th International Conference, TSD 2007, Pilsen, Czech Republic, September 3–7, 2007: Proceedings*, 607–612. Berlin & Heidelberg: Springer international publishing.
- Schneiderová, Eva. 1993. K užívání zájmena ten (v přívlastkové pozici) v mluvených projevech [The use of the pronoun ten (in the attributive position) in spoken language]. *Naše řeč* 76. 31–37.
- Uhlířová, Ludmila. 1992. Ten nějaký / nějaký ten a případy podobné [The one / some one and similar cases]. *Naše řeč* 75. 247–257.

Fabio Mollica, Sören Stumpf

Families of constructions in German

A corpus-based study of constructional phrasemes with the pattern [X_{NP} attribute]

Abstract: This paper focuses on morphosyntactically and semantically-pragmatically similar constructional phrasemes (Dobrovolskij 2011) that can be described as a family of constructions within the framework of Construction Grammar. These are constructional phrasemes in which a lexically open noun phrase is specified by a lexically fixed postponed attribute. The constructions have the syntactic pattern [X_{NP} attribute] and share more or less the same semantics and pragmatics. As an example, three members of the family are characterized by their formal structure and their semantic-pragmatic properties ([X_{NP} *pur*], [X_{NP} *par excellence*], [X_{NP} *wie er/sie/es im Buche steht*]). The aim is to describe these similar constructions with regard to their network-like relationships. The article aims to determine usage restrictions and preferred fillers on the basis of the *German Reference Corpus (DeReKo)* and with the help of the tool *Lexical Pattern Analyzer (lexpan)*, which permits an automatic quantification of the lexical fillers.

Keywords: Construction Grammar, corpus analysis, constructional phrasemes, schematic idioms, families of constructions.

1 Introduction

A look at current Construction Grammar research shows that one of the most important and cross-theoretical questions is “how speakers’ knowledge of language is organized in the construct-i-con” (Hilpert 2014: 71) as “a large repository of form-meaning pairs” (Hilpert 2014: 57). The important point is that the

Note: This article is the result of close collaboration between the two authors; however, for academic purposes only, Fabio Mollica is responsible for Sections 1, 2 and 5, and Sören Stumpf for Sections 3 and 4.

Fabio Mollica, University of Milan, Fabio.Mollica@unimi.it
Sören Stumpf, Trier University, stumpf@uni-trier.de

<https://doi.org/10.1515/9783110520569-004>

constructicon “is not a flat list or even an unordered ‘bag of constructions’, but instead a highly structured, hierarchical network in which constructions are interlinked” (Hilpert 2014: 57). Within the framework of Construction Grammar, related constructions which share more or less the same formal as well as semantic-pragmatic characteristics are considered a family of constructions (Ruiz de Mendoza, Luzondo Oyón, and Pérez Sobrino 2017: 2).

However, with regard to German, until now there have been only a few studies that undertake a systematic analysis of families of constructions within the framework of Construction Grammar. This is due to the fact that most of these studies restrict themselves to the description of individual constructions and rarely extend the investigation to related phenomena. However, the longer-term goal should be to derive the systematic characteristics of the linguistic system from the analysis of individual constructions (Stefanowitsch 2006: 153). If we look at related constructions, we see that the focus in the area of grammar is particularly on the description of construction networks within related argument structure constructions (cf. e.g. Engelberg et al. 2011; Proost 2015, 2017; De Knop and Mollica 2017). Only a few studies are devoted to the analysis of phraseological families of constructions (cf. e.g. Staffeldt 2011; Auer 2016; Mollica 2020).

This article deals with a complex family of constructional phrasemes (Dobrovolskij 2011), which consists of a large number of partially lexicalized phrasemes. The constructions analyzed have the syntactic commonality that a noun phrase is specified by a postponed attribute. They are all based on the abstract pattern [X_{NP} attribute]. The examples 1–9 show concrete realizations of these individual constructional phrasemes¹:

- (1) *Die Grünen hätten mit ihrem Verhalten in der vergangenen Woche in Deutschland „Politikverdrossenheit pur“ herbeigeführt, und „hängen an der Macht bis zur Selbstaufgabe“.*

‘With their behavior in the past week the Greens had brought about “pure disenchantment with politics” in Germany, and “they are clinging to power to the point of abandoning their ideals”.’

(Nürnberg Nachrichten, 19 November 2001)

- (2) *Seitdem entfaltet sich vor der atemlosen spanischen Öffentlichkeit eine **Justiz- und Politikpötte sondergleichen**.*

¹ The list is certainly not complete. It is not difficult to find other, similar constructions (e.g. [X_{NP} mit Leib und Seele], [X_{NP} wie im/aus dem Bilderbuch], [X_{NP} ohne Ende]).

‘Since then, an unheard of judicial and political farce has been unfolding before the breathless Spanish public.’

(Die Zeit, 27 January 1995)

- (3) *Wenn es um Sex-Appeal geht, sind Sahra Wagenknecht (Linke) und Jürgen Trittin (Grüne) das **Politiker-Traumpaar schlechthin**.*

‘When it comes to sex appeal, Sahra Wagenknecht (The Left) and Jürgen Trittin (Greens) are undoubtedly the perfect political couple.’

(Hamburger Morgenpost, 23 July 2013)

- (4) *Perikles war ein **Politik-Maestro par excellence**.*

‘Pericles was a political maestro par excellence.’

(Die Südostschweiz, 14 December 2011)

- (5) *Es hat somit eine gewisse regulatorische Zentralisierung, aber keine administrative Zentralisierung stattgefunden. Dies ist „**Politikverflechtung**“ in **Reinkultur**.*

‘There has thus been a certain amount of regulatory centralization but no administrative centralization. This is “political interdependence” in its purest form.’

(Neue Zürcher Zeitung, 7 January 2015)

- (6) *Dass der Staat in schlechten Zeiten mit Konjunkturprogrammen die Wirtschaft unterstützen und die Kaufkraft stärken soll, das ist **sozialdemokratische Politik in Reinform**.*

‘That the state should support the economy in bad times and strengthen purchasing power with economic stimulus programs, that is social democratic policy in its purest form.’

(Die Zeit [Online-Ausgabe], 27 November 2008)

- (7) *Dass Royal bis nach dem letzten Urnengang wartete, unterstreicht, dass sie **Politikerin durch und durch** ist.*

‘That Royal waited until after the last ballot emphasises the fact that she is a politician through and through.’

(Spiegel-Online, 18 June 2007)

- (8) *Die Zutaten, die man für einen **Washingtoner Politikskandal erster Güte** benötigt, liegen schon bereit: die größte Unternehmenspleite der amerikanischen Geschichte, eine enge direkte Verbindung zwischen den Bankrotteuren*

und dem Präsidenten sowie eine Reihe von Wahlkampfspenden.

‘We already have all the ingredients required for a classic Washington political scandal: the largest corporate failure in American history, a close direct link between the bankrupt people and the President, and a series of campaign donations.’

(Spiegel-Online, 10 January 2002)

(9) *Kurz: Der Ökobonus ist grüne Politik, wie sie im Buche steht.*

‘In short: The eco bonus is a textbook example of Green policy (lit. as it stands in the book).’

(die Tageszeitung, 8 April 2008)

In the following, the aim is to present a theoretical classification of this constructional family within the framework of a usage-based and cognitively oriented Construction Grammar (Goldberg 1995, 2006). To do so, a description of these related constructions with regard to their network-like relationships will be given. With [X_{NP} *pur*], [X_{NP} *par excellence*] and [X_{NP} *wie er/sie/es im Buche steht*] three constructional phrasemes have been selected as examples and are analyzed with regard to their formal structure and their semantic-pragmatic characteristics. On the basis of the *German Reference Corpus (DeReKo)*² and with the help of the tool *Lexical Pattern Analyzer (lexpan)*³ usage restrictions and preferred fillers are identified. The analysis shows that the constructional phrasemes are not fully equivalent despite their common core meaning. The study also aims to contribute to a more intensive exchange between phraseology and Construction Grammar. The constructional family can be used to show how phraseology, which is traditionally dedicated to lexically filled units, and Construction Grammar, whose research focuses on abstract structures, can complement each other.

2 Theoretical framework

2.1 Constructional phrasemes

The above-mentioned examples 1–9 are so-called “Phrasem-Konstruktionen” (‘constructional phrasemes’), which Dobrovol’skij defines as follows:

² cosmas2.ids-mannheim.de/cosmas2-web (22 July 2019).

³ uww.ids-mannheim.de/lexpan (22 July 2019).

PhK können als Konstruktionen definiert werden, die als Ganzes eine lexikalische Bedeutung haben, wobei bestimmte Positionen in ihrer syntaktischen Struktur lexikalisch besetzt sind, während andere Slots darstellen, die gefüllt werden müssen.

‘Constructional phrasemes can be defined as constructions that have a lexical meaning as a whole, whereby certain positions in their syntactic structure are lexically filled, while others represent slots that must be filled.’ [our translation]

Dobrovol’skij (2011: 114)

Such structures, which in traditional German phraseology research are also referred to as “Phraseoschablonen” (‘phraseological templates’) (Fleischer 1997), “Modellbildungen” (‘pattern formation’) (Burger 2015) and “Wortverbindungenmuster” (‘word combination patterns’) (Steyer 2013), can be regarded as the central subject of investigation in Construction Grammar, where they are known under the terms “formal idioms” (Fillmore, Kay, and O’Connor 1988), “schematic idioms” (Croft and Cruse 2004) or “constructional idioms” (Langacker 1987).

In many grammar models phrasemes are banished in the lexicon because they are considered peripheral phenomena of a grammar description that is predominantly compositional, so-called “Kerngrammatik” (‘core grammar’) (Rostila 2012: 263). In Construction Grammar phrasemes gain in importance (cf. e.g. Fillmore, Kay, and O’Connor 1988; Fillmore 1988) because, as Staffeldt (2018: 144) rightly observes, they represent prototypical constructions (in the sense of Construction Grammar), i.e. pairs of forms and meanings whose formal and/or semantic-pragmatic particularities are not predictable.

According to Ziem (2018a: 3), phrasemes even represent perhaps the most important driving force for Construction Grammar, because – as the author explains – a grammar model that can adequately capture the syntactic and/or semantic idiosyncrasies of a language can also describe its regular structures. This turns out to be a necessity since many phenomena due to their syntagmatic and formal complexity – as is the case with constructional phrasemes – cannot be described as regular units of the lexicon without constraints (Ziem 2018a: 4).

The aim of Construction Grammar is to cover all structures, the so-called constructions, completely and uniformly belonging to the language system – even those that are not rule-based. Constructions, which are “conventionalized pairings of form and function” (Goldberg 2006: 3), can exhibit different degrees of abstraction and schematicity.

It is obvious that constructional phrasemes – because of their form and function – are of considerable interest for this model, since they can be classified at the interface between phraseology and syntax (cf. Fleischer 1997: 130–131; Burger 2018: 31; Ziem 2018a: 6, 2018b: 29–30; Stein and Stumpf 2019: chap-

ter 2.3). Due to their partially specified form, constructional phrasemes are hardly ever the object of investigation of traditional grammar models (Ziem 2018a: 6). They stand in fact on the continuum between idiomatic structures on the one hand and regular patterns or rules on the other (Ziem 2018a: 4). Construction Grammar rejects a strict separation between grammar and lexicon and postulates the so-called lexicon-grammar continuum (cf. Croft and Cruse 2004: 255–256; Goldberg 2013: 17; Ziem and Lasch 2013: chapter 8.1).

In traditional phraseology research, such model-like word combinations were already accorded a special status in the 1970s and 1980s (cf. Häusermann 1977; Fleischer 1982), but it is only due to the discussion with Construction Grammar that these structures have become of considerable interest to phraseology (cf. Mellado Blanco 2015; Schafroth 2015; Stumpf 2015b, 2017, 2021; Ziem 2018a, 2018b; Dobrovól'skij 2020). This may also be due to their partially specified lexical structure. Thus, they can be described neither as fully specified nor as abstract or schematic constructions (i.e. as lexemes or as purely grammatical units). Rather, they have both lexical and grammatical peculiarities that depend on the particular individual construction (Ziem 2018a: 6–7).

Construction Grammar and phraseology thus share the interest in such partly specific constructions and attempt “to grasp the regular in the irregular” (Ziem 2018a: 8; our translation); however, the view is taken from different perspectives, because – as Staffeldt (2018: 151) explains – phraseology is more interested in the slots to be filled lexically, while a construction grammatical analysis focuses more on the different degrees of abstraction of the constructions and the relations between the constructions. However, both approaches are also interested in the formal and semantic-pragmatic limitations of individual constructional phrasemes. In the present study, both perspectives are combined. The focus is on the slots of the constructional phrasemes which have to be filled lexically and their restrictions as well as on the relations between the constructions under analysis.

2.2 Families of constructions

Constructions that have formal-structural as well as semantic-pragmatic similarities are defined as “families of constructions” (Ruiz de Mendoza Ibáñez, Luzondo Oyón, and Pérez Sobrino 2017: 2; Diessel 2019: 199) within Construction Grammar. Constructions should in fact not be regarded as isolated units; rather they form groups “of formally and functionally connected configurations” (Ruiz de Mendoza Ibáñez, Luzondo Oyón, and Pérez Sobrino 2017: 2) in the so-called “constructi-

con”, which is a highly structured network in which constructions are organized prototypically and connected to each other through links and relations (Goldberg 1995: 51, 79–81, 2019: 34–37).

Here the concepts of “family” and “resemblance” are of relevance. Goldberg and Jackendoff (2013: 536) note with reference to Wittgenstein (1955) and Rosch and Mervis (1975): “By ‘family’ we have in mind the sort of family resemblances recognized to exist in nonlinguistic categories”. And they explain this concept on the basis of the resultative construction: “We see no choice but to treat the resultative as made up of a family of subconstructions, united by related but not identical syntax and by related but not identical semantics” (Goldberg and Jackendoff 2013: 563). In particular, we understand the family resemblance in reference to Engelberg et al. (2015) as a reflexive, symmetrical and non-transitive relationship. Such a relationship can be summarized in the words of Proost and Winkler, in relation to the argument structure constructions, in the following way:

Das heißt, jedes Argumentstrukturmuster ist sich selbst hinreichend ähnlich, und wenn zwei Muster einander ähnlich sind, so gilt diese Relation in beide Richtungen. Die Nicht-Transitivität der Ähnlichkeitsbeziehung besagt, dass aus der Tatsache, dass sowohl a^1 und a^2 als auch a^2 und a^3 einander ähnlich sind, nicht folgt, dass auch a^1 und a^3 einander ähnlich sind, weil die Ähnlichkeit von a^1 und a^2 einerseits und die Ähnlichkeit von a^2 und a^3 andererseits durchaus auf anderen Attributen beruhen kann.

‘That is, each argument structure pattern is sufficiently similar to itself; and if two patterns are similar to each other, this relation is valid in both directions. The non-transitivity of the resemblance relation means that it does not follow from the fact that both a_1 and a_2 as well as a_2 and a_3 are similar to each other, that a_1 and a_3 are also similar to each other, because the similarity of a_1 and a_2 on the one hand and the similarity of a_2 and a_3 on the other may well be based on other attributes.’ [our translation]

Proost and Winkler (2015: 10)

As Proost and Winkler (2015: 8–9) point out, families are made up of several members who share formal and semantic-pragmatic characteristics, but not all members of the same family have to share the same property. According to this understanding of family resemblance “the individual members of a family must be sufficiently similar to other members of the family, but not to all of them” (Proost and Winkler 2015: 9; our translation).

Family resemblances have been described above all in the field of grammar (cf. e.g. Goldberg and Jackendoff 2013; Ruiz Mendoza Ibáñez, Luzondo Oyón, and Pérez Sobrino [eds.] 2017; Mollica 2014), especially in argument structure constructions (cf. e.g. Engelberg et al. 2015; Proost 2015, 2017; Proost and Winkler 2015). Only a few studies investigate the resemblances within phraseologi-

cal units (cf. e.g. Staffeldt 2011; Bücken 2014; Auer 2016; Mollica 2020). Therefore, we focus our study on this topic, which is also relevant from a construction grammatical point of view.

The phraseological units [X_{NP} *pur*], [X_{NP} *par excellence*], [X_{NP} *wie er/sie/es im Buche steht*], [X_{NP} *schlechthin*], [X_{NP} *durch und durch*], [X_{NP} *erster Güte*] etc. are constructional phrasemes: the lexically open noun phrase is specified by a postponed lexically fixed attribute.

However, the units under examination can all be traced back to a common abstract construction with the following schematic form and abstract meaning:

Form: [X_{NP} attribute]

Meaning/Function: The postnominal attribute intensifies the X_{NP} -position so that it is considered to be (proto-)typical, particularly pronounced and/or unsurpassable). An evaluative stance on the part of the speaker is expressed.

According to Ruiz de Mendoza Ibáñez, Luzondo Oyón, and Pérez Sobrino (2017: 4), the study of families of constructions in terms of their structural and semantic characteristics and of the relationships between members of the family is “a useful analytical pathway to put to the test the explanatory potential of CxG approaches, thereby contributing to their theoretical development”. The holistic description of constructions within a family can also be very useful for describing lexically partially and fully specified constructions such as phrasemes, in order to record analogies and differences in related constructions.

In order to determine family resemblances and divergences within the constructional phraseme [X_{NP} attribute] presented above at the macro-level (for the different construction levels see Section 4), it is necessary to examine the individual family members with regard to their meaning and the fillers (and their possible constraints) that can occupy the X_{NP} -position. It is also interesting to observe whether the individual structures are more or less productive; because the more fillers that can occupy a slot, the more productive the construction is (Ziem 2018a: 10), whereby the schematicity of a construction (and the resulting possibility to fill its slots with new lexical units) correlates with its entrenchment (Ziem and Lasch 2013: 103).

3 Case studies

3.1 Methodological approach

To determine formal-structural as well as semantic-pragmatic resemblances and differences within the family members of [X_{NP} attribute], we refer to the *German Reference Corpus* and the tool *Lexical Pattern Analyzer (lexpan)*, developed in the project “Usuelle Wortverbindungen” (‘multi-word expressions in common usage’).⁴ *Lexpan* allows automatic counting of lexical fillers in a slot as well as the determination of the extent of their occurrence within a slot (Steyer 2012: 305, 2013: 122). The slot analysis and the associated filler tables are a useful tool for researching family-like constructional phrasemes. As Steyer (2011: 230) explains, the slot analyses provide empirical indications of the nature of lexical fillers and the scale of typicality or productivity. They help to identify possible semantic-pragmatic differences within a constructional family. According to the postulate of non-synonymy (Goldberg 1995: 67), the constructional phrasemes we deal with here cannot be regarded as full equivalents despite their common core meaning.

Slot analyses have already been carried out in a preliminary investigation for most of the constructions shown at the beginning (examples 1–9) of this paper. Since a detailed description of all the family members would go beyond the scope of the current study, the constructional phrasemes [X_{NP} *pur*] (lit. [X_{NP} pure]), [X_{NP} *par excellence*] (lit. [X_{NP} par excellence]) and [X_{NP} *wie er/sie/es im Buche steht*] (lit. [X_{NP} as he/she/it stands in the book]) have been selected for the analysis below.⁵ The selection of these phrasemes can be justified as follows: on the one hand, these constructions are structurally different from one another. And on the other hand, [X_{NP} *pur*] differs from the other two from a semantic-pragmatic point of view. [X_{NP} *par excellence*] and [X_{NP} *wie er/sie/es im Buche steht*], however, are very similar, if not almost identical, in their core meaning.

⁴ In addition, examples from the database for Spoken German (dgd.ids-mannheim.de/dgd/pragdb.dgd_extern.welcome, 22 July 2019) as well as evidence from social media (e.g. tweets) will be used to illustrate the fact that the constructional phrasemes also occur in spoken language.

⁵ Only the literal translation of the constructional phrasemes in English – and not a functional equivalent – is given here.

3.2 Corpus-based analysis of the constructions [X_{NP} *pur*], [X_{NP} *par excellence*] and [X_{NP} *wie er/sie/es im Buche steht*]

3.2.1 Morphosyntactic characteristics

From a morphosyntactic point of view, the three constructions differ with regard to attribute type. We find in [X_{NP} *pur*] a postponed uninflected adjective attribute, in [X_{NP} *par excellence*] a prepositional phrase borrowed from the French and in [X_{NP} *wie er/sie/es im Buche steht*] a phraseological subordinate clause. For all three constructional phrasemes, the nominal fillers are usually singular; in [X_{NP} *pur*] the nouns are used without an article and in [X_{NP} *par excellence*] and [X_{NP} *wie er/sie/es im Buche steht*] with an indefinite article. The selection of the articles results in particular from the semantics of the lexical fillers (see section 3.2.2). It should also be mentioned that in the *pur*-construction the adjective attribute can also stand in the position before the noun (*pure Freude* instead of *Freude pur*). However, a quantitative corpus analysis shows that in many cases the postposition is much more frequent than the preposition (e.g. in constructions with the nouns *Stimmung* ‘atmosphere’, *Spannung* ‘suspense’ and *Dramatik* ‘drama’)⁶ (Stumpf 2017: 324).

If we look more closely at the syntactic embedding of these constructions, we can see that the three constructional phrasemes frequently occur as predicative complements, mostly in connection with the copula verb *sein* ‘to be’ (see examples 10–12). The constructions are thus embedded in the more abstract copula construction, which can be schematized as follows: [Y_{NP} copula verb_[sein] X_{NP} *pur* / *par excellence* / *wie er/sie/es im Buche steht*].⁷

(10) [*Die Olympia-Bewerbung*]_Y *ist* [*Stadtmarketing*]_X ***pur***.

‘The bid for the Olympics is pure city marketing.’

(Mannheimer Morgen, 29 March 2003)

(11) Für viele Deutschschweizer *ist* [das Tessin]_Y [***der Ferienkanton***]_X ***par excellence***.

⁶ *pure Stimmung* (6 instances, 1 %) versus *Stimmung pur* (625 instances, 99 %); *pure Spannung* (34 instances, 1 %) versus *Spannung pur* (1,819 instances, 99 %); *pure Dramatik* (18 instances, 2 %) versus *Dramatik pur* (748 instances, 98 %).

⁷ In our explanations, we refer to the realization of the constructional phrasemes [X_{NP} attribute] within the copula-construction, which can be considered its prototypical application.

‘For many German-speaking Swiss, Ticino is the holiday canton par excellence.’

(St. Galler Tagblatt, 23 May 1997)

- (12) [*Heinz Hoschek*]_y *ist* [*ein Verlierer*]_x, *wie er im Buche steht*.

‘Heinz Hoschek is a textbook example of a loser (lit. as he stands in the book).’

(Nürnberger Zeitung, 3 December 2004)

In the *pur*-construction, in many cases, pronouns (usually *das* ‘that’) appear as subjects which refer to a previous text passage (see example 13). In addition, the construction is used as a subject for example with the verb *herrschen* (‘to prevail’) (usually in connection with an adverbial complement) (see example 14) and as an accusative complement, for example with the verbs *bieten* ‘to offer’, *versprechen* ‘to promise’ and *erleben* ‘to experience’ (see examples 15–17).

- (13) [*Relaxen am Sandstrand zwischen Palmen*]_y: [*Das*]_{pronoun} *ist* [*Karibik-Urlaub*]_x *pur am Playa Bavaro in der Ferienregion von Punta Cana*.

‘Relaxing on the sandy beach among palm trees: That is the very essence of a Caribbean holiday (lit. a Caribbean holiday pure) on Playa Bavaro in the holiday region of Punta Cana.’

(Hamburger Morgenpost, 15 March 2009)

- (14) [*Erleichterung*]_x *pur herrscht* [*im Lager der Braunschweiger Basketballer*]_y *nach dem 72:70-Sieg in Trier*.

‘There is complete and utter (lit. pure) relief in the Braunschweig basketball players’ camp after the 72:70 victory in Trier.’

(Braunschweiger Zeitung, 10 October 2012)

- (15) [*Kokolores*]_x *pur boten* [*die zwei „durchgeknallten“ Kirchspielkehrer Manfred Wolfs und Konrad Stephan*]_y [...].

‘The two “crazy” Kirchspielkehrer Manfred Wolfs and Konrad Stephan had nothing to offer but pure nonsense.’

(Rhein-Zeitung, 06 February 2006)

- (16) [*Das Hotel am Berg*]_y *verspricht* [*Abenteuer*]_x *pur*.

‘The Hotel am Berg promises pure adventure.’

(NEWS, 14 June 2006)

- (17) Hier *erleben* [die kleinen und großen Besucher]_Y [**Schneespaß**]_X **pur** und lernen spielerisch das Skifahren.

‘Here, children and grown-ups can have enormous (lit. pure) fun in the snow and learn to ski while playing.’

(Luxemburger Tageblatt, 22 November 2013)

Both [X_{NP} *par excellence*] and [X_{NP} *wie er/sie/es im Buche steht*] can be found as appositions (see examples 18–19). However, the *par excellence*-construction can also be used with verbs like *gelten* ‘to be considered’ (see example 20), *bezeichnen* ‘to call’, *sich erweisen* ‘to turn out to be’ and *entpuppen* ‘to turn out to be’ in connection with a predicative complement (introduced by the particle *als*).

- (18) [*Elke Strathmann*]_Y, 56, [**Karrierefrau**]_X **par excellence**, muss sich wohl einen neuen Job suchen.

‘Elke Strathmann, 56, a career woman *par excellence*, will probably have to look for a new job.’

(Süddeutsche Zeitung, 15 April 2014)

- (19) [*Aglaiä Konrad*]_Y, [**eine Globetrotterin**]_X, **wie sie im Buche steht**, präsentiert ihre Architekturfotografien von Weltstädten in Europa, Japan, China und Südamerika aus der Fussgänger-Perspektive [...].

‘Aglaiä Konrad, a textbook example of a world traveler (lit. as she stands in the book), presents her architectural photographs of cosmopolitan cities in Europe, Japan, China and South America from the perspective of a pedestrian.’

(Neue Zürcher Zeitung, 10 October 2011)

- (20) [*Winterkorn*]_Y gilt als [**Automann**]_X **par excellence**.

‘Winterkorn is considered a car guy *par excellence*.’

(VDI nachrichten, 10 November 2006)

3.2.2 Semantic and pragmatic characteristics

Particularly with regard to the semantics of the superordinate constructions in which the constructional phrasemes are embedded, it can be said that the *par excellence* and *wie er/sie/es im Buche steht*-constructions have more or less the same core meaning, which must be differentiated from that of the [X_{NP} *pur*]-construction.

The meaning of [X_{NP} *par excellence*] and [X_{NP} *wie er/sie/es im Buche steht*] can be paraphrased as follows: ‘The entity passing into the Y_{NP} -position (hyponym) represents a typical, exemplary version of the X_{NP} -position (hyperonym) or, within the X_{NP} -position class, an entity of the highest degree (establishing relations between two entities)’.

[X_{NP} *pur*] means as much as: ‘The entity in the X_{NP} -position is to be characterized as genuine, the highest degree of a particular property’; often an entity (Y_{NP}) is attributed a certain characteristic by [X_{NP} *pur*] (as a predicative complement). It is interesting that the postnominal use of *pur* compared to the prenominal use leads to a stronger intensification (e.g. *Romantik* ‘romance’ < *pure Romantik* < *Romantik pur*). Perhaps this is also due to the fact that a postnominal positioning of an adjective violates the rules of German syntax (cf. Stumpf 2015a: chapter 6). According to Duden (2011: 50), the postposition serves as an expressive stylistic device in colloquial language. In other words, the postposition has a semantic-pragmatic added value compared to the preposition, which is typical for referential phrasemes in general (cf. Kühn 1985, 1994). *Romantik pur* is more emotional, more passionate, more sensitive etc. than *pure Romantik*. This also corresponds to Goldberg’s (1995) principle of non-synonymy discussed above.

As already mentioned, the slot analysis of [X_{NP} *pur* / *par excellence* / *wie er/sie/es im Buche steht*] in particular provides indications of possible differences in the constructions. Tables 1–3 show a selection of the fillers which can be found in the X_{NP} -position.

Tab. 1: Fillers for [X_{NP} *pur*] (section)

Search query: # pur

The search query includes 9,871 KWICs and 2,187 different fillers

Filler	Frequency	Percentage
<i>Natur</i> (‘nature’)	748	7.56
<i>Spannung</i> (‘suspense’)	591	5.99
<i>Kultur</i> (‘culture’)	293	2.97
<i>Abstiegskampf</i> (‘ranking battle’)	262	2.65
<i>Dramatik</i> (‘drama’)	217	2.20
<i>Entspannung</i> (‘relaxation’)	164	1.66
<i>Romantik</i> (‘romance’)	149	1.51
<i>Stress</i> (‘stress’)	136	1.38
<i>Action</i> (‘action’)	135	1.37

Search query: # pur

The search query includes 9,871 KWICs and 2,187 different fillers

Filler	Frequency	Percentage
<i>Luxus</i> ('luxury')	129	1.31
<i>Erholung</i> ('recreation')	127	1.29
<i>Unterhaltung</i> ('entertainment')	126	1.28
<i>Gänsehaut</i> ('goose bumps')	115	1.17
<i>Lebensfreude</i> ('joie de vivre')	106	1.07
<i>Sonne</i> ('sun')	102	1.03
<i>Emotionen</i> ('emotions')	101	1.02

Tab. 2: Fillers for [X_{NP} *par excellence*] (section)**Search query: # par excellence**

The search query includes 9,674 KWICs and 6,079 different fillers

Filler	Frequency	Percentage
<i>Beispiel</i> ('example')	43	0.44
<i>Land</i> ('country')	38	0.39
<i>Kino</i> ('cinema')	29	0.30
<i>Ort</i> ('place')	28	0.29
<i>Fahrmaschine</i> ('driving machine')	26	0.27
<i>Film</i> ('movie')	25	0.26
<i>Klassiker</i> ('classic')	25	0.26
<i>Stadt</i> ('city')	25	0.26
<i>Wachstumsbranche</i> ('growth industry')	25	0.26
<i>Erfolgsgeschichte</i> ('success story')	20	0.21
<i>Europäer</i> ('European')	20	0.21
<i>Künstler</i> ('artist')	20	0.21
<i>Medium</i> ('medium')	20	0.21
<i>Kunst</i> ('art')	19	0.20
<i>Partei</i> ('party')	18	0.19
<i>Institution</i> ('institution')	17	0.18

Tab. 3: Fillers for [X_{NP} *wie er/sie/es im Buche steht*] (section)

Search query: # wie er/sie/es im Buche steht
The search query includes 3,091 KWICs and 1,999 different fillers

Filler	Frequency	Percentage
<i>Stadt</i> ('city')	40	1.29
<i>Architektur</i> ('architecture')	33	1.07
<i>Macho</i> ('macho')	15	0.49
<i>Mann</i> ('man')	14	0.45
<i>Selbmademann</i> ('self-made man')	14	0.45
<i>Familienbetrieb</i> ('family business')	13	0.42
<i>Hypochonder</i> ('hypochondriac')	12	0.39
<i>Dieb</i> ('thief')	11	0.36
<i>Held</i> ('hero')	10	0.32
<i>Beruf</i> ('job')	9	0.29
<i>Betrüger</i> ('fraud')	9	0.29
<i>Bürger</i> ('citizen')	9	0.29
<i>Gentleman</i> ('gentleman')	9	0.29
<i>Text</i> ('text')	9	0.29
<i>Derby</i> ('derby')	8	0.26
<i>Patriarch</i> ('patriarch')	8	0.26

Here we can already see that all three constructional phrasemes are very productive. However, the productivity of [X_{NP} *par excellence*] (9,674 tokens, 6,079 types) and [X_{NP} *wie er/sie/es im Buche steht*] (3,091 tokens, 1,999 types) is higher than of [X_{NP} *pur*] (9,871 tokens, 2,187 types).⁸ We examined the first 250 fillers with regard to semantic categories such as 'person', 'concretum/object' and 'abstractum'. It turns out that the slot of the *pur*-construction is almost exclusively filled by abstracta (e.g. *Stress* 'stress', *Abwechslung* 'change', *Zuversicht* 'confidence'). The construction therefore does not serve to categorize (perceptible) objects (i.e. con-

8 It should be noted that the number of hits is slightly distorted, since among the concrete filler elements there are also words that indicate a different syntactic realization of the construction. In these variants, the nominal reference word is not placed immediately before the attribute. Cf. for example the word *so* 'thus' in the context of [X_{NP} *wie er/sie/es im Buche steht*]: „Lieber Wolfgang Röken, lieber Wolfgang, ein netter Kerl, so wie er im Buche steht.“ (Protokoll der Sitzung des Parlaments Landtag Nordrhein-Westfalen, 25 March 2010) ['Dear Wolfgang Röken, dear Wolfgang a nice guy, just like a textbook example.'].

creta or living beings) but abstract entities (feelings, [mental] states, [socio-]political ideas/systems etc.). Furthermore, fixed, fully lexicalized word combinations also seem to exist (*Natur pur* ‘Nature pure’, *Spannung pur* ‘pure suspense’), which, because of their high frequency of occurrence, have to a certain extent become conventionalized. In contrast, in the *par excellence*-construction there are hardly any restrictions regarding the instances that can enter the X_{NP} -position; there are persons (*Künstler* ‘artist’, *Entertainer* ‘entertainer’, *Politiker* ‘politician’), concreta/objects (*Stadt* ‘city’, *Theater* ‘theater’, *Schnellimbiss* ‘fast food restaurant’) and abstracta (*Erlebnis* ‘adventure’, *Skandal* ‘scandal’, *Unsinn* ‘nonsense’). [X_{NP} *wie er/sie/es im Buche steht*] is characterized by the fact that nouns referring to persons most often function as filling elements (*Gastronom* ‘restaurateur’, *Feuerwehrmann* ‘firefighter’, *Berlinerin* ‘Berlin woman’), and much more frequently than in the *par excellence*-construction.

In our opinion it is also worth taking a look at the stylistic peculiarities of the three constructions. Thus, in the *pur*-construction there is a for German rather unusual postposition of the adjective attribute. The construction is therefore also the subject of lay language criticism, which judges the postposition to be ungrammatical (Stumpf 2017: 317–319). The *par excellence*-construction is – probably due to its borrowing from French – described as “educated language”⁹ in Duden online. [X_{NP} *wie er/sie/es im Buche steht*] is an idiom of biblical origin, which is described as “colloquial” in Duden (2013: 140). Conspicuous at the formal level is the retention of the dative-*e* (cf. Stumpf 2015a: section 5).

Finally, from a pragmatic-functional point of view, the three expressions can be seen to have certain characteristics in common: all three constructions express an evaluative speaker stance, namely one of intensification (on [X_{NP} *pur*] see also Dürscheid 2002: 67). The evaluation of a fact can – under circumstances also depending upon the noun used – be positive (*Entspannung* ‘relaxation’, *Begeisterung* ‘enthusiasm’, *Sinnlichkeit* ‘sensuality’ *pur*, *Hörgenuss* ‘listening pleasure’, *Körperbeherrschung* ‘body control’, *Goalgetter* ‘scorer’ *par excellence*, *Held* ‘hero’, *Gentleman* ‘gentleman’, *Führungsspieler* ‘leader’ *wie er/sie/es im Buche steht*), but in some cases also negative (*Enttäuschung* ‘disappointment’, *Tristesse* ‘tristesse’, *Verzweiflung* ‘despair’ *pur*, *Bananenrepublik* ‘banana republic’, *Etikettenschwindel* ‘fraud’, *Trauerspiel* ‘tragedy’ *par excellence*, *Macho* ‘macho’, *Nazi* ‘Nazi’, *Faulpelz* ‘slacker’ *wie er/sie/es im Buche steht*).¹⁰ On the one hand, the evaluative intention thus depends on the connotation of the nouns that fill the

⁹ www.duden.de/rechtschreibung/par_excellence (22 July 2019).

¹⁰ Here it can be observed that the slot of [X_{NP} *wie er/sie/es im Buche steht*] tends to have more negative than positive connotations with proper names.

slot. On the other hand, the (evaluative) function is also reinforced by the frequent embedding of the constructional phraseme in a copula construction, within which a pronoun refers to the event to be evaluated (in most cases *Das ist/war* [‘that is/was’] [X_{NP} *pur / par excellence / wie er/sie/es im Buche steht*]) (see examples 21–23):

- (21) *[Es kommen dauernd neue Verfehlungen ans Tageslicht. In keinem europäischen Land wäre er noch tragbar]_Y – [das]_{pronoun} ist [Berlusconi-Politik]_X pur.*

‘More and more offences are constantly coming to light. In no European country would he still be tolerated – that is pure Berlusconi policy.’

(NEWS, 29 January 2004)

- (22) *[Saftiges Fleisch vom Lamm, dazu die appetitliche Frische des Granatapfels]_Y: [Das]_{pronoun} ist [libanesishe Küche]_X par excellence.*

‘Juicy lamb, plus the delicious freshness of pomegranate: this is classic Lebanese cuisine.’

(Sonntagsblick, 7 February 2010)

- (23) *Vor allen Dingen haben CDU und CSU erklärt, [dass sie auf keinen Fall die Rücknahme der gegenseitigen Unterhaltspflicht von Eltern und Kindern akzeptieren]_Y. [Das]_{pronoun} ist [Familienpolitik der CDU/CSU]_X, wie sie im Buche steht.*

‘First and foremost, the CDU and CSU have stated that they under no circumstances accept the withdrawal of the required mutual financial responsibility maintenance obligations of parents and children. This is a textbook example of CDU / CSU family policy (lit. as it stands in the book).’

(Protokoll der Sitzung des Parlaments Landtag Nordrhein-Westfalen, 16 November 2003)

In addition, concrete evidence can be found for all three constructional phrasemes that can be classified as conceptually spoken (see examples 24–29). The constructions are therefore not limited to conceptually written texts.

- (24) Conversation in a driving lesson (FOLK_E_00146_SE_01_T_01)

0237 RK so pass auf

0238 (0.3)

0239 RK da sind se voll an arbeiten °h und das is **stress**
pur

‘so watch out / there they are busy working and that is pure stress’

(25) Tweet from FC Augsburg¹¹



FC Augsburg
@FCAugsburg

Folgen

90' Ekstase pur mit dem Abpfiff. Jetzt alles geben und das Momentum auf unsere Seite holen. WAS FÜR EIN FIGHT! #FCARBL #DFBPokal

'90 minutes of pure ecstasy with the final whistle. Give your all now and get the momentum on our side. WHAT A FIGHT!'

(26) Interview (IS--_E_00093_SE_01_T_01)

0069 S1 aber dein bruder war kein kommunist mehr

0070 S2 nein den kommunismus legte er sehr schnell ab er wurde ein **kapitalist par excellence**

'but your brother was no longer a communist / no, he very quickly discarded communism and became a capitalist par excellence'

(27) Tweet from Inge Hannemann¹²



Inge Hannemann
@IngeHannemann

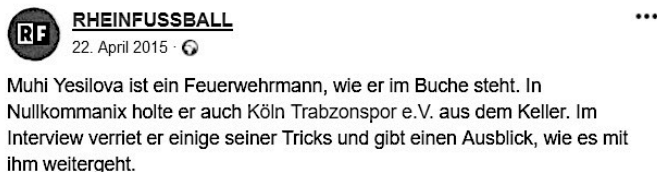
Folgen

Ein Bürowahnsinn par excellence. Oder wie erschaffe ich interne Arbeitsbeschaffungsmaßnahmen?: "
#Jobcenter geben 60 Millionen Euro aus, um 18 Millionen einzutreiben" #hartzIV

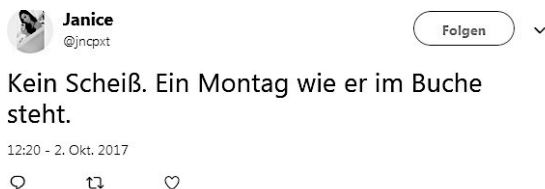
'A classic case of Bureaucracy gone mad. Or how do I create internal job creation measures? Job centers spend 60 million euros to collect 18 million'

¹¹ twitter.com/FCAugsburg/status/1113179379440517120 (22 July 2019).

¹² twitter.com/IngeHannemann/status/1100651843057782785 (22 July 2019).

(28) Facebook post from RHEINFUSSBALL¹³

‘Muhi Yesilova is a textbook example of someone who can put out. In no time he also saved Köln Trabzonspor e.V. from a lower ranking. In the interview, he revealed some of his tricks and tells us what the future looks like for him.’

(29) Tweet from Janice¹⁴

‘No shit. A textbook example of a Monday (lit. as it stands in the book).’

It is worth mentioning that all three constructions often function as headings. In these, there is usually no Y_{NP} -position, whereby the constructions are intended to arouse the interest of the recipient and provide an incentive for him to read the whole article. The reader wants to know what exactly *Entspannungsstress pur* ‘relaxation stress pure’ refers to (see example 30), what *Erfolgsgeschichte* ‘success story’ is rated as an *Erfolgsgeschichte par excellence* (see example 31) and what kind of *Betrug* ‘fraud’ it is, for it to be described as *Betrug, wie er im Buche steht* (see example 32).

¹³ de-de.facebook.com/RHEINFUSSBALL/posts/muhi-yesilova-ist-ein-feuerwehrmann-wie-er-im-buche-steht-in-nullkommanix-holte-/628417827289284/ (22 July 2019).

¹⁴ twitter.com/jncpxt/status/914933081534017536 (22 July 2019).

- (30) Headline in the *Süddeutsche Zeitung* newspaper¹⁵



- (31) Headline in the *Fränkische Nachrichten* newspaper¹⁶



- (32) Headline in the *Neue Ruhr Zeitung* newspaper¹⁷



¹⁵ www.sueddeutsche.de/leben/schoen-doof-entspannungsstress-pur-1.3655660 (22 July 2019).

¹⁶ www.fnweb.de/fraenkische-nachrichten_artikel,-lauda-koenigshofen-eine-erfolgsgeschichte-par-excellence_action,fotos_arid,1324944.html (22 July 2019).

¹⁷ www.nrz.de/region/niederrhein/humor/betrug-wie-er-im-buche-steht-id209234403.html (22 July 2019).

4 The constructional family [X_{NP} attribute]: the different levels

While the constructional phrasemes [X_{NP} *pur*], [X_{NP} *par excellence*], [X_{NP} *wie er/sie/es im Buche steht*], [X_{NP} *schlechthin*], [X_{NP} *durch und durch*], [X_{NP} *erster Güte*] etc. show formal-structural differences with regard to the realization of the postponed attribute, they also share the intensifying function described above which expresses the speech act of evaluation. The constructional phrasemes can therefore be traced back to the more common abstract construction [X_{NP} attribute] which is a frequent and productive pattern of German. On the one hand, numerous constructional phrasemes are formed according to this pattern, which, due to their formal-structural and semantic-pragmatic resemblances, can be regarded as members of a family and described as a complex network. On the other hand, our corpus analysis shows that their slots are not subject to severe semantic restrictions; for this reason, these constructional phrasemes can be regarded as highly productive.

According to Traugott, we differentiate between several construction levels in the evaluating constructional family [X_{NP} attribute]:

- *macro-constructions*: meaning-form pairings that are defined by structure and function [...]
- *meso-constructions*: sets of similarly-behaving specific constructions
- *micro-constructions*: individual construction-types
- *constructs*: the empirically attested tokens [...]

Traugott (2008: 236)

On the macro-level – at the highest and schematic level – the existence of the completely abstract, non-lexicalized construction [X_{NP} attribute] is postulated, which is an object of investigation not of phraseology, but of Construction Grammar. Partly lexicalized multi-word expressions like [X_{NP} *pur*], [X_{NP} *par excellence*], [X_{NP} *wie er/sie/es im Buche steht*], [X_{NP} *in Reinform*], [X_{NP} *durch und durch*] etc. result from this macro-construction and represent the meso-constructions, which – due to their lexically fixed components – can be regarded as phraseological units. These different constructional phrasemes form a network, a family of formal-structurally and semantical-pragmatically /functionally interconnected constructions. Micro-constructions can be understood as fully lexicalized constructions that are entrenched and

conventionalized because they have a high token frequency.¹⁸ The constructs, the respective corpus examples, clearly express the degree of entrenchment of a micro-construction (see Fig. 1).

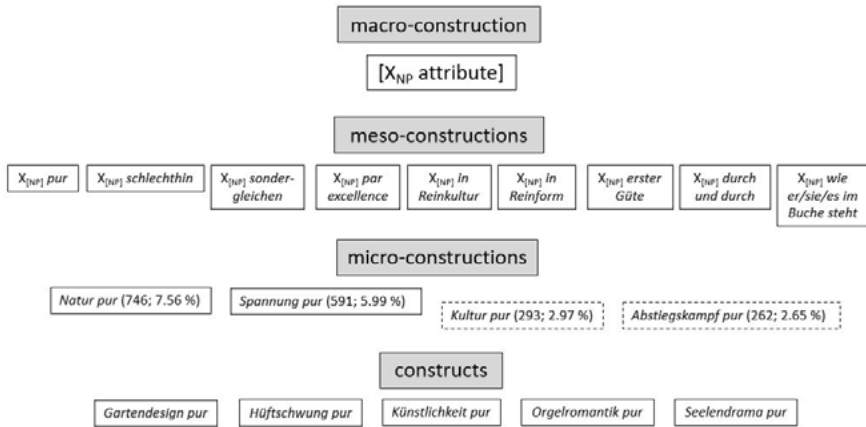


Fig. 1: The constructional family [X_{NP} attribute]

The high token frequency, which basically can only be assessed in relation to other instantiations of the construction, can be determined with the help of slot analyses. In our opinion, *Natur pur* ‘Nature pure’ and *Spannung pur* ‘pure suspense’ in the constructional phraseme [X_{NP} pur] have a very high frequency and a very high percentage in relation to other filling elements, so that they can be described as micro-constructions compared to constructs such as *Gartendesign* ‘garden design’, *Hüftschwung* ‘swivel of the hips’, *Künstlichkeit* ‘artificiality’, *Orgelromantik* ‘Organ romance’ and *Seelendrama* ‘psychological drama’ pur. The question that inevitably arises here is the demarcation between constructs and micro-constructions. In other words: are such examples as *Kultur pur* and *Abstiegskampf pur* instantiations of the meso-construction [X_{NP} pur] or, due to their non-negligible frequency, already completely lexically entrenched and

¹⁸ It should be noted here that micro-constructions are regarded in different ways in the literature. Following Fried (2013: 437), we understand them to be “substantive constructions”, i.e. fully lexicalized multiword connections. Hoffmann and Bergs (2018: 8), on the other hand, consider them to be merely “concrete constructions with limited generality and abstractness [...] such as *She refused him a kiss.* or *They refused her the answer [...]*”.

thus individual micro-constructions?¹⁹ This question cannot be answered at this point; we need more extensive corpus-based studies which also look at other constructions in order to draw or postulate possible boundaries within the different construction levels.

5 Conclusions

In our contribution we have examined the constructional phrasemes [X_{NP} *pur*], [X_{NP} *par excellence*] and [X_{NP} *wie er/sie/es im Buche steht*]. Construction Grammar was chosen as the description model because it has a number of advantages in the description of phrasemes. While phraseology focuses on descriptions of usage restrictions and slot preferences for each individual constructional phraseme, Construction Grammar enables a uniform description of constructions that are similar in form and function, so that they are (no longer) considered separately – as in traditional phraseology – but can be seen as a network or as members of the same family. Furthermore, Construction Grammar makes it possible to postulate different levels of constructions or constructions of different schematicity, so that phraseological phenomena can be viewed from a different and also more abstract perspective.

Thus, the constructional phrasemes [X_{NP} *pur*], [X_{NP} *par excellence*] and [X_{NP} *wie er/sie/es im Buche steht*] represent more concrete instantiations of the abstract construction [X_{NP} attribute], which has turned out to be a frequent and productive pattern; according to this schema, a series of additional constructional phrasemes exist in German, such as [X_{NP} *durch und durch*], [X_{NP} *erster Güte*], [X_{NP} *in Reinkultur*] and [X_{NP} *sondergleichen*]. Since these lexically partly specified phenomena have a similar form and meaning (they consist formally of a nominal phrase and a postnominal attribute and have an intensifying function), we have regarded them as a family of constructions.

Constructions, however, are interconnected in the constructicon; thus, traditional phraseology would place the constructional phraseme [X_{NP} *wie er/sie/es im Buche steht*] among the comparative phrasemes without clearly emphasizing the formal and functional resemblances with the evaluative construction [X_{NP} attribute]. The description inventory of Construction Grammar, by contrast, enables a better definition of the analogies between the constructions. Thus, the

¹⁹ For this reason, the two examples are shown in figure 1 in dotted boxes between the micro level and the constructs level.

constructional phraseme [X_{NP} *wie er/sie/es im Buche steht*] has prototypical properties of comparative phrasemes, but also shares features with the constructional family described in this article, whereby it can perhaps be regarded – due to its form – as a peripheral member.²⁰

In this article we hope to have shown that the concept of family resemblance can also be useful in phraseological questions, as it places phrasemes in a new light.

References

- Auer, Peter. 2016. *Wie geil ist das denn?* – Eine neue Konstruktion im Netzwerk ihrer Nachbarn. *Zeitschrift für germanistische Linguistik* 44. 69–92.
- Bücker, Jörg. 2014. Konstruktionen und Konstruktionscluster: Die Zirkumposition von XP *her* im gesprochenen Deutsch. In Alexander Lasch & Alexander Ziem (eds.), *Grammatik als Netzwerk von Konstruktionen? Sprachwissen im Fokus der Konstruktionsgrammatik*, 117–135. Berlin & Boston: de Gruyter.
- Burger, Harald. 2015. *Phraseologie. Eine Einführung am Beispiel des Deutschen*. 4., neu bearbeitete Auflage. Berlin: Schmidt.
- Burger, Harald. 2018. Wie weit, wie eng soll der Objektbereich der Phraseologie sein? Beobachtungen eines Zeitzeugen zu Konzeptionen im Wandel. In Natalia Filatkina & Sören Stumpf (eds.), *Konventionalisierung und Variation. Phraseologische und konstruktionsgrammatische Perspektiven*, 19–38. Berlin: Lang.
- Croft, William & Alan D. Cruse. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- De Knop, Sabine & Fabio Mollica. 2017. The family of German dative constructions. In Francisco Ruiz de Mendoza Ibáñez, Alba Luzondo Oyón & Paula Pérez Sobrino (eds.), *Constructing Families of Constructions*, 205–239. Amsterdam & Philadelphia: Benjamins.
- Diessel, Holger. 2019. *The Grammar Network. How Linguistic Structure is Shaped by Language Use*. Cambridge: Cambridge University Press.
- Dobrovolskij, Dmitrij. 2011. Phraseologie und Konstruktionsgrammatik. In Alexander Lasch & Alexander Ziem (eds.), *Konstruktionsgrammatik III. Aktuelle Fragen und Lösungsansätze*, 111–130. Tübingen: Stauffenburg.
- Dobrovolskij, Dmitrij. 2020. German constructional phrasemes with deictic elements in bilingual dictionaries. In Paola Cotta Ramusino & Fabio Mollica (eds.), *Contrastive Phraseology: Languages and Cultures in comparison*, 115–127. Cambridge: Cambridge Scholars.
- Duden. 2011. *Richtiges und gutes Deutsch. Das Wörterbuch der sprachlichen Zweifelsfälle*. 7. Auflage. Mannheim: Dudenverlag.

20 However, the constructional family also includes simple lexemes (adjectives) with an evaluative function such as *pur* ‘pure’, *rein* ‘pure’, *perfekt* ‘perfect’, *(proto-)typisch* ‘(proto-)typical’ etc., which can be used in similar contexts (*das war Spaß pur!* ‘this was pure fun!’ versus *das war purer/reiner/perfekter Spaß* ‘it was pure/perfect fun!’).

- Duden. 2013. *Redewendungen. Wörterbuch der deutschen Idiomatik*. 4., neu bearbeitete und aktualisierte Auflage. Berlin: Dudenverlag.
- Dürscheid, Christa. 2002. *Polemik satt und Wahlkampf pur* – Das postnominale Adjektiv im Deutschen. *Zeitschrift für Sprachwissenschaft* 21. 57–81.
- Engelberg, Stefan, Svenja König, Kristel Proost, & Edeltraud Winkler. 2011. Argumentstrukturmuster als Konstruktionen? Identität – Verwandtschaft – Idiosynkrasien. In Stefan Engelberg, Anke Holler & Kristel Proost (eds.), *Sprachliches Wissen zwischen Lexikon und Grammatik*, 71–112. Berlin & Boston: de Gruyter.
- Fillmore, Charles, Paul Kay & Mary Catherine O'Connor. 1988. Regularity and Idiomaticity in Grammatical Constructions. The Case of *let alone*. *Language* 64. 501–538.
- Fleischer, Wolfgang. 1982. *Phraseologie der deutschen Gegenwartssprache*. Leipzig: VEB Bibliographisches Institut.
- Fleischer, Wolfgang. 1997. *Phraseologie der deutschen Gegenwartssprache*. 2., durchgesehene und ergänzte Auflage. Tübingen: Niemeyer.
- Fried, Mirjam. 2013. Principles of Constructional Change. In Thomas Trousdale & Graeme Hoffmann (eds.), *The Oxford Handbook of Construction Grammar*, 419–437. Oxford: Oxford University Press.
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.
- Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Goldberg, Adele E. 2013. Constructionist Approaches. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*, 15–31. Oxford: Oxford University Press.
- Goldberg, Adele E. 2019. *Explain me this. Creativity, Competition and the Partial Productivity of Constructions*. Princeton: Princeton University Press.
- Goldberg, Adele E. & Ray Jackendoff. 2013. The English Resultative as a Family of Constructions. *Language* 80. 532–568.
- Häusermann, Jürg. 1977. *Phraseologie. Hauptprobleme der deutschen Phraseologie auf der Basis sowjetischer Forschungsergebnisse*. Tübingen: Niemeyer.
- Hilpert, Martin. 2014. *Construction Grammar and its application to English*. Edinburgh: Edinburgh University Press.
- Hoffman, Thomas & Alexander Bergs. 2018. A Construction Grammar Approach to Genre. *CogniTextes* 18. journals.openedition.org/cognitextes/1032. (accessed 10 October 2019).
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar*. Vol. 1: *Theoretical Prerequisites*. Stanford: Stanford University Press.
- Leibniz-Institut für Deutsche Sprache. 2019. *Deutsches Referenzkorpus*. Mannheim: Leibniz-Institut für Deutsche Sprache. <http://www.ids-mannheim.de/DeReKo> (accessed 2 October 2019).
- lexpan. Lexical Pattern Analyzer. Ein Analysewerkzeug zur Untersuchung syntagmatischer Strukturen auf der Basis von Korpusdaten*. Developed by the project „Usuelle Wortverbindungen“. Mannheim: Leibniz-Institut für Deutsche Sprache. <http://www.ids-mannheim.de/lexpan/> (accessed 2 October 2019).
- Mellado Blanco, Carmen. 2015. Phrasem-Konstruktionen und lexikalische Idiom-Varianten: der Fall der komparativen Phraseme des Deutschen. In Stefan Engelberg, Meike Meliss, Kris-

- tel Proost & Edeltraud Winkler (eds.), *Argumentstruktur zwischen Valenz und Konstruktion*, 217–235. Tübingen: Narr.
- Mollica, Fabio. 2014. Der Dativus ethicus im Deutschen aus konstruktionsgrammatischer Sicht. *Zeitschrift für germanistische Linguistik* 42. 349–378.
- Mollica, Fabio. 2020. Die Phrasem-Konstruktion [X_{NPnom} sein_{Kopula} mir_{Exp} (Modalpartikel) Det_{Lein(e)} Y_{NP}]_{Exkl} und ihre Relationen innerhalb der Ethicus-Konstruktion und der Dativ-Familie. *Linguistische Berichte* 261. 47–83.
- Mollica, Fabio & Elmar Schafroth. 2018. Der Ausdruck der Intensivierung in komparativen Phrasem-Konstruktionen im Deutschen und im Italienischen: eine konstruktionsgrammatische Untersuchung. In Kathrin Steyer (ed.), *Sprachliche Verfestigung. Wortverbindungen, Muster, Phrasem-Konstruktionen*, 103–136. Tübingen: Narr.
- Proost, Kristel. 2015. Die SUCH-Konstruktion als eine Familie von Argumentstrukturmustern. *OPAL – Online publizierte Arbeiten zur Linguistik* 1. 33–78. [ids-pub.bszbw.de/frontdoor/deliver/index/docId/3719/file/Proost_Die_Such_Konstruktion_2015.pdf](https://pub.ids-bszbw.de/frontdoor/deliver/index/docId/3719/file/Proost_Die_Such_Konstruktion_2015.pdf) (accessed 2 October 2019).
- Proost, Kristel. 2017. The role of verbs and verb classes in identifying German search-construction. In Francisco José Ruiz de Mendoza Ibáñez, Alba Luzondo Oyón & Paula Pérez Sobrino (eds.), *Constructing Families of Constructions Analytical perspectives and theoretical challenges*, 17–51. Amsterdam & Philadelphia: Benjamins.
- Proost, Kristel & Edeltraud Winkler. 2015. Familienähnlichkeiten deutscher Argumentstrukturmuster. Definitionen und grundlegende Annahmen. *OPAL – Online publizierte Arbeiten zur Linguistik* 1. 3–12. <https://pub.ids-mannheim.de//laufend/opal/opal15-1.html> (accessed 2 October 2019).
- Rosch, Eleanor & Carolyn B. Mervis. 1975. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* 7. 573–605.
- Rostila, Jouni. 2012. Phraseologie und Konstruktionsgrammatik. Konstruktionsansätze zu präpositionalen Funktionsverbgefügen. In Michael Prinz & Ulrike Richter-Vapaatalo (eds.), *Idiome, Konstruktionen, „verblühte rede“*. Beiträge zur Geschichte der germanistischen Phraseologieforschung, 263–282. Stuttgart: S. Hirzel.
- Ruiz de Mendoza Ibáñez, José, Alba Luzondo Oyón & Paula Pérez Sobrino (eds.). 2017. *Constructing Families of Constructions*. Amsterdam & Philadelphia: Benjamins.
- Schafroth, Elmar. 2015. Italian Phrasemes as Constructions: How to Understand and Use Them. *Journal of Social Sciences* 11. 317–337. thesaipub.com/abstract/10.3844/jssp.2015.317.337 (accessed 10 October 2019).
- Staffeldt, Sven. 2011. Die phraseologische Konstruktionsfamilie [X Präp Hand Verb]. *Zeitschrift für germanistische Linguistik* 39. 188–216.
- Staffeldt, Sven. 2018. *Gebrauchsemantik von Hand. Korpusbasierte Studien zu somatischen Phraseologismen des Deutschen mit der Konstituente Hand*. Tübingen: Stauffenburg.
- Stefanowitsch, Anatol. 2006. Konstruktionsgrammatik und Korpuslinguistik. In Kerstin Fischer & Anatol Stefanowitsch (eds.), *Konstruktionsgrammatik. Von der Anwendung zur Theorie*, 151–176. Tübingen: Stauffenburg.
- Stein, Stephan & Sören Stumpf. 2019. *Muster in Sprache und Kommunikation. Eine Einführung in Konzepte sprachlicher Vorgeformtheit*. Berlin: Schmidt.
- Steyer, Kathrin. 2011. Von der sprachlichen Oberfläche zum Muster. Zur qualitativen Interpretation syntagmatischer Profile. *Travaux neuchâtelois de linguistique* 55. 219–239.
- Steyer, Kathrin. 2012. Sprichwortstatus, Frequenz, Musterbildung. Parömiologische Fragen im Lichte korpusmethodischer Empirie. In Kathrin Steyer (ed.), *Sprichwörter multilingual*.

- Theoretische, empirische und angewandte Aspekte der modernen Parömiologie*, 287–314. Tübingen: Narr.
- Steyer, Kathrin. 2013. *Usuelle Wortverbindungen. Zentrale Muster des Sprachgebrauchs aus korpusanalytischer Sicht*. Tübingen: Narr.
- Stumpf, Sören. 2015a. *Formelhafte (Ir-)Regularitäten. Korpuslinguistische Befunde und sprachtheoretische Überlegungen*. Frankfurt am Main: Lang.
- Stumpf, Sören. 2015b. *Kann Jogi Weltmeister?* – Phraseologische und konstruktionsgrammatische Überlegungen zu einer aus (laien-)sprachkritischer Sicht „agrammatischen“ Konstruktion. *Aptum. Zeitschrift für Sprachkritik und Sprachkultur* 11. 1–20.
- Stumpf, Sören. 2017. *Phraseologie pur* – Die Konstruktion $X_{[Nomen]} pur$ als produktive und keineswegs ungrammatische Modellbildung. *Deutsche Sprache* 45. 317–334.
- Stumpf, Sören. 2021. *Star Wars ist eben mehr so ein Männerding*. Eine Korpusuntersuchung zur Syntax, Semantik und Pragmatik der Konstruktion [X Kopula ein Y-Ding]. *Deutsche Sprache* 49. 24–45.
- Traugott, Elizabeth Closs. 2008. Grammaticalization, constructions and the incremental development of language: suggestions for the development of degree modifiers in English. In Regine Eckhardt, Gerhard Jäger & Tonjes Veenstra (eds.), *Variation, selection, development: probing the evolutionary model of language change*, 219–250. Berlin & New York: de Gruyter.
- Wittgenstein, Ludwig. 1955. *Philosophical investigations*. Oxford: Blackwell.
- Ziem, Alexander. 2018a. Construction Grammar meets Phraseology: eine Standortbestimmung. *Linguistik online* 90. 3–19. bop.unibe.ch/linguistik-online/article/view/4316 (accessed 3 October 2019).
- Ziem, Alexander. 2018b. *Tag für Tag Arbeit über Arbeit*: konstruktionsgrammatische Zugänge zu Reduplikationsstrukturen im Deutschen. In Kathrin Steyer (ed.), *Sprachliche Verfestigung. Wortverbindungen, Muster, Phrasem-Konstruktionen*, 25–48. Tübingen: Narr.
- Ziem, Alexander & Alexander Lasch. 2013. *Konstruktionsgrammatik. Konzepte und Grundlagen gebrauchsbasierter Ansätze*. Berlin & Boston: de Gruyter.

Elizaveta Onufrieva

Constructional phrasemes in Modern Greek: structure and meaning

Abstract: This study aims to explore the class of constructional phrasemes in the Modern Greek language. The analysis of a group of Modern Greek constructional phrasemes shows that they fall into three structural types, depending on the type of lexemes involved in their fixed component. The differences between the types of constructional phrasemes suggest that they should be positioned on the syntax-lexicon continuum in an extended rather than a local way. Different constructional phrasemes may have a similar typical meaning and form in this way distinct semantic groups, some of which are discussed in this paper. The study concludes with a description of two Modern Greek constructional phrasemes denoting multitude.

Keywords: constructional phrasemes, phraseology, Modern Greek.

1 Introduction

The term “constructional phrasemes”, or “phraseme-constructions” (*Phrasem-Konstruktionen* in German, *frazeologizmy-konstruktsii* in Russian), proposed by Dobrovol’skij (2011: 114), is considered by some linguists to be the most suitable one to describe the structures that occupy on the syntax-lexicon continuum an intermediate point between fully lexically specified constructions, traditionally studied in phraseology, and free syntactic constructions, pertaining to the domain of syntax (Mollica and Schafroth 2018: 104).

Constructional phrasemes are included in the classification of phraseological units by Baranov and Dobrovol’skij (2016: 69) as a separate class of phraseologisms along with idioms, collocations, proverbs, grammatical phrasemes and situational clichés. According to the definition provided by these scholars (2016: 88), such phrasemes constitute “syntactically autonomous expressions with a fixed composition, in which certain slots have to be filled (with non-propositional actants X, Y or propositional actants P). The fixed elements of a constructional phraseme, along with its syntax, have a joint meaning, close to a

Elizaveta Onufrieva, Lomonosov Moscow State University, lisa.onufrieva@gmail.com

<https://doi.org/10.1515/9783110520569-005>

lexical one.¹”

Constructional phrasemes consist of a fixed component and a variable one and can be conceived of as “patterns with ‘open slots’” (Baranov and Dobrovol’skij 2016: 86). The fixed component is reproduced in speech as a prefabricated unit of language and determines the typical meaning of the construction – the meaning which remains the same in every distinct context where the construction occurs. The variable component is not lexically specified and is inserted into the construction depending on the context and the speaker’s intention.

The meaning conveyed by the fixed component of a constructional phraseme is, as a rule, clear to the speaker of the language even if the empty slots are labelled with symbols (X / Y in case of non-propositional actants or P in case of propositional actants), rather than filled with specific lexemes. At the same time, the phraseme becomes a complete utterance and expresses a complete thought only after its fixed component is filled with a certain lexeme or a certain proposition. In other words, “constructional phrasemes are similar to what is called ‘propositional form’ in logic – a proposition, the terms of which (not necessarily all) are replaced by variables” (Baranov and Dobrovol’skij 2016: 86).

One of the crucial characteristics of constructional phrasemes is that there are semantic constraints imposed by the fixed component on fillers of the open slot. It may be said that the open slot of a construction has a meaning of its own, which is set for it by the fixed component of the construction:

... a word may occur in a construction if it is semantically compatible with the meaning of the construction (or, more precisely, with the meaning assigned by the construction to the particular slot in which the word appears).

Stefanowitsch and Gries (2003: 213)

Idiomaticity of a constructional phraseme is confined only to its fixed component, while the lexical variables are used in their literal meaning. This is an important defining criterion that allows us to distinguish constructional phrasemes from lexical idioms built on the basis of one and the same structural model.

Structures that fall within the class of constructional phrasemes by Baranov and Dobrovol’skij have long held the attention of Russian scholars, starting with Shvedova (1958, 1960) and Shmelev (1960). More recently, interest in such structures has been stimulated by the development of machine translation systems and the advent of text corpora.

1 All the translations in this paper are the author’s, unless otherwise indicated.

Most studies on phraseologised syntactic constructions come from syntactic theories and approaches, in particular from the field of research into the semantics of syntax and from Construction Grammar as a syntax-oriented kind of Cognitive linguistics. On the part of phraseologists, there are considerably fewer studies on phraseologised syntactic structures, since phraseology traditionally had a greater interest in “classical” phraseologisms – idioms, proverbs, etc., – while phraseologised syntactic units were largely neglected and not included in phraseological classifications. As Dobrovolskij points out:

so far phraseologists have not given due attention to the units of this type: the semantics of such constructions has been described quite roughly, while there has been practically no research on the constraints that affect empty slots and on respective selective preferences.

Dobrovolskij (2016: 18)

Phraseologisms in general and constructional phrasemes in particular are one of the most theoretically under-explored aspects of the Modern Greek linguistics. As Tresorukova notes:

in the Greek linguistic tradition, there does not exist a unified semantic classification of phraseological units, and the number of scholars concerned with theoretical issues of phraseology is insignificant.

Tresorukova (2015: 556)

The few studies providing original classifications of Modern Greek phraseologisms (e.g. by Motsiou) make no reference to syntactic idioms. Mini (2009: 57), while describing the basic tenets of Construction Grammar, though mentions Fillmore’s lexically open idioms and provides a Greek translation for the term (*λεξικώς ανοιχτές φράσεις*), does not give an example of such idioms in Modern Greek.²

Constructional phrasemes of the Modern Greek language have no full description in either dictionaries or in grammars. Some Modern Greek constructional phrasemes are missing from dictionaries. Many others are included in dictionaries with their slots filled with a certain actant, which does not allow the user of the dictionary to understand that there is a productive phraseological pattern behind a seemingly fixed construction. Occasionally, constructional

² At the same time, some Greek linguists note the irregular syntax of many set expressions and use such terms as “schemas”, i.e. notions that are closely related to syntactic idioms. Namely, Setatos in some of his studies (Setatos 1996–1997: 234) refers to “schemas” (*σχήματα*) existing in Modern Greek.

phrasemes, in particular those consisting solely of function words, are described in manuals on syntax of the Modern Greek language as special, exceptional cases of irregular behavior of function words.

The absence of systematic descriptions of constructional phrasemes in Modern Greek suggests the need of a dedicated study.

2 Constructional phrasemes in Modern Greek

2.1 Research focus, methodology and data sources

The purpose of the present paper is to explore the nature of constructional phrasemes of the Modern Greek language and, in particular, to investigate their structure and semantics.

A total of fifty Modern Greek constructional phrasemes were selected for this study. The principal sources were the main monolingual dictionaries of Modern Greek, such as the *Dictionary of Standard Modern Greek* by the Institute for Modern Greek Studies of the Aristotle University of Thessaloniki (2017), the *Dictionary of Modern Greek* by Babiniotis (2005) and the *Practical Dictionary of Modern Greek* by the Academy of Athens (ed. Charalambakis, 2014). Also, the works by Skartsis (1997) and Katos (2017) were used, as well as the data from the online dictionaries *slang.gr* and *el.wiktionary.org*. Apart from the dictionaries of Modern Greek, constructional phrasemes were collected from the texts of Greek literary works along with their translations into other languages, as well as from literary texts, translated into Greek from other languages.

The analysis of the constructions found was carried out using the *Greek Web Corpus (elTenTen14)*, a monolingual corpus of Modern Greek which is available at <https://the.sketchengine.co.uk>. The present corpus, created in August 2014, is made up of texts collected from the Internet and is currently the largest corpus of Modern Greek with more than 1.6 billion words.³ The corpus was used to count frequency of the constructions found and to analyze fillers of their open slots, as well as to determine the typical meaning of the constructions on the basis of multiple contexts where they occur and to deduce their additional semantic nuances.

³ Compare to the *National Corpus of Modern Greek Language (ΕΘΕΓ)*, which currently contains only about 52.2 million words.

2.2 Structure of Modern Greek constructional phrasemes

2.2.1 Structural types

Modern Greek constructional phrasemes differ with regard to the nature of elements involved in their fixed component. The analysis of the fixed component of the selected constructional phrasemes shows that they fall into three structural types.

The first type involves phrasemes, whose fixed component consists solely of function words (conjunctions, particles, prepositions) and / or pronouns (see Table 1).

Tab. 1: Constructional phrasemes whose fixed component consists solely of function words and / or pronouns

Constructional phraseme	Literal translation	Generalised meaning
$Ti X / P_1 \tau i Y / P_2$	what X what Y	equality of X / P ₁ and Y / P ₂
$K\acute{\alpha}\theta\epsilon X \kappa\alpha\iota Y$	every X and Y	permanence of Y with respect to X
$A\upsilon\tau\acute{o}\varsigma \kappa\iota \alpha\nu (\delta\epsilon\nu) P$	he / this and if (not) P	intensiveness of P

The idiomaticity of constructional phrasemes of the first type is exclusively based on their deviant syntax, e.g. reduction of certain lexemes (some constructional phrasemes of this type constitute reduced syntactic schemata), irregular placement of function words in the fixed component or idiosyncratic use of function words.

For instance, the interrogative pronoun τi in Modern Greek is normally used in questions and exclamations (similar to the English *what*), but in the repetitive construction $Ti X / P_1 \tau i Y / P_2$ the pronoun loses its interrogative force, and the construction begins to convey the meaning of equality between two (or more) objects, as in (1).

- (1) ***Ti χωρισμός, τι θάνατος. Η λέξη αλλάζει μόνο.***

What separation, what death. The word changes only.

‘Separation and death are the same. Only the words are different.’

(elTenTen14, 1925367471)

The meaning of equality is not typically expressed by the pronoun τi , and some dictionaries (e.g. the Dictionary of the Academy of Athens) do not list it among

other meanings of *τι*. Therefore, it might be said that we are dealing here with a case of idiosyncratic use of a function word.

The phraseme *Κάθε X και Y* typically conveys the meaning of permanent co-existence of two objects, actions or phenomena, and their mutual dependence, which is illustrated in (2) – *κάθε τηλέφωνο και βίζιτα* (literally, ‘every phone and visit’).

- (2) *Τελικά είχαμε καταλήξει στο συμπέρασμα πως μάλλον δε θα δεχόταν εδώ στο σπίτι τους εραστής της. Το πιθανότερο να πήγαινε εκείνη στα σπίτια τους. «Κάθε τηλέφωνο και βίζιτα» λέγαμε. (Κ. Μουρσελάς, Βαμμένα κόκκινα μαλλιά)*

‘In the end, we concluded that she didn’t entertain her boyfriends in her home. More likely she went to their place. **“Every phone call is a date,”** we said.’

(Κ. Μουρσελάς, *Βαμμένα κόκκινα μαλλιά* / K. Mourselas, *Red Dyed Hair*, transl. from Greek: F. A. Reed)

The phraseme *Κάθε X και Y*, marked “phraseologism” in the dictionary (Charalambakis 2014: 737), is considered by the Greek syntactician Tzartanos (2016: 160) to be a case of lexical reduction and simultaneously a case of special use of the conjunction *και* (‘and’).

Another example of a constructional phraseme of the first type is the construction *Αυτός κι αν (δεν) P*, typically denoting intensiveness of some action or state, as in (3).

- (3) ***Αυτός κι αν έχει ταλέντο στις πωλήσεις.***

He and if has a talent for sales.

‘He has a great talent for sales.’

(eITenTen14, 294577510)

The placement of the constituents in the fixed component of the phraseme *Αυτός κι αν (δεν) P* is specific to that particular construction and, therefore, cannot be considered regular.

Constructional phrasemes of the first group partially correspond to the structures that are included by the *Russian Academic Grammar* (Shvedova 1980: 382) into the class of syntactic idioms as a subclass under the name of “sentences with a phraseologised structure.”

The second structural type involves constructional phrasemes, whose fixed component contains semantically bleached content words (see Table 2).

Tab. 2: Constructional phrasemes whose fixed component contains semantically bleached content words

Constructional phraseme	Literal translation	Generalised meaning
<i>Σιγά (να) μην (και) P</i>	quietly (to) not (and) P	negation of P
<i>Μη σώσεις και P</i>	don't manage and P	lack of interest on the part of the speaker in the action, expressed by P

In contrast to the constructional phrasemes of the first type, whose fixed component consists exclusively of function words and / or pronouns, constructional phrasemes of the second type may contain in their fixed part words that formally pertain to the category of content ones, but these words in this particular case have lost (partially or fully) their lexical meaning and / or part-of-speech characteristics. Phrasemes of the second type are highly non-compositional and are often characterised by highly irregular syntax, while their possible prototypes – structures with regular syntax – hardly allow for hypothetical reconstruction.

For example, the Modern Greek word *σιγά* is an adverb of manner that typically means ‘quietly, without making much noise’ or ‘slowly’ (Charalambakis 2014: 1431). However, in the phraseme *Σιγά (να) μην (και) P*, this adverb loses its semantic and part-of-speech characteristics, as it becomes a part of a structure that is used to produce polemic negation, as in (4).

- (4) *Αμ, τότε για ποιον ήρτε; Για την Ουράνα; Σιγά να μην ήρτε γι' αυτήν!*
 Eh, then who did he come for? For Urana? *Quietly* to not came for her!
 ‘Eh, then who did he come for? For Urana? **Fat chance that he came to see her!**
 (Μ. Λουντέμης, *Ένα παιδί μετράει τ'άστρα*)

The phraseme *Μη σώσεις και P* can serve as another example of the second type of constructional phrasemes. The Greek word *σώσεις* is a form of the colloquial verb *σώνω*, which in turn has its origins in the verb *σώζω* with the main dictionary meaning ‘to save’ (Charalambakis 2014: 1551, 1553). The verb *σώνω* has developed its own meanings over time, but most of them are manifested only in set expressions, e.g. the *Dictionary of the Academy of Athens* lists under the lemma of *σώνω* idioms only. The phraseme *Μη σώσεις και P* is typically used to indicate the speaker’s complete disregard for the action or state expressed by the variable P, as in (5).

- (5) ... και απειλήσε να μην πάει στο Σότσι για τους G8. **Μη σώσει και πάει!**
 ... and threatened not to go to Sochi to the G8. Not *manage*_{-3SG} and *go*_{-3SG}!
 ‘... and threatened not to go to Sochi to the G8. **I couldn’t care less if he doesn’t go!**’
 (elTenTen14, 370496048)

As the words *σιγά* and *σώσεις* have evidently lost their meaning, it is hardly possible to provide an adequate literal translation for the constructions *Σιγά (να) μην (και) P* and *Μη σώσεις και P* in (4–5).

The second type of constructional phrasemes partially corresponds to the structures that are referred to by the *Russian Academic Grammar* (Shvedova 1980: 217) as “formations in which the grammatical or literal lexical meanings of the components expressing certain subjective-modal meanings have been lost or weakened.”

Finally, the third group includes constructional phrasemes, whose fixed component contains content words (see Table 3).

Tab. 3: Constructional phrasemes whose fixed component contains content words

Constructional phraseme	Literal translation	Generalised meaning
<i>Να Χ να μάλαμα</i>	here X here gold	negative evaluation of X
<i>Πιο Χ πεθαίνεις</i>	more X you die	a high degree of the quality denoted by the variable X
<i>Χ και τα μάτια σου</i>	X and your eyes	the speaker’s request to pay attention to X

Constructional phrasemes of the third type may also exhibit certain syntactic particularities, but their idiomaticity has to do not so much with irregular syntax, as with the specifics of lexical elements involved in their fixed component. Velichko (2016: 63), describing syntactic phraseologised structures of the Russian language, points out that they, as opposed to lexical phraseologisms, “are devoid of imagery.” Indeed, lack of imagery is typical of constructional phrasemes of the first and the second types, while the same cannot be said of the phrasemes of the third type.

Constructional phrasemes of the third type are not devoid of imagery. Their fixed component contains one or more content words, but these are not semantically bleached (as is the case with phrasemes of the second type), but, on the contrary, are metaphorically reinterpreted and create bright imagery. It is im-

agery that distinguishes the third group of constructional phrasemes from the previous groups.

Constructional phrasemes of the third type are close to lexical idioms. Metaphorically reconsidered lexical constituents of these constructional phrasemes determine their figurative inner form and by means of the inner form their actual meaning. What distinguishes constructional phrasemes of the third type from lexical idioms is that the former constitute productive patterns.

For instance, the noun *μάλαμα* ('gold') in the fixed part of the phraseme *Να X να μάλαμα!* is used as an "ideal" evaluative stereotype. The phraseme compares the lexeme that fills the open slot with gold and conveys the meaning of negative evaluation of X as of an object that does not live up to the evaluative stereotype, as in (6).

- (6) ... ο Γιωργάκης δεν καταλαβαίνει τίποτα. **Να αρχηγός, να μάλαμα.**
 ... Yorgakis does not understand anything. Here leader, here gold.
 '... Little Yorgos understands nothing. **Some leader!**'
 (elTenTen14, 26862905)

The verb *πεθαίνεις* ('die') in the phraseme *Πιο X πεθαίνεις* serves as an illustration of the physical limit of a person's capacities, and such an inner form conveys the meaning of the highest degree of the quality denoted by the variable X, as in (7).

- (7) *Γιατί τα άρθρα σας είναι ανυπόγραφα; Πιο κίτρινος τύπος πεθαίνεις.*
 Why are your articles unsigned? More yellow press die._{2SG}.
 'Why are your articles unsigned? **That's the worst yellow press possible.**'
 (elTenTen14, 424612367)

The lexeme *μάτια* ('eyes') in the phraseme *X και τα μάτια σου* evokes the picture of a steady look, and such an inner form expresses the actual meaning of the need to pay attention to X or to care for X.

- (8) **Το ευρώ και τα μάτια σας.** Χωρίς ευρώ, δεν θα έχουμε ψωμί... .
 The euro and the eyes your. Without the euro, we will not have bread... .
 '**Take care of the euro.** Without the euro, we won't have bread... .'
 (elTenTen14, 263382091)

As demonstrated above, Modern Greek constructional phrasemes differ as to how they are structured. Nonetheless, all of them, different with respect to their structure and degree of syntactic deviance, possess a semantically restricted variable component, which makes it possible to include them all in one class of phraseologisms.

2.2.2 Modern Greek constructional phrasemes on the syntax-lexicon continuum

The notion of the syntax-lexicon continuum is used in Cognitive Grammar and Construction Grammar to indicate the absence of sharp boundaries between lexis and syntax. The syntax-lexicon continuum, in our case, can be represented as a scale that has purely schematic abstract constructions at one extreme (lexically non-specified schemas) and lexically specified phraseologisms at the other.

Constructional phrasemes, as partially filled schemas, occupy on this scale an intermediate place between fully lexically specified constructions, traditionally studied in phraseology, and free syntactic constructions, pertaining to the domain of syntax (Mollica and Schafroth 2018: 104; Benigni et al. 2015: 282) (see Figure 1).

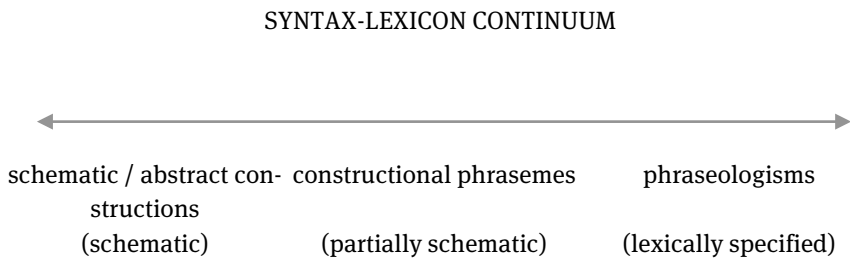


Fig. 1: Continuum between schematic / abstract constructions and phraseologisms (Stathi 2011: 151; Benigni et al. 2015: 282)

The structural and semantic differences between the types of constructional phrasemes in Modern Greek enable us to position them on the scale of the syntax-lexicon continuum extendedly rather than locally, i.e. different structural types of constructional phrasemes form a continuum of their own (see Figure 2).

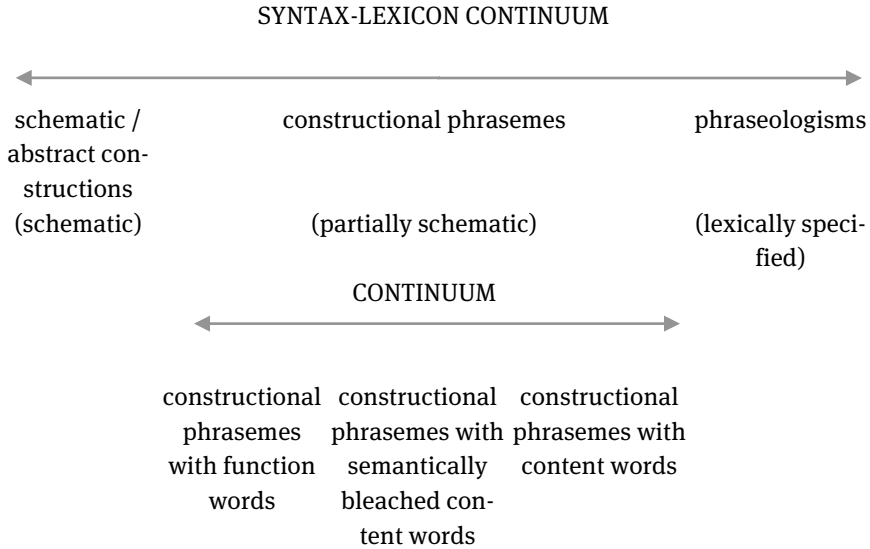


Fig. 2: Modern Greek constructional phrasemes on the syntax-lexicon continuum (Stathi 2011: 151; Benigni et al. 2015: 282)

Constructional phrasemes with function words in the fixed component will be closest on the syntax-lexicon continuum to schematic / abstract constructions, as their fixed component does not contain full content words and their idiomatity is based mostly on their syntactic non-compositionality (expressed to varying degrees). Constructional phrasemes that have content words in their fixed part are closer to the lexical end of the continuum. Like idioms, such constructional phrasemes have reinterpreted lexical elements in the fixed part. Constructional phrasemes with semantically bleached content words occupy an intermediate position between the first two types of phrasemes. Since their constituents have lost partially or fully their lexical meaning, the contribution they make to the overall meaning of these constructions is less evident. All the three types of constructional phrasemes share the property of having a semantically restricted variable component and irregular syntax (to different extents).

2.3 Semantics of Modern Greek constructional phrasemes

According to the definition of constructional phrasemes provided by Baranov and Dobrovol'skij (2016: 88), “the fixed elements of a constructional phraseme, along with its syntax, have a joint meaning, close to a lexical one.”

As Apresian (2015: 96) notes, the evolution of constructions from standard to non-standard and further on to phraseologisms is accompanied by “the narrowing of the meaning.” The analysis of the Modern Greek constructional phrasemes shows that every constructional phraseme has one permanent typical meaning, set by its fixed component for all cases of its use. In my view, this is an important characteristic that distinguishes constructional phrasemes from other structures with irregular syntax that are often polysemous.

The typical meaning of a constructional phraseme, along with its semantic nuances, can be deduced on the basis of multiple contexts where the phraseme occurs, which presupposes the use of the corpus data. A considerable problem arises in distinguishing between constructional phrasemes and homonymous structures of regular syntax. Syntactic homonymy considerably slows down the process of establishing frequencies of occurrence of a particular constructional phraseme in the corpus and requires that each context be read individually. The group that poses the most difficulty is that of phrasemes that consist solely of function words, as in this case the percentage of homonymous patterns may exceed 90%.

Different constructional phrasemes may have a similar typical meaning and form in this way distinct semantic groups. The selected constructional phrasemes of the Modern Greek language can be divided into various semantic groups, some of which are provided below.

1. Evaluation.

The phrasemes in this group are judgments about whether “something is good or bad” (Hunston 2004: 157). For instance, the phrasemes *X να σου πετύχει* and *Χαρά σε X* always express negative evaluation of X, as of an object that does not live up to the evaluative stereotype (see examples 9–10).

(9) *Ακούς τον αθεόφοβο, καλόγερος να σου πετύχει!*

Listen to that unholy, monk to you happen_{-3SG!}

‘Just look at that unholy man, **some monk!**’

(N. Καζαντζακής, *Ο Χριστός ξανασταυρώνεται*)

- (10) *Κοίτα τον κακομοίρη πώς καμπουρίζει, κοίτα κανιά, χαρά στο λεβέντη!*
 Look at that poor guy, he is slouching, look at these legs, joy to the handsome fellow!
 ‘Look at that poor guy, he is slouching, look at these legs. **Hardly a handsome man!**’
 (N. Καζαντζακής, *Ο Χριστός ξανασταυρώνεται*)

2. High degree of quality.

Another large group of Modern Greek constructional phrasemes is the phrasemes that typically denote a high degree of some quality expressed by the variable X. One of these is the phraseme *Εκεί να δεις X*, illustrated in (11).

- (11) *Έγινα φορτοεκφορτωτής. Εκεί να δεις ευτυχία!*
 I became a freight handler. There to see-2SG happiness!
 ‘I got a job as a freight handler. **Real happiness!**’
 (S. Alexievich, *Το τέλος του κόκκινου ανθρώπου*, transl. from Russian: A. Ιωαννίδου)

Another example of a constructional phraseme denoting a high degree of some quality is the construction *Πιο X πεθαίνεις*, which is evidently a constructional calque from French (*Plus X que Y tu meurs*) or Italian (*Più X di Y si muore*). The construction imposes semantic constraints on fillers of its open slot, which have to denote some gradable quality. When the open slot is filled with lexemes belonging to the parts of speech that do not normally denote gradable quality (e.g. a verb or a numeral), the construction coerces the meaning of these lexemes into the meaning of gradable quality, as in (12).

- (12) *Πιο 1984 πεθαίνεις.*
 More 1984 die-2SG.
 ‘**That’s pure 1984.**’
 (elTenTen14, 686923709)

3. Multitude.

The phrasemes in this semantic group typically denote a large number of objects, indicated by the variable X, as, for example, the phraseme *X να δουν τα*

μάτια σου (13). Two phrasemes of this group will be further discussed in section 3.

- (13) *Χιλιάδες μουσεία. Αμέτρητα! ... Μουσεία να δουν τα μάτια σου.*
Thousands of museums! Countless! ... Museums to see the eyes your.
'Thousands of museums! Countless! ... **You got to see the number of museums there.**'
(elTenTen14, 538532309)

4. Equality.

An example of a phraseme with this meaning is the Modern Greek constructional phraseme *Τι X / P₁ τι Y / P₂*, which denotes equivalence between two different objects or two different actions or states (see example in 14).

- (14) *Αν δεν αγαπάς, τι να ζεις, τι να μη ζεις.*
If you don't love, what to live-2SG what to not live-2SG.
'If you don't love, **you might as well be dead.**'
(Κ. Μουρσελάς, *Βαμμένα κόκκινα μαλλιά*)

5. Permanence of a quality.

The phrasemes in this group denote permanence of some quality or action, as, for example, the phraseme *X τον ανεβάζω, X / Y τον κατεβάζω* that conveys the meaning of a permanent characteristic (15).

- (15) *Ανήθικο τον ανέβαζε, τρακαδόρο τον κατέβαζε.*
Unethical him lifted up, moocher him took down.
'**She constantly called him unethical and a moocher.**'
(Κ. Μουρσελάς, *Βαμμένα κόκκινα μαλλιά*)

6. Negation.

Finally, a large group of Modern Greek constructional phrasemes includes the phrasemes that typically denote negation of different kinds. An example of these is the phraseme *Πολύ που* P (16).

(16) *Πολύ που τον τρέμω.*

Much that him tremble_{1SG}.

'I am in no way afraid of him.'

(Μ. Λουντέμης, *Ένα παιδί μετράει τ'άστρα*)

Phrasemes inside each semantic group, having a similar meaning, are not necessarily absolute synonyms, since apart from the typical meaning every constructional phraseme has individual semantic nuances.

The grouping based on semantics can be very efficient in foreign language teaching and learning. Velichko, who has for many years been studying syntactic phraseology of the Russian language and teaching Russian as a foreign language, believes that a classification of phraseologised syntactic constructions based on semantics is much preferable to one based on structural characteristics:

The most efficient and useful classification seems the semantic one, which takes into account the cognitive approach to the language Semantic classification is important for a better theoretical understanding of phraseologized structures, as it makes it possible to define more precisely their semantic specifics and function in speech.

Velichko (2016: 124)

The semantic grouping listed above is based on the current sample of phrasemes and is subject to change as empirical data accumulate. It can also vary depending on the researcher's needs.

3 Analysis of Modern Greek constructional phrasemes denoting multitude

In this section I undertake an analysis of two Modern Greek constructional phrasemes that denote multitude:

- *Από X άλλο τίποτα*
*as for X else nothing

- *X να δουν τα μάτια σου*
*X to see the eyes your

The analysis of the two phrasemes consisted of several steps:

- to determine the structure of the fixed component;
- to see the lexemes that fill the slot and to count their frequency;
- to find semantic and part-of-speech restrictions that are imposed by the fixed component on fillers of the open slot;
- to determine the typical meaning of the construction on the basis of multiple contexts where the construction occurs.

3.1 The constructional phraseme *Από X άλλο τίποτα*

The fixed component of the construction *Από X άλλο τίποτα* consists of three lexical elements (“lexical anchors”): the preposition *από*, the pronoun *άλλο* (‘other / else’) and the pronoun *τίποτα* (sometimes *τίποτε*) (‘nothing’).

It can be assumed that the preposition *από* in the construction is a reduced form of the adverbial preposition *εκτός από* or *πέρα(ν) από* (‘apart from’, ‘except for’). In this case, the construction possibly originates from the regular Modern Greek sentence *Εκτός από X δεν έχουμε τίποτα άλλο* – ‘Apart from X we do not have anything else’ / ‘Except for X we do not have anything else.’ Alternatively, the preposition *από* may be used in the construction in one of its own dictionary meanings, which is ‘as for,’ ‘with respect to,’ and serves to restrict the focus to a particular object or person. In this case, the prototypical construction may be *Από X – δεν έχουμε τίποτα άλλο* – ‘As for X, we do not have anything else.’ The second version seems more convincing and below I will translate *από* as ‘as for.’

The phraseme *Από X άλλο τίποτα* is included in dictionaries of Modern Greek. The *Dictionary of the Academy of Athens* (2014: 96) lists it under the lemma of *άλλος* (‘other’) with a note “Φρ.” (‘phrase’) and an omission mark in place of the only variable. The definition provided by this dictionary for the phraseme is ‘about something existing in plenty.’ The meaning of multitude, conveyed by the phraseme, comes with that of intensification. Another dictionary, the *Dictionary of Standard Modern Greek* (2017), lists the construction under the lemma of *τίποτε* (‘nothing’) with the definition ‘to indicate that something that is denoted by the noun exists to a high degree.’

The corpus *elTenTen14* yielded a total of 847 contexts with the phraseme *Από X άλλο τίποτα*. The most frequent lexical variables in the construction are shown in Table 4 in the nominative case.

Tab. 4: Frequency of lexemes that fill the open slot of the construction *Από Χ άλλο τίποτα*

Word	Translation	Number of occurrences
λόγια	words	24
τέτοιος	such	17
ιδέες	ideas	16
αυτός	this	12
δικαιολογίες	excuses	12
θεωρία, -ες	theory, -ies	11
υπομονή	patience	10
επιλογές	choices	8
όνειρα	dreams	8
όρεξη	appetite	8
προβλήματα	problems	8
φαντασία	fantasy	8
ευκαιρίες	opportunities	7
ιστορίες	stories	6
ποικιλία, -ες	diversity	6

The open slot of the construction in most cases is filled with nouns (singular or plural) and pronouns. Based on corpus data, the construction can be schema-tised as follows:

από + Noun / Noun Phrase / Pronoun + *άλλο* + *τίποτα*

The lexemes that fill the valency slot of the construction denote abstract notions (e.g. *υπομονή* – ‘patience’, *όρεξη* – ‘appetite’), as well as the groups of animate or inanimate objects that share the same characteristics (e.g. *λόγια* – ‘words’, *προβλήματα* – ‘problems’).

The following examples from the corpus demonstrate the meaning of multi-tude expressed by the phraseme:

(17) **Από θεωρίες, άλλο τίποτα.** *Ο καθένας έχει τη δικιά του.*

As for theories, else nothing. Everybody has one of their own.

‘As for theories, there are lots of them. Everybody has one of their own.’

(elTenTen14, 1590305676)

- (18) *Κι από παρουσίες σταρ, άλλο τίποτα: από Γιόκο Όνο και Λάιζα Μινέλι μέχρι Ζακ Έφρον ...*
 And as for stars' appearances, else nothing: from Yoko Ono and Liza Minnelli to Zac Efron ...
 'And as far as stars' appearances are concerned, there were plenty of **them**: from Yoko Ono and Liza Minnelli to Zac Efron ...'
 (elTenTen14, 1640209569)

When its variable component is filled, the construction can become a rather extended utterance. In (19), the open slot of the construction is filled with 21 words.

- (19) *Από φανφάρες του τύπου «τουρισμός και πολιτισμός είναι η βαριά βιομηχανία της Ελλάδας, το όχημα που θα μας βγάλει από την κρίση», άλλο τίποτα!*
 As for highfalutin statements like “tourism and culture are the heavy industry of Greece, the vehicle that will get us out of the crisis,” else nothing!
 ‘**There is nothing but highfalutin statements like “tourism and culture are the heavy industry of Greece, the vehicle that will get us out of the crisis”!**’
 (elTenTen14, 1625747866)

Occasionally, extra components can be introduced into the construction in the nominative case, in particular, those referring to the object or place being characterised, as in (20–21).

- (20) *Από παρατσούκλια οι Αιγινήτες άλλο τίποτα!*
 As for nicknames the Aeginians else nothing!
 ‘The Aeginians **have plenty of nicknames!**’
 (elTenTen14, 469055990)
- (21) *Από Ναπολέοντες αυτό το σπίτι άλλο τίποτα!*
 As for Napoleons this house else nothing!
 ‘In this house, **there is nothing but Napoleons!**’
 (elTenTen14, 1587128448)

Despite the possibility of such an extension of the construction by including extra components, the construction has a stable structure – the rearrangement

of the structural elements is not common (the variant form *Άλλο τίποτα από Χ* with the meaning of multitude occurred in the corpus 7 times only).

The phraseme *Από Χ άλλο τίποτα* typically denotes multitude, predominance, excess of a certain object or a phenomenon, while indirectly it implies lack of other, more desirable objects or phenomena. Quite often, the lexeme that fills the open slot has negative connotations, and the phraseme expresses a shade of negative evaluation (see examples 22–23 below).

(22) *Τον ρώτησα αν είχαν κανένα καβγαδάκι, πώς και έφυγε. — Εντάξει, **από καβγαδάκια άλλο τίποτα...*** (Κ. Μουρσελάς, *Βαμμένα κόκκινα μαλλιά*)
 ‘Maybe they’d had a little spat, I ask, maybe that was the reason. — Sure, **all the little spats you like ...**’ (Κ. Mourselas, *Red Dyed Hair*, transl. from Greek: F. A. Reed)

(23) ... *Άλλωστε **από παλιανθρώπους άλλο τίποτα** σε τούτο τον τόπο, Αμπρόσιο.* (Μ. Βάργκας Γιόσα, *Πότε πήραμε την κάτω βόλτα*; transl. from Spanish: Τ. Παναγοπούλου)
 “... Y además en este país **hay canallas para regalar ...**.” (Μ. Vargas Llosa, *Conversaciones en La Catedral*)
 “...Anyway, **there’s a surplus of swine** in this country” (Μ. Vargas Llosa, *Conversation in the Cathedral*, transl. from Spanish: G. Rabassa)

In other cases, positive evaluation is produced, which is illustrated below by (24), in particular by the use of the interjection *δόξα τω Θεώ* ‘thank goodness.’

(24) *Κι **από βιβλία** το σπίτι τους, *δόξα τω Θεώ, **άλλο τίποτα!***
 And as for books their house, Glory to God, else nothing!
 ‘And **as for books**, thank goodness, **there were plenty of them** in their house!’
 (Μ. Ιορδανίδου, *Λωξάντρα*)*

Judging by the contexts provided above, we can conclude that the phraseme may serve as a means of evaluation, either positive or negative, depending on the context where the construction occurs and / or the lexeme that fills the open slot.

3.2 The constructional phraseme *X να δουν τα μάτια σου*

The fixed part of the constructional phraseme *X να δουν τα μάτια σου* is represented by five lexemes: the multifunctional particle *να* ('to'), the verb *δουν* ('see', 3rd person plural dependent form), the definite article *τα*, the noun *μάτια* ('eyes') and the pronoun *σου / σας* ('your').

The phraseme is included in the dictionaries under the lemma of the verb *βλέπω* ('see') with a note “ΦΡ” ('phrase') and the following definitions: 'to indicate great abundance' (DSMG 2017), 'to show multitude of something' (Babiniotis 2005: 371), 'to underline the large number, multitude' (Charalambakis 2014: 330). The variable component in the three dictionaries bears an omission mark. The phraseme occurs in the corpus 66 times, being considerably less frequent than the previously described phraseme *Από X άλλο τίποτα*.

Table 5 shows the lexemes that are used in the corpus to fill the open slot of the construction more than once. The words are shown in the nominative.

Tab. 5: Frequency of lexemes that fill the open slot of the construction *X να δουν τα μάτια σου*

Word	Translation	Number of occurrences
<i>ξύλο</i>	whacking	3
<i>κόσμος</i>	people	2
<i>λεφτά</i>	money	2
<i>ουρές</i>	queues	2

The corpus data show that the open slot of the construction is always filled with a noun or a noun phrase. Based on this, the construction can be schematised as follows:

Noun / Noun Phrase + *να* + *δουν* + *τα* + *μάτια* + *σου / σας*

The following are examples from the corpus *elTenTen14*:

- (25) ... *βγάξαμε υπερ-κέρδη, – κέρδη να δουν τα μάτια σας!* – εκατοντάδες εκατομμυρίων ...
 ... we'd been earning superprofits, – profits to see the eyes your! – hundreds of millions ...

‘... we’d been earning superprofits, – **innumerable profits** – hundreds of millions...’

(elTenTen14, 134581194)

- (26) *Οι κάτοικοι του νησιού ήταν από νωρίς εκεί, **κοσμοσυρροή να δουν τα μάτια σου.***

The residents of the island were there from early morning, throng to see the eyes your.

‘The residents of the island were there from early morning, **you should have seen the number of people gathered.**’

(elTenTen14, 406984021)

The phraseme *X να δουν τα μάτια σου* has a rigid structural form – if the fixed and the variable components change places, the construction loses the meaning of multitude and is to be interpreted literally, as in (27).

- (27) *... λύσ’ τους, **να δουν τα μάτια μου τ’ αγαπητά μου αδέρφια.***

... free them, to see the eyes my my beloved brothers.’

‘... free them, **so that my eyes would see my beloved brothers.**’

(elTenTen14, 254570951)

The corpus does not provide contexts in which extra components are inserted into the construction, and the open slot is rarely filled with an extended phrase. Example 28 below shows one of the few contexts where the open slot of the construction is filled with an extended piece of language (10 words).

- (28) ***Αγωνιστές, αντιστασιακούς και αντιφασίστες της εκπάγλου κατοπινής πολύ “δημοκρατικής” μάσας να δουν τα μάτια σου.***

Fighters, oppositionists and antifascists of the dazzling posterior very “democratic” grub to see the eyes your.

‘**There were so many fighters, oppositionists and antifascists of the dazzling, very “democratic” grub that followed.**’

(elTenTen14, 425793667)

Occasionally, when the open slot of the construction is filled with two nouns at once, one of the nouns is placed separately at the end of the construction, as in (29).

(29) **Ψέμμα να δουν τα μάτια σου και υποκρισία.**

Lies to see the eyes your and hypocrisy.

‘So many lies and so much hypocrisy.’

(elTenTen14, 1456979365)

Similarly to the phraseme *Από X άλλο τίποτα*, the construction *X να δουν τα μάτια σου* may express, apart from the meaning of multitude, the meaning of intensification, which can be clearly observed in some contexts, such as (30).

(30) – **Κλάμα να δουν τα μάτια σου, ξάδελφε. Με λυγμούς.** (Κ. Μουρσελάς, *Βαμμένα κόκκινα μαλλιά*)

“– **Crying like you'd never believe**, cousin. Sobbing.” (Κ. Mourselas, *Red Dyed Hair*, transl. from Greek: Fred A.Reed)

The construction can also serve as a means of evaluation, either positive or negative, depending on the context and the lexeme that fills the open slot. Among the fillers of the open slot, we can see both lexemes with clearly negative connotations (e.g. *ξύλο* – ‘whacking,’ *ουρές* – ‘queues,’ *βία* – ‘violence,’ *ζητιανιά* – ‘beggary’) and lexemes with positive connotations (e.g. *χαρά και ευτυχία* – ‘joy and happiness’).

The analysis of the phrasemes examined in 3.1–3.2 leads to the following observations:

1. The two phrasemes show significantly different frequencies of use, which should be taken into account when teaching or learning Modern Greek as a foreign language.
2. The open slot of the phraseme *Από X άλλο τίποτα* may be filled with an extended piece of language (21 words in ex. 19), and besides, some extra components may intervene between the parts of the phraseme. The open slot of the second phraseme, *X να δουν τα μάτια σου*, is commonly filled with a single word or a shorter noun phrase (10 words in ex. 28), while the fixed and the variable components always co-occur adjacently, and no extra components can intervene between the parts of the phraseme.
3. Both phrasemes can serve as a means of either positive or negative evaluation, depending on the context where they occur and / or the lexeme that fills the open slot.
4. Both phrasemes, in addition to the meaning of multitude, may express the meaning of intensification.

5. The phraseme *Από Χ άλλο τίποτα*, typically denoting excess of a certain object or of a certain phenomenon, may indirectly imply lack of other objects or phenomena more desirable to the speaker. The phraseme *Χ να δουν τα μάτια σου* does not have this additional meaning.

4 Concluding remarks

This paper has been concerned with the description of Modern Greek constructional phrasemes – a class of phraseologisms that has until now not received significant attention from scholars concerned with issues in Modern Greek phraseology and Modern Greek linguistics in general. Although such a study is hampered by the absence of corresponding previous research on the part of linguists and the lack of a systematic approach to such units in both Greek lexicography and grammar, it is of considerable theoretical interest, contributes to the further development of the Modern Greek phraseology, and is important for those who learn or teach Modern Greek as a foreign language.

Based on a set of phrasemes collected from dictionaries of Modern Greek and texts of Greek literary works, this study provides a description of the structure and semantics of these phraseologisms.

It has been argued that Modern Greek constructional phrasemes do not form a homogeneous group of linguistic units and can be divided into three structural types, depending on the nature of elements involved in their fixed component: function words and / or pronouns, semantically bleached content words, or full content words. The lexical make-up of the fixed component of constructional phrasemes is directly related to their imagery potential, which is different for each structural type. The differences found between these three types suggest that the class of constructional phrasemes should be positioned on the syntax-lexicon continuum in an extended rather than a local way.

It has been demonstrated that different constructional phrasemes may have a similar typical meaning and can be grouped on the basis of their semantics, though phrasemes belonging to the same semantic group are not necessarily full synonyms and may differ in semantic nuance or frequency of use. Various semantic groups have been discussed, such as evaluation, high degree of quality, negation, etc.

The paper has also provided more specific analyses of two Modern Greek constructional phrasemes denoting multitude. Each analysis involved the description of the fixed component of the construction and the lexemes that fill the open slot, the examination of semantic and part-of-speech constraints im-

posed on fillers of the open slot, the determination of the typical meaning, and additional semantic nuances of the construction and the analysis of its frequency of use.

This study is currently based on a set of fifty Modern Greek constructional phrasemes, but the current set is expected to be expanded in the future.

References

- Apresian, Valentina. 2015. *Ustupitel'nost': mekhanizmy obrazovaniya i vzaimodejstviya slozhnyh znachenij v yazyke* [Concession: the mechanisms of formation and interaction of complex meanings in language]. Moscow: Yazyki slavyanskoj kul'tury.
- Babinotis, Georgios. 2005. *Leksiko tis Neas Ellinikis Glossas* [Dictionary of the Modern Greek language]. Athens: Kentro Lexikologias.
- Baranov, Anatolij & Dmitrij Dobrovol'skij. 2016. *Osnovy frazeologii (kratkie kurs)* [Basics of Phraseology (a concise course)]. Moscow: Flinta.
- Benigni, Valentina, Paola Cotta Ramusino, Fabio Mollica & Elmar Schafroth. 2015. How to Apply CxG to Phraseology: A Multilingual Research Project. *Journal of Social Sciences* 11(3). 275–288.
- Charalambakis, Christoforos (ed.). 2014. *Khristiko leksiko tis Neoellinikis Glossas* [Practical dictionary of the Modern Greek language]. Athens: Ethniko tipografio.
- Dobrovol'skij, Dmitrij. 2011. Phraseologie und Konstruktionsgrammatik. In Alexander Lasch & Alexander Ziem (eds.), *Konstruktionsgrammatik III: Aktuelle Fragen und Lösungsansätze*, 110–130. Tübingen: Stauffenburg Linguistik.
- Dobrovol'skij, Dmitrij. 2016. Grammatika konstruktivnykh frazeologicheskikh konstruktsij i frazeologicheskogo slovarstva [Construction Grammar and Phraseology]. *Voprosy yazykoznaniya* 3. 7–21.
- Hunston, Susan. 2004. Counting the uncountable: problems of identifying evaluation in a text and in a corpus. In Alan Partington, John Morley & Louann Haarman (eds.), *Corpora and Discourse*, 157–188. Bern: Peter Lang.
- Katos, Georgios. 2017. *Leksiko tis laikis kai perithoriakis glossas* [Dictionary of the colloquial and marginal speech]. Thessaloniki: Centre for the Greek Language.
<http://georgakas.lit.auth.gr/dictionaries/index.php/anazitisi/g-katou> (accessed 30 August 2019).
- Leksiko tis Kinis Neoellinikis* [Dictionary of Standard Modern Greek]. 2017. Thessaloniki: Institute for Modern Greek Studies of the Aristotle University of Thessaloniki.
http://www.greek-language.gr/greekLang/modern_greek/tools/lexica/triantafyllides/ (accessed 30 August 2019).
- Mini, Marianna. 2009. *Glossologiki kai psikhologiki meleti ton pagiomenon rimatikon fraseon me stathero ipokimeno tis Neas Ellinikis: morfosintaktiki analisis, simasioleksiki diavathmisi kai epeksergasia apo paidia skholikis ilikias* [A linguistic and psychological study of the fixed verb phrases with a fixed subject in Modern Greek: a morphosyntactic analysis, a lexical-semantic classification and interpretation by school children]. Athens: University of Patras PhD thesis.

- Mollica, Fabio & Elmar Schafroth. 2018. Der Ausdruck der Intensivierung in komparativen Phrasem-Konstruktionen im Deutschen und im Italienischen: eine konstruktionsgrammatische Untersuchung. In Kathrin Steyer (Hrsg.), *Sprachliche Verfestigung. Wortverbindungen, Muster, Phrasem-Konstruktionen*, 103–136. Tübingen: Narr.
- Motsiou, Victoria. 1987. Eleftheria kai desmefsi sti glossa (i periptosi ton idiotismon) [Freedom and restraint in language (the case of idiomatizations)]. *Dodoni* 16. 231–255.
- Setatos, Michalis. 1996–1997. I diavathmisis stin kini neoelliniki [Gradation in Standard Modern Greek]. *Scientific annals of the Faculty of Philosophy of the University of Thessaloniki*. Volume 6 by the Department of Philology. 47–60.
- Shmelev, Dmitrij. 1960. O «svyazannykh» sintaksicheskikh konstruktivnykh v russkom yazyke [On “bound” syntactic constructions in Russian”. *Voprosy yazykoznaviya* 5. 47–60.
- Shvedova, Nataliya. 1958. O nekotorykh tipakh frazeologizirovannykh konstruktivnykh v stroye russkoy razgovornoy rechi [On certain types of phraseologized constructions in spoken Russian discourse]. *Voprosy yazykoznaviya* 2. 93–100.
- Shvedova, Nataliya. 1960. *Ocherki po sintaksisu russkoy razgovornoy rechi* [An essay on the syntax of spoken Russian discourse]. Moscow: Publishing House of the Academy of Sciences of the USSR.
- Shvedova, Nataliya (ed.). 1980. *Russkaya grammatika*. Tom II. *Sintaksis* [Russian Grammar. Volume II. Syntax]. Moscow: Nauka.
- Skartsis, Sokratis. 1997. *Frasis tou laikou logou* [Phrases of the colloquial speech]. Athens: Ellinika grammata.
- Sthali, Katerina. 2011. Idiome und Konstruktionsgrammatik: Im Spannungsfeld zwischen Lexikon und Grammatik. In Alexander Lasch & Alexander Ziem (eds.), *Konstruktionsgrammatik III: Aktuelle Fragen und Lösungsansätze*, 149–163. Tübingen: Stauffenburg Linguistik.
- Stefanowitsch, Anatol & Stephan Th. Gries. 2003. Collocations: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8. 209–243.
- Tresorukova, Irina. 2015. Katataksi ton fraseologismon tis ellinikis glossas: prosengisi sto thema (me vasi ti rosiki theoria tis glossologias) [Classification of the phraseologisms of the Greek language: an approach to the subject (based on the Russian theory of linguistics)]. In Konstantinos Dimadis (ed.), *Proceedings of the 5th European Congress of Modern Greek Studies. Continuities, Discontinuities, Ruptures in the Greek World (1204–2014): Economy, Society, History, Literature*. Vol. 4. Thessaloniki: European Society of Modern Greek Studies.
- Tzartzanos, Achillefs. 2016. *Neoelliniki sintaxis tis Kinis Dimotikis* [Modern Greek Syntax (of Common Demotic)]. Thessaloniki: Kiriakidis Editions.
- Velichko, Alla. 2016. *Predlozheniya frazeologizirovannoy struktury v russkom yazyke: strukturno-semanticheskoye i funktsionalno-kommunikativnoye issledovaniye* [Sentences of phraseologized structure in Russian language: a structural semantic and functional-communicative study]. Moscow: MAK Press.

Valentina Piunno

Coordinated constructional intensifiers: patterns, function and productivity

Abstract: This investigation proposes a corpus-based description of a particular type of Italian phraseological construction, having a coordinated syntactic structure and playing the role of an intensifier. Through the analysis of data extracted from a corpus of Italian, the contribution aims at i) identifying and analysing different types of coordinated constructional intensifiers, ii) showing a formal representation of their syntactic-semantic schemas, iii) classifying them into different types on the basis of their functional, syntactic and semantic properties. The paper is structured as follows: the first section contains a brief description of the main types of intensifying strategies identified in the literature, deserving a special attention to Italian lexical strategies. The second section is devoted to the description of the general theoretical framework of this contribution, and to the analysis and definition of constructional intensifiers. The third section briefly describes the methods of data extraction and analysis. The fourth section is devoted to the analysis of Italian coordinated intensifying constructions, which are classified into completely and partially filled units, according to a set of parameters (i.e. lexical variation, syntactic cohesion, idiomatism and lexicalisation). Then, the set of coordinated constructional intensifiers is evaluated against the presence of an abstract scheme, semantic and syntactic analysability, semantic predictability, productivity, and schematicity.

Keywords: coordinated constructional intensifiers, partially and completely filled constructions, pattern schematicity, semantic predictability, degrees of lexicalisation.

1 Intensification strategies

Intensification strategies are linguistic resources that languages use to superficially convey the intensifying value. An intensifier is a “device that scales a quality, whether up or down or somewhere between the two” (Bolinger 1972: 17). Intensifying tools can appear under several forms, varying on the basis of

Valentina Piunno, Tuscia University, valentina.piunno@unitus.it

<https://doi.org/10.1515/9783110520569-006>

language features and of the speaker's choice. As Labov (1984: 48) suggests, the set of markers of intensity is not closed: “[i]ntensity is signaled by a large and miscellaneous class of devices, ranging from the most peripheral of prosodic variations to the most central categories of the grammar [...]: verbal and non-verbal gestures; expressive phonology, including sudden changes in length, pitch, duration, and vowel quality; repetition; and the use of quantifiers”. Among the most commonly used linguistic strategies, it is possible to highlight i) syntactic intensification (e.g. dislocations, topicalisations, etc.), ii) phonological intensification (e.g. intonation), iii) morphological intensification (e.g. prefixes, superlative suffixes, partial reduplication) and iv) lexical intensification (e.g. full reduplication). Thus, intensifiers can vary in nature, depending on the intensifying strategy admitted (and preferred) in a language: “il existe une foule de lexies pouvant exprimer l’Intensification, mais le choix de la lexie appropriée dépend de la lexie dont le locuteur veut intensifier le sens” (Mel’čuk and Polguère 2007: 20). Among others, the class of *lexical intensifiers* represents a common and productive strategy of intensification across different languages,¹ and for this reason it will be the object of study of this work.

Lexical intensifiers are linguistic units able to convey an intensifying meaning to the lexemes with which they occur. This set of intensifiers is mainly composed of adjectives (1) and adverbial (2) items (in italics, in the examples) (Bolinger 1972).²

(1) *big*_{INTENS} problem, *deep*_{INTENS} interest, *true*_{INTENS} friend, *total*_{INTENS} failure

(2) *deeply*_{INTENS} grateful, *extremely*_{INTENS} rare, *perfectly*_{INTENS} awful, *too*_{INTENS} right³

In such cases, adverbial and adjectival items usually express a quality (i.e. *big*, *deep*, *true*) or a quantity (i.e. *too*, *all*, *totally*). In the case of *quality intensifiers*, the intensification is based on a relation of similarity involving semantic prototypicality (Rapatel 2015) and categorisation of the intensified item by prototypes (Anscombe and Tamba 2013): *a true friend* is ‘the best representative of the

¹ In some cases it is even one of the most employed. For example, according to Bolinger (1972) English intensifiers are mainly lexical in nature.

² As Bolinger points out, “[m]anifestations of degree and intensity are commonly associated with adjectives and adverbs, not so commonly with nouns and verbs” (1972: 15).

³ In particular, as far as adverbial intensifiers are concerned, Bolinger notes that “investigation will probably reveal that virtually any adverb modifying an adjective tends to have or to develop an intensifying meaning” (1972: 23).

friend category' in the terms of Rosch (1978).⁴ *Quantity intensifiers* are able to make the intensified item gradual and to (metaphorically) refer to the highest degree of a certain quantity (e.g. *totally serious*). Furthermore, intensification is a functional-semantic category (of amplification or attenuation) and encompasses a number of different scales (Van Os 1989).⁵ According to Bolinger (1972: 17), it is possible to “distinguish intensifiers according to the region of the scale that they occupy”: e.g. *boosters* (in the upper section), *compromisers* (in the middle), *diminishers* (the lower section), *minimizers* (the “lower end of the scale”) (Bolinger 1972: 17).⁶

As far as the distributional properties are concerned, lexical intensifiers can co-occur with several kinds of linguistic items, such as nouns (i.e. *deep*_{intensifier} *interest*_{intensified}), adjectives (e.g. *dead*_{intensifier} *tired*_{intensified}), adverbs (e.g. *truly*_{intensifier} *never*_{intensified}), verbs (e.g. *to like*_{intensified} *a lot*_{intensifier}), pronouns (e.g. *absolutely*_{intensifier} *nothing*_{intensified}), but also word combinations and predicative items:

- word combinations (e.g. *terrible*_{intensifier} *Monday morning feeling*_{intensified})
- a predication (Labov 1984) (e.g. *I am so*_{intensifier} *tired of this!*)
- a question (i.e. *What the hell*_{intensifier} *is going on here?*)
- a negation (i.e. *I cannot see a bloody thing*_{intensifier})

With respect to lexical selection, it is important to note that intensifiers tend to single out gradable items: i.e. it. *una bella*_{INTENS} *nevicata* (‘a quantitatively important snowfall’, *lit.* a beautiful snowfall). In some cases, it is nevertheless possible to apply intensification to non-gradable lexical elements, coercing the semantics of the intensified item: i.e. it. *bel*_{INTENS} *guaio* (‘an important trouble’, *lit.* a beautiful trouble).

It is worth noting here that such co-occurrence properties can often be associated with some more general and abstract combinatorial schemas: in fact, the group of lexical intensifiers includes a set of co-occurrence patterns which are characterised by common syntactic and semantic features. Such abstract configurations may be described as *constructions*, in the terms of models of Construction Grammar (Fillmore, Kay, and O’Connor 1988; Goldberg 1995,

⁴ As far as the syntactic position of the intensifier is concerned, it is often subject to restrictions. In Italian, for example, the adjectival intensifying item usually falls to the left of the adjectival or nominal element (i.e. *un perfetto*_{INTENS} *idiota* vs. *un idiota perfetto*_{INTENS}).

⁵ In particular, Van Os (1989: 118) highlights a set of different “degrees of intensification”: *absolut* ‘absolute’, *approximativ* ‘approximate’, *extrem hoch* ‘extremely high’, *hoch* ‘high’, *gemäßigt* ‘moderate’, *schwach* ‘attenuating’, *minimal* ‘minimal’.

⁶ For the purpose of this study, only the first group will be considered.

2006) and of Categories and Constructions Grammar (Simone 2007). The starting assumption of the current study is that some classes of lexical intensifiers derive from the transition between two different levels (Simone 2007): i) the purely linguistic level, and ii) the cognitive-conceptual (pre-linguistic) level. The former is the level of individual linguistic sequences (i.e. *constructs*), which can be associated with linguistic templates representing peculiar semantic values (i.e. *constructions*) at the cognitive conceptual level; the latter is able to structure and organise the semantic information associated with lexemes (Simone 2007). Specific syntactic-semantic patterns can therefore identify the original templates of different but conceptually similar constructions. In particular, this contribution aims at examining *constructional intensification* as a set of lexically-filled and semi-filled patterns characterised by the presence of almost two lexical slots (i.e. the slot of the *intensifier*_A and the one of the *intensified*_B: e.g. *deep*_A *interest*_B), and which have an intensifying function.

2 Constructional intensifiers

Constructions have been the object of many recent studies and are generally defined as “relatively stable combinations of words that tend to co-occur in discourse and are treated [...] as units on some level of representation” (Simone 2007: 210). Constructions are characterised by a ready-made syntactic pattern which, through an increasing frequency of use (Bybee 2010) and a process of routinisation (Detges and Waltereit 2002: 181), becomes fixed and specialises in conveying specific a meaning – which is thus “constructional”, in the sense that it is specifically associated with the construction and “fades away outside of it” (Simone 2007: 215).

For the purpose of this analysis, *constructional intensifiers* (hereinafter CIs) can be defined as stable combinations of lexical elements and/or part of speech tags which convey the specific semantic value of intensification, and show variable degrees of productivity (Bybee 2010: 94) in terms of syntactic patterns. Different types of CIs and constructional intensifying strategies have been identified in the literature. Among the most popular we can mention reduplication, polar extreme enforcement phrases and coordinated items:⁷

⁷ It is worth noting that this graph only includes a part of possible constructional phenomena used as intensifiers in Italian as well as in other languages. A number of other interesting intensifying phenomena could be mentioned, such as *comparative phraseological constructions*

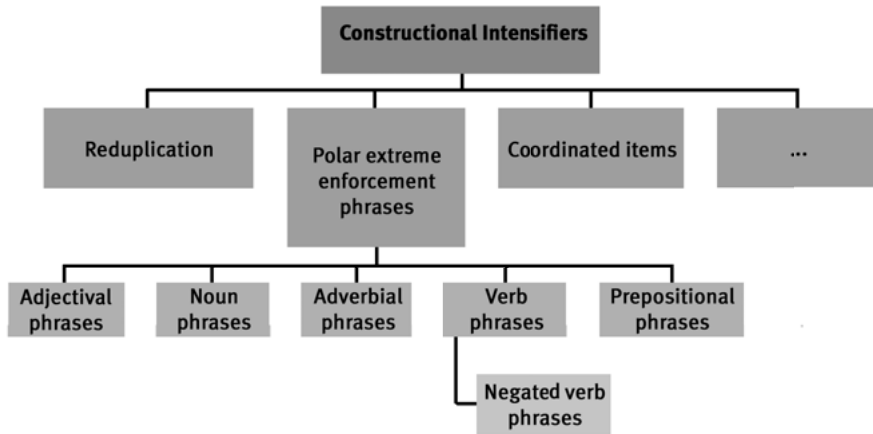


Fig. 1: Constructional intensifying formats

Reduplication may be understood as one of the most productive CI formats of the world languages (Marantz 1982),⁸ and a number of different reduplication formats may exist (Moravcsik 1978).⁹ Reduplication may be distinguished on the basis of i) the repeated element (e.g. *partial*¹⁰ and *total reduplication*), ii) the lexicalisation degree of the sequence (e.g. it may be a phenomenon of *discourse*, or a lexical unit definitely lexicalised and ascribable to the system), iii) the morpho-syntactic structure (namely, the parts of speech involved in the sequence, e.g. Adjective₁ + Adjective₁ (e.g. *same same*). It is worth noting that, on the basis of the language and on the type of repetition, reduplication may bear different functions: the intensifying meaning is the most frequent, but other values are

(Mollica and Schafroth 2018) and *light nouns* (Simone and Masini 2014), among others. However, they will not be considered for the purpose of this study.

8 As Rubino (2005) points out, 85% of the languages included in the WALS employ reduplication formats as a mechanism of intensification.

9 As far as Italian reduplication is concerned, cf. among others, De Santis (2011), as well as Grandi (2017) for some more general considerations (and references therein).

10 We highlight here the distinction between partial and total reduplication, even though partial reduplication cannot be included in the CI strategies. As a matter of fact, partial reduplication is generally considered as a morphological phenomenon, as it typically involves morphemic units.

also possible.¹¹ What is interesting about intensifying reduplication is that the intensifying value is not conveyed by the lexical elements involved, but by the syntactic format of the combination (that is, the construction itself).

The second type of construction is represented by the so-called *polar extreme enforcement* phrases (Lehmann 2005) and is used with the same function as a superlative: “[g]iven a predicate that may be true of its argument to different degrees, there may be a default value of that predicate for that class of arguments, and there may be particular individuals that the predicate is true of to a higher degree or even to the highest conceivable degree” (Lehmann 2005). The construction may contain sequences of two or more elements. Among the most studied Italian intensifying phrases it is possible to mention the following:¹²

Tab. 1: Some examples of Italian intensifying phrases¹³

Phrase	Pattern	Italian examples
Adjectival phrase	Adj + Adj _{INTENS} ¹⁴	innamorato <i>pazzo</i> ‘madly in love’ (<i>lit.</i> in love mad)
	Adv _{INTENS} + Adj	<i>terribilmente</i> dispiaciuto ‘awfully sorry’
	Adj _{INTENS} + NOUN	<i>completo</i> fallimento ‘dead failure’ (<i>lit.</i> complete failure)
Noun phrase	NOUN + Noun _{INTENS}	freddo <i>cane</i> ‘freezing cold’ (<i>lit.</i> cold dog)
	NOUN + Adj _{INTENS <relational>}	freddo <i>glaciale</i> ‘freezing cold’ (<i>lit.</i> cold glacial)
	la/il grande _{INTENS} + Noun ¹⁵	<i>la grande</i> bellezza ‘the great beauty’ (<i>lit.</i> the big beauty)
	Adj + Past Participle _{INTENS}	un attore <i>nato</i> ‘a very good actor’ (<i>lit.</i> an actor born)
	un/una + Noun _{INTENS} + di + Noun	<i>un tesoro di</i> ragazza ‘a very nice girl’ (<i>lit.</i> a treasure of girl)

11 For example, reduplication can mark the future tense (Stolz et al. 2011), imperfective aspect (Hardy and Montler 1988), and plurality (Dixon 2004). For a list of different functions connected with reduplicative constructions, cf. Inkelas and Downing (2015a, 2015b).

12 This list only includes a small set of the most analysed intensifying constructions, and it is not fully representative of the whole group of intensifying configurations of Italian.

13 Abbreviations: Adj: adjective; Det: determiner; Neg: negation; Prep: preposition.

14 Cf. Berlanda (2013) for an in-depth analysis of this structure.

15 Cf. Grandi (2017: 57).

Phrase	Pattern	Italian examples
Prepositional phrase	Prep + (Adj) + Noun/Adj INTENS ¹⁶	<i>da matti</i> 'crazy' (<i>lit.</i> of crazy), <i>a tutta birra</i> 'full speed' (<i>lit.</i> at all beer)
	Prep + (Adj) + Verb INTENS ¹⁷	<i>da morire</i> 'very much' (<i>lit.</i> to die), <i>a tutto andare</i> 'full speed' (<i>lit.</i> at all going)
	di un/una + Adj + Adj INTENS	<i>di un rosso acceso</i> 'bright red' (<i>lit.</i> of a red bright)
	di un/una + Noun + Adj INTENS	<i>di una pazienza angelica</i> 'very patient' (<i>lit.</i> of a patience angelic)
	di + Noun + di + Noun ¹⁸ del tutto INTENS + Adj ¹⁹	<i>di giorno in giorno</i> 'day by day' (<i>lit.</i> of day of day) <i>del tutto ridicolo</i> 'completely ridiculous' (<i>lit.</i> of the all ridiculous)
Verb phrase	Neg + Verb + Det + Noun INTENS ²⁰	<i>non capire un cavolo</i> 'to not understand a thing' (<i>lit.</i> to not understand a cabbage)
	Verb + di INTENS + Noun	<i>riempire di critiche</i> 'ply with critiques' (<i>lit.</i> to fill with critiques)

Coordinated intensifiers are *binomial* constructions i) composed generally of two lexical items (usually belonging to the same class) which are coordinated by a conjunction, and ii) playing the role of intensifiers.

(3) *bello e buono*
'downright' (*lit.* beautiful and good)

The pairs of coordinated units are irreversible from the syntactic point of view (cf. Malkiel 1959), and are characterised by specific semantic relations (§ 4.1). As a matter of fact, a number of studies have devoted a great attention on the analysis of the so-called *binomial* sequences, in many different languages.²¹ Never-

¹⁶ Cf. Benigni (2017) and Piunno (2018a) for the analysis of this particular type of intensifier.

¹⁷ This pattern also includes the intensifying structure [*a tutto* + Verb_{infinitive}]_{INTENS}, which has been extensively analysed by López Meirama (2020) for Spanish.

¹⁸ Cf. Piunno (2018a) for Italian. This peculiar construction has also been analysed for Spanish in terms of Construction Grammar by López Meirama and Mellado Blanco (2018). This configuration semantically recalls the intensifying configuration [Noun + *su* + Noun], which has been extensively analysed in Schafroth (2020).

¹⁹ This construction has been analysed for Italian by Cimaglia (2011).

²⁰ Cf. Piunno (2018b) for Italian. An in-depth analysis of negated constructions in the framework of Construction Grammar is also found in Mellado Blanco (2020b).

²¹ Among others, cf. in particular the pioneering study of Malkiel (1959), the analysis of Gustafsson (1975), and very recent work by Kopaczyk and Sauer (2017), containing an in-depth

theless, the set of coordinated CI is still neglected in the literature. For this reason they have been selected as the object of this analysis. In particular, the study aims at identifying their peculiar semantic, syntactic and functional properties, from the general perspective of Construction Grammar.

3 Methodology and data extraction

This contribution takes a usage-based approach based on the analysis of data extracted from a corpus of contemporary Italian. In particular, data have been gathered by means of *Sketch Engine* (Kilgarriff et al. 2004, 2014), a software which is able to easily extract syntagmatic word clusterings from corpora. In particular, *Sketch Engine* is an online text analysis tool, containing a number of freely available ready-made corpora. The corpus that has been selected for the study is the *Italian Web corpus itTenTen16*²², a collection of texts extracted from the web, containing 4,9 billion words. Data have been extracted thanks to the frequency lists obtained through the CQL advanced query, which allows the user to combine different morpho-syntactic constraints, in order to obtain specific PoS-grams patterns. In particular, the CQL query is able to select sequences of part-of-speech tags, by specifying the position, the lexical category, and the specific lemma that the extracted sequences should include. For the purposes of this analysis, coordinated CIs have been detected through a PoS-based query which included the PoS-grams patterns and the coordinating conjunction *e* ‘and’: e.g. [Adjective + *e* + Adjective], [Past participle + *e* + Adjective/Past participle]. For reasons of space, this analysis does not provide a complete representation of all the possible alternative configurations involving the coordinated pattern. In particular, it excludes less frequent syntactic patterns employing nouns and adverbs as their lexical constituents, as well as more complex sequences involving modifiers. In fact, as far as the latter structure is concerned, adverbial modifiers can occasionally appear in the preadjectival position: e.g. [Adjective + *e* + **Adverb** + Adjective]. According to corpus data, such configurations mostly employ the adverb *più* ‘more’ (e.g. *nuovo e più moderno* ‘new and

analysis of the literature devoted to the general topic of binomial sequences. For Italian, cf. in particular Masini (2006) and references therein.

²² The *itTenTen16* corpus belongs to the *TenTen* family corpus, a set of comparable corpora freely available in *Sketch Engine* (Jakubíček et al. 2013). It is a large web-text corpus representing a balanced and representative sample of non-specialised varieties of language.

more modern')²³ and a number of other quantity and quality adverbs (e.g. *molto* 'much', *ben* 'well'), and generally give rise to less stable and rather unfrequent configurations, which can be used to further intensify the semantic value of the construction.²⁴

It is worth noting that a part of the qualitative analysis has been driven taking into account two lexicographic works: i) *CombiNet*, an Italian combinatory dictionary which has been recently developed²⁵, and ii) *GRADIT*²⁶, an Italian dictionary of language usage containing a large quantity of words and a fine-grained description of fixed combinatorial sequences.

The most frequently occurring 600 units of each pattern have been collected from the corpus. Extracted data have been qualitatively evaluated against the following parameters:

- the semantic class of the lexemes included in the construction and the overall constructional meaning (§ 4.1)
- the degree of fixedness of the sequence and cohesion of its single components (§ 4.2)
- the presence of a multitude of exemplars having similar patterns, that could be attributed to a single abstract constructional scheme (§ 4.2)
- the functional properties of the sequences (§ 4.3)

4 Coordinated constructional intensifiers

Coordinated CIs are characterised by a specific morpho-syntactic pattern composed of (almost) three syntactic slots:

23 The data extracted from the corpus show that *più* 'more' is by far the most frequently employed adverb in the preadjectival position (in a set of 2,000 "modified" constructions it occurs in 50% of cases), followed by other types of intensifiers (or minimisers) referring to quantity or quality. The examples show the most frequently occurring ones:

- (1) *semplice e **facilmente** comprensibile* 'easy and easily understandable'
- (2) *chiaro e **ben** visibile* 'clear and clearly visible' (*lit.* clear and well visible)
- (3) *sano e **perfettamente** funzionante* 'safe and perfectly working'
- (4) *fondamentale e **assolutamente** necessario* 'fundamental and absolutely necessary'

24 More corpus-based research into the different features of such configurations is needed to verify whether the lesser internal cohesion between constituents can be correlated to poor levels of lexicalisation (and constructionalisation) of the syntactic sequence.

25 Cf. Simone and Piuino (2017), Lenci et al. (2017), Piuino (2016).

26 *Grande Dizionario Italiano dell'Uso* (De Mauro 1999).

(4) a. EN: *sick and tired*
 A B C

b. IT: *vivo e vegeto*
 A B C
 ‘*alive and kicking*’

- i. Slot A: the first position is filled with a lexical item (generally an adjective, a past participle and an adverb; only rarely a noun);
- ii. Slot B: the second slot contains the coordinating conjunction *e* ‘and’;
- iii. Slot C: the third slot employs another lexical item, belonging to the same morpho-syntactic category as A, and sharing some of its semantic peculiarities, but differing in terms of lexical unit.

The sequence may be composed of two coordinated adjectives, past participles or adverbs, as the examples below show:

(5) [Adjective/Past Participle + *and* + Adjective/Past Participle]²⁷

IT: *vero e proprio, sano e salvo, fatto e finito*

‘out-and-out’ (*lit.* true and proper), ‘safe and sound’ (*lit.* healthy and safe), ‘done and dusted’ (*lit.* done and ended)

EN: *armed and ready, sick and tired, open and shut*

(6) [Adverb + *and* + Adverb]

IT: *solo e soltanto, forte e chiaro*

‘exclusively’ (*lit.* only and just), ‘loud and clear’

EN: *far and wide, above and beyond*

The constructional pattern may sometimes include nominal elements (7), or may be more complex in syntactic terms (8):

(7) [Noun + *and* + Noun]

IT: *anima e corpo*

‘body and soul’ (*lit.* soul and body)

EN: *body and soul, hammer and tongs*

(8) [Preposition + Noun + *and* + (Preposition) + Noun]

²⁷ For the purpose of this analysis only this pattern has been taken into account.

IT: *in lungo e in largo, in fretta e furia, d'amore e d'accordo*
 'far and wide' (*lit.* in long and in large), 'all of a sudden' (*lit.* in hurry and haste'), 'in peace and harmony' (*lit.* of love and of agreement)

It is worth noting that coordinated CIs never consist of the mere reduplicating of lexical elements, as happens for many intensifying constructions involving repetition. The two lexical elements employed often belong to the same semantic sphere, and may be in a particular paradigmatic relation between each other. Both the lexical elements involved and the coordinated syntactic scheme play a pivotal role in the attribution of the intensifying value to the whole construction. The entire sequence has an intensified meaning in the sense that the meaning of the entire construction is intensified with respect to the meaning of the single lexical items it includes. The peculiarity of the construction is represented by i) the semantic relation linking lexemes A and C, which can differ on the basis of single constructs, ii) its schematicity and productivity in terms of new forms, and iii) its functional value.

4.1 Semantic relations between the coordinated lexical items

Coordinated CIs generally include two lexical elements, often sharing some specific semantic values. This section is devoted to the analysis of the specific semantic features of such constructions, with particular reference to those which are fully lexicalised and fixed (thus, also emerging as dictionary entries). This choice aims at demonstrating that a sort of regularity characterises some sequences: i) at times, even the most cohesive and entrenched phraseological units may be attributed to or recall a sort of abstract scheme, and ii) sequences differing from each other in terms of lexical items and meaning may sometimes share a common semantic constructional scheme.

In particular, it is possible to highlight four groups of coordinated CIs, on the basis of the degree of semantic overlap between the lexical elements (A and C) involved in the construction:

- i. $A = C \rightarrow$ Lexemes A and C are *absolute synonyms* (Lyons 1981);
- ii. $A \simeq C \rightarrow$ Lexemes A and C are *completely synonyms* (Lyons 1981);
- iii. $A \approx C \rightarrow$ Lexemes A and C are *near (or quasi-) synonyms*;
- iv. $A \neq C \rightarrow$ Lexemes A and C belong to the same morpho-syntactic category, they share some formal features, but their meaning is semantically unrelated.

It is worth noting that the boundaries between the different groups are not strictly defined, as it is possible to identify sequences simultaneously representing more than one subset. This is particularly true for iii), which encompasses different kinds of phenomena.

In the first case, the two terms “have the same distribution and are completely synonymous in all their meanings and in all their contexts of occurrence” (Lyons 1981: 148); they cannot be distinguished either denotatively or connotatively, as they are completely interchangeable, as in the following example:

- (9) a. *solo e soltanto*
 ‘exclusively’ (*lit.* only and just)
 b. *felice e contento*
 ‘extremely happy’ (*lit.* happy and content)

In such cases the lexemes involved in the coordinated structure are absolutely interchangeable. Thus, the construction could be defined as *semantic reduplication*, since it involves different – but completely overlapping – lexemes. For this reason some scholars define constructions of this kind as *tautological* (Kopaczyk and Sauer 2017): in fact, they are able to convey exactly the same meaning by means of two different lexemes.

The second type describes a relation of complete synonymy between A and C; in particular, the two terms share “a certain range of contexts” (Lyons 1981: 148) and are completely intersubstitutable exclusively in that range of contexts, as in the following examples:

- (10) a. *grande e grosso*
 ‘big and strong’ (*lit.* great and big/robust)
 b. *in fretta e furia*
 ‘all of a sudden’ (*lit.* in hurry and haste)

Thus, even though *grande* ‘great’ and *grosso* ‘big/robust’ share many different contexts, they cannot substitute each other in the totality of their uses.²⁸

In the third case, the coordinated structure employs two lexical units which have a similar meaning – cognitive and denotational meaning in terms of Cruse (1986) and Halliday (1976). Nevertheless, the two lexemes are not collocational-

²⁸ This is the case, for example, of *mare grosso* ‘rough sea’ (*lit.* sea big/robust), or *sale grosso* ‘course salt’ (*lit.* salt big/robust), where *grosso* cannot be substituted by *grande*.

ly interchangeable,²⁹ because their meaning is not strictly symmetrical (Carter 1987) and they are only approximately equivalent (Landheer 1989). Furthermore, in the great majority of coordinated constructions including near synonyms, C is semantically surplus with respect to A (e.g. it may denote a qualifying value, or an aspectual information), but it can also represent a partial semantic reduplication of its meaning. In some cases, C may add a semantic trait or semantic information which is not included in A, thus intensifying its value by semantic addition, as in the following examples:

- (11) a. *sano e salvo*
 ‘safe and sound’ (*lit.* healthy and safe)
sano [+ healthy], *salvo* [+ safe]
 b. *unico e irripetibile*
 ‘unique and unrepeatable’
unico [+ uniqueness, + singular], *irripetibile* [- repetition]

In (11a) *salvo* ‘safe’ adds a specific trait of ‘free from hurt or damage’, which is not properly included in *sano* ‘healthy’. The semantic role of the C unit is particularly evident in sequences like (11b), where the lexeme (*irripetibile* ‘unrepeatable’) is able to mark the highest degree of a quality (the *uniqueness*). The lexeme C may also denote a partial reduplication of the semantics expressed by A: this means that C adds semantic information which is generally already included in the meaning of A, as in the following examples:

- (12) a. *vivo e vegeto*
 ‘alive and kicking’ (*lit.* alive and thriving)
vivo: [+ animate], *vegeto*: [+ animate] [+ thriving]
 b. *primo e unico*
 ‘first and unique’
primo: [+ preceding, + singular], *unico*: [+ uniqueness, + singular]
 c. *fatto e finito*
 ‘done and dusted’ (*lit.* done and ended)
fatto: [+ completed], *finito* [+ concluded]

For example, in the sequence *vivo e vegeto*, *vegeto* denotes a thriving entity, which not only is ‘alive’ (i.e. *vivo*), but also ‘healthy’. In (12b) something that is *primo e unico* ‘first and unique’, is not only the single exemplar preceding all

²⁹ Cf. Conzett (1997), Partington (1998), Tognini Bonelli (2001), and Xiao and McEnery (2006).

others, but also the single (and unique) one of its type. Thus, in (12c), if something is *ended* ('finito') or concluded, then it presupposes that it has been *done* ('fatto') or completed. This construction has another important peculiarity: C may add specific aspectual information, which typically denotes a point of completion. In such cases the lexeme in the C slot is expressed as the past participle and expresses the potential end-point of an event or apex of a status, as happens in the following examples:

- (13) a. *morto e sepolto*
 'dead and buried'
 b. *cornuto e mazziato*
 'to add insult to injury' (*lit.* cuckolded and beaten)
 c. *calzato e vestito*
 'completely' (*lit.* shod and dressed)

Thus, if something is *dead and buried* ('*morto e sepolto*'), it is completely finished, so that it cannot happen again. The intensifying meaning is attributed by C, which denotes the ending point of the process; it is worth noting that, even though the sequence may acquire a metaphorical meaning, the syntactic order of the lexemes reflects the exact temporal succession of events. As will be shown in the following sections, the aspectual feature often affects the whole construction, which tends to lose its primary semantic properties, thus acquiring the telic meaning of completion. Thus, the construction itself contains end-point oriented aspectual information.³⁰

As far as the fourth group is concerned, lexemes A and C may be semantically unrelated, even though they share the part-of-speech categorisation:

- (14) a. *bello e buono*
 'downright' (*lit.* beautiful and good)
 b. *sempre e solo*
 'only ever' (*lit.* always and only)
 c. *nudo e crudo*
 'true' (*lit.* nude and raw)

In most cases such constructions are characterised by syntactic fixedness and high degrees of lexicalisation. The lexical items involved do not share any pecu-

30 As far as the relationship between intensification and aspectual information (e.g. duration) is concerned, cf. also the examples analysed by López Meirama (2020) for Spanish.

liar semantic features, even though they may show sound similarities: there is nothing in common between *nudo* ‘nude’ and *crudo* ‘raw’, but they are partially alike in phonological terms.

Finally, it is important to bear in mind that the different sets of intensifiers show a varying distribution in terms of number of exemplars: not surprisingly, while groups i), ii) and iv) are generally less frequently attested, the clusters represented in iii) tend to be the most frequently associated with coordinated CIs.

4.2 Schematism and productivity

As far as the degrees of lexicalisation and fixedness are concerned, through the data-extraction process two types of CIs have emerged:³¹

- i. *completely filled intensifying constructions*, which are fully lexically specified (they are invariable in terms of lexical features), semantically unpredictable (they may have metaphorical or idiomatic meanings), and morpho-syntactically fixed,
- ii. *partially filled intensifying constructions*, which are characterised by a lower degree of lexical specification (thus, they are flexible to variation), and a higher degree of semantic and lexical predictability.

On the one hand, the subset of *completely filled intensifying constructions* typically includes fully lexicalised phraseological constructions, which not only are syntactically invariable – as they are characterised by a cohesive force (Simone 2007) and do not allow any syntactic variation) –, but also semantically fixed – in the sense that they have acquired a specific and invariable semantic value, which does not generally correspond to the primary meaning of its components. Thus, they may be idiomatic, as they “may specify a semantics (and/or pragmatics) that is distinct from what might be calculated from the associated semantics” of its components (Fillmore, Kay and O’Connor 1988: 501). As words,

31 For the distinction between lexically specified and partially lexically filled constructions, cf. among others, Fillmore, Kay and O’Connor (1988), Goldberg (2006), Michaelis (2019). Partially lexically specified constructions represent a dominant topic in the field of constructionist phraseology, where they are generally defined as ‘phraseological constructions’ (*Phrasem-Konstruktionen*, cf. Dobrovolskij 2011, Steyer 2018; *construcciones fraseológicas*, cf. Mellado Blanco 2020a, Mellado Blanco 2020b) or ‘multiword patterns’ (*Wortverbindungsmuster*, cf. Steyer 2013, 2018). As far as Italian is concerned, cf. Piuanno (2018a, 2020), as well as Mollica and Schafroth (2018) in the Italian-German contrastive approach.

completely filled constructions are a stable part of the lexicon and, as such, “they have to be learned separately as individual whole facts” (Fillmore, Kay and O’Connor 1988: 504). It is not uncommon to find them as dictionary entries (cf. *CombiNet*) or as examples of multiword units (cf. *GRADIT*). In the following, some examples of the completely filled intensifying constructions which have been collected through the corpus are represented:

- (15) a. *bello e buono*
 ‘downright’ (*lit.* beautiful and good)
 b. *adulto e vaccinato*
 ‘grown up’ (*lit.* adult and vaccinated)
 c. *calzato e vestito*
 ‘completely’ (*lit.* shod and dressed)
 d. *vero e proprio*
 ‘out-and-out’ (*lit.* true and proper)
 e. *nudo e crudo*
 ‘true’ (*lit.* nude and raw)
 f. *puro e duro*
 ‘out-and-out’ (*lit.* pure and hard)
 g. *grande e grosso*
 ‘big and strong’ (*lit.* great and big/robust)
 h. *sano e salvo*
 ‘safe and sound’ (*lit.* healthy and safe)

The group of completely filled intensifying constructions is unproductive in terms of creation of new forms and *extensibility* (Barðdal 2008), since they do not originate new intensifiers through their syntactic-semantic schema. This means that, though they are characterised by the same morphosyntactic pattern, they tend not to share any lexical or semantic features with other exemplars of the same kind. Their meaning can be idiosyncratic, and they are often opaque and unanalyzable: thus, it is not always possible to recognise and to treat distinctly its single components (Traugott and Trousdale 2013). For example, the phraseological construction *bello e buono* ‘downright’ (*lit.* beautiful and good) is not analyzable in semantic terms, since it has acquired a functional value (the one of intensification) and its component parts have lost their original meaning, as the example below shows:

- | | | | | | |
|-----------------|-------------|----------------|--------------|----------|--------------|
| (16) <i>Non</i> | <i>dire</i> | <i>falsità</i> | <i>belle</i> | <i>e</i> | <i>buone</i> |
| Not | say | lies | beautiful | and | good |

‘Don’t say outright lies’
(itTenTen16, 50490970)

On the other hand, the group of *partially filled intensifying constructions* is represented by partially schematic word combinations characterised by a flexible syntactic pattern and showing variable degrees of productivity (Bybee 2010: 94). As *constructions* (Fillmore, Kay and O’Connor 1988; Goldberg 1995, 2006) they represent a form-meaning pairing, able to convey specific semantic values. Since they are only *partially* filled in lexical terms, such constructions contain a fixed position and variable ones (Bybee 2010):

(17) [*fixed slot* + VARIABLE SLOT_{SEMANTIC/MORPHO-SYNTACTIC RESTRICTION}]

The fixed slot is totally invariable, and the variable position can be filled with various types of lexical units, often responding to specific semantic and/or morpho-syntactic restrictions. It is worth noting that, in the specific case of CIs, the fixed slot is generally filled with words having a polysemous meaning, or at least being characterised by ‘heterosemy’ (cf. Roberts 2010: 46), in that they are able to vary their semantics on the basis of the context of use (e.g. *puro* ‘pure’, *semplice* ‘simple’, *grande* ‘big’, *chiaro* ‘clear’, *forte* ‘strong’). If involved in intensifying constructions of this kind, such lexemes tend to lose their primary concrete meaning, and are generally used in figurative or metaphorical contexts.

It is extremely important to note that partially filled units do not produce permanent effects on the lexicon, since they do not create additional units of the language as a system. The constructs originating from abstract constructions are typically unstable in terms of fixedness and can be intended as discourse-based³² phenomena (cf. Simone 2007). Nevertheless, new instances of partially filled constructions can always represent good candidates for inclusion in the system: frequency of use (Bybee 2010) and routinisation (Detges and Waltereit 2002) can in fact improve the probability of lexicalisation occurring.

In the following example, the Italian CI [*puro e X*] (*lit.* pure and) represents the fixed slot. The variable slot can be filled with adjectival lexemes having a

³² This notion recalls the distinction between *discourse* and *system*, which goes back to the tradition of studies on enunciation (Benveniste 1967; Berrendonner 2002; Blanche-Benveniste 2003) and those of Givón (1979) and Hopper and Thompson (1984). The distinction is based on the belief that language resources are not all equally accessible to the speaker, and can be in fact distributed over two different levels: the *system* resources and those of *discourse*. Among others, cf. Simone (2007) on this topic.

qualifying meaning (e.g. *santo* ‘saint’, *incontaminato* ‘uncontaminated’, *sincero* ‘sincere’, *cristallino* ‘crystal clear’), as the examples below show. The construction acquires the peculiar meaning of “totally/extremely X_{ADJECTIVE}” in combination with this class of adjectives:

(18) [*puro e* + X_{ADJECTIVE<qualifying>}] = ‘totally/extremely X_{ADJECTIVE}’

Tab. 2: Partially filled constructions employing the adjective *puro* in the first slot

Partially filled unit	Absolute frequency	Normalised frequency
1. <i>puro e semplice</i> ‘pure and simple’	11,109	1.89
2. <i>puro e santo</i> ‘pure and saint’	383	0.07
3. <i>puro e naturale</i> ‘pure and natural’	372	0.06
4. <i>puro e incontaminato</i> ‘pure and uncontaminated’	360	0.06
5. <i>puro e sincero</i> ‘pure and sincere’	343	0.06
6. <i>puro e sano</i> ‘pure and healthy’	292	0.05
7. <i>puro e cristallino</i> ‘pure and crystal clear’	275	0.05
8. <i>puro e perfetto</i> ‘pure and perfect’	224	0.04
9. <i>puro e vero</i> ‘pure and true’	222	0.04
10. <i>puro e limpido</i> ‘pure and limpid’	221	0.04
11. <i>puro e fresco</i> ‘pure and fresh’	199	0.03
12. <i>puro e genuino</i> ‘pure and genuine’	186	0.03
13. <i>puro e libero</i> ‘pure and free’	176	0.03
14. <i>puro e innocente</i> ‘pure and innocent’	161	0.03
15. <i>puro e trasparente</i> ‘pure and transparent’	143	0.02
16. <i>puro e disinteressato</i> ‘pure and disinterested’	143	0.02
17. <i>puro e casto</i> ‘pure and chaste’	141	0.02

The adjective *puro*, which is highly polysemous, together with the coordinating conjunction *e* ‘and’, is used in the construction to express a great degree or intensity of the quality denoted by the adjective represented in the third slot:

(19) *Una serata di puro e passionale tango*
 an evening of pure and passionate tango
 ‘An evening of extremely passionate tango’
 (itTenTen16, 4532359)

- (20) *Una riserva di energia pura e naturale*
 A reserve of energy pure and natural
 ‘A reserve of totally natural energy’
 (itTenTen16, 1529564)

A similar pattern is represented by the coordinated construction employing the polysemous adjective *semplice* ‘simple’, as in the following examples:

- (21) [*semplice e* + X_{ADJECTIVE<qualifying>}] = ‘totally/extremely X_{ADJECTIVE}’

Tab. 3: Partially filled constructions employing the adjective *semplice* in the first slot

Partially filled unit	Absolute frequency	Normalised frequency
1. <i>semplice e chiaro</i> ‘simple and clear’	3,741	0.63791
2. <i>semplice e intuitivo</i> ‘simple and intuitive’	3,340	0.56953
3. <i>semplice e lineare</i> ‘simple and linear’	1,873	0.31938
4. <i>semplice e genuino</i> ‘simple and genuine’	1,695	0.28903
5. <i>semplice e naturale</i> ‘simple and natural’	1,629	0.27777
6. <i>semplice e facile</i> ‘simple and easy’	1,272	0.21690
7. <i>semplice e comprensibile</i> ‘simple and understandable’	871	0.14852

However, in this case the adjective in the third slot (C) reduplicates the meaning of *semplice* ‘simple’, adding a specific semantic surplus. The construction is used to express a high degree of the quality denoted by the adjective represented in both A and C slots:

- (22) *E’ forse l’antivirus più semplice e facile da installare*
 is maybe the antivirus more simple and easy to install
 ‘It is maybe the simplest antivirus to install’
 (itTenTen16, 781079)
- (23) *La procedura è semplice e intuitiva*
 The procedure is simple and intuitive
 ‘The procedure is very intuitive’
 (itTenTen16, 58616138)

In some cases, the adjective included in the fixed slot may be characterised by a peculiar semantic trait and this may have effects on its combinatorial proper-

ties. The example below concerns the adjective *lungo* ‘long’, which is used in the coordinated intensifying construction to describe a process which lasts for a long time, and at the same time to qualify it as negative. This is confirmed by the adjectives it is most frequently associated with: as the examples below show, they tend to convey negative evaluations (e.g. *complesso* ‘complex’, *faticoso* ‘strenuous’, *difficile* ‘difficult’, *tortuoso* ‘contorted’).

(24) [*lungo e* + X_{ADJECTIVE<qualifying_negative>}] = ‘extremely X_{ADJECTIVE}’

Tab. 4: Partially filled constructions employing the adjective *lungo* in the first slot

Partially filled unit	Absolute frequency	Normalised frequency
1. <i>lungo e complesso</i> ‘long and complex’	3,449	0.58812
2. <i>lungo e faticoso</i> ‘long and strenuous’	2,725	0.46466
3. <i>lungo e difficile</i> ‘long and difficult’	2,622	0.44710
4. <i>lungo e tortuoso</i> ‘long and contorted’	1,009	0.17205
5. <i>lungo e articolato</i> ‘long and articulated’	896	0.15278
6. <i>lungo e impegnativo</i> ‘long and demanding’	849	0.14477
7. <i>lungo e laborioso</i> ‘long and laborious’	811	0.13829
8. <i>lungo e costoso</i> ‘long and expensive’	795	0.13556
9. <i>lungo e doloroso</i> ‘long and painful’	792	0.13505
10. <i>lungo e complicato</i> ‘long and complicated’	738	0.12584
11. <i>lungo e noioso</i> ‘long and boring’	712	0.12141

Selected data show that in the third slot there is a predominance of adjectives having a negative connotation – or a negative *semantic prosody*³³, in the terms of Louw (1993) and Sinclair (1998).

The different groups of partially filled constructions share similar morpho-syntactic and semantic features; in particular, their semantics and syntactic configuration is able to convey particular sets of meanings. This means that, at the linguistic level, some linguistic structures (syntactic *formats*, in terms of Simone 2007), tend to be typically associated with specific semantic values; at the cognitive-conceptual level, semantic and syntactic pairings – or, semantic

³³ *Semantic prosody* has been defined as the positive or negative evaluation conveyed by words. It is a discourse function used to describe implicitly the communicative purposes of the speaker (Lindquist and Levin 2018).

and syntactic *formats*, Simone 2007) – are able to structure and organise the semantic information associated with a lexeme. New word combinations can develop from specific existing constructions, and patterns may conventionalise through repetition (Bybee 2010). Thus, frequency of use is concerned with pattern productivity (Traugott and Trousdale 2013: 18): new constructions may “spread by gradually increasing their frequency of use over time” (Bybee and McClelland 2005: 387). An increase in the frequency of use of a construction and of new instances produced from it³⁴ may correspond to an expansion of domains of use of that construction and a consequent increase in productivity (Traugott and Trousdale 2013: 18).³⁵ The notion of *pattern productivity* recalls other peculiarities of partially filled constructions, and in particular their *schematicity*. Thus, a further distinction has to be made between completely and partially filled coordinated CIs. As shown, the former are characterised by a strict reduction of flexibility, as they are fixed and invariable; in fact, complete CIs are non-schematic syntactically complex units (behaving as a *single complete lexical units* in the terms of Lipka 2002) which have undergone a process of lexicalisation, becoming both syntactically and semantically fixed (Blank 2001, Trousdale and Traugott 2013). Even though single constructs may originate from schematic and productive constructions (Goldberg 1995), they have become fully lexicalised units, which are semantically unanalysable and generally lacking productivity and schematicity.³⁶ As Trousdale and Traugott (2013: 177) point out “[t]hese are now non-compositional atomic referring micro-constructions that are relics from the loss of the subschema that sanctioned them”. On the contrary, partially filled constructions are a sort of *constructional sets* (Traugott and Trousdale 2013), “a series of related constructions at different degrees of schematicity, clustered around a particular node in the constructional network” (Traugott and Trousdale 2013: 28). They represent linguistic schemas, “abstractions across sets of constructions which are (unconsciously) perceived by language-users to be closely related to each other in the constructional network” (Traugott and Trousdale 2013: 14). Partially filled CIs may be characterised by different degrees of abstraction and lexical specificity; their schematic construction is often analysable in syntactic and semantic terms. Even though single constructs are instances of discourse and dispel after the utterance, coordinated

³⁴ As Traugott and Trousdale (2013: 18-19) note, the difference between a construction and its instances may be compared to the type/token distinction (Baayen 2001; Bybee 2010).

³⁵ Cf. also Ziem and Lasch (2013), and López Meirama and Mellado Blanco (2018).

³⁶ An in-depth analysis of the developments of lexicalisation is offered in Trousdale and Traugott (2013).

constructional patterns may become conventionalised and their abstract schema can extend to a range of less general linguistic sequences. Thus, a number of single instances of lexically specified constructions may arise on the basis of the productivity of the abstract pattern (López Meirama and Mellado Blanco 2018). In particular, the scheme of coordinated intensifiers is frequently attested in Italian, and comprehends a number of more specific constructions, through which it is possible to obtain different linguistic instances.

It is therefore possible to draw a lexicon-syntax *continuum* (that in some cases could also be understood as a diachronic derivation), where the two different classes of CIs can be placed. The *continuum* takes into account the linguistic level (the level of lexically filled single constructs, that is, the individual instances originated from the abstract schemes). Single constructs are represented as characterised by different degrees of i) lexical variation, ii) syntactic cohesion, iii) idiomaticity and iv) lexicalisation. At the extreme poles of the *continuum* completely and partially filled intensifying constructions can be placed, as the graph shows:

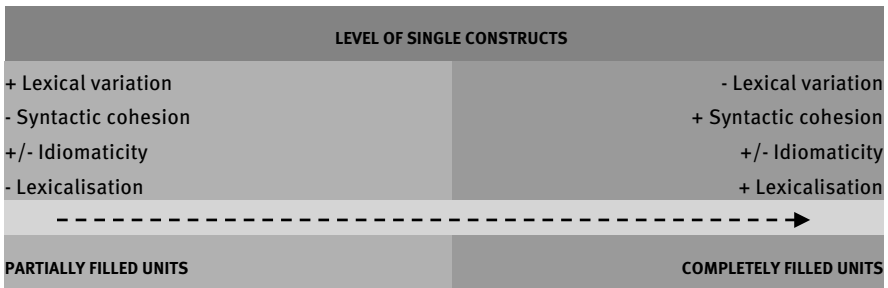


Fig. 2: *Continuum* of lexicalisation

On the left pole, partially filled CIs are characterised by i) lexical variation (the presence of at least one empty position), (ii) a low degree of syntactic cohesion between constituents, (iii) different degrees of idiomaticity (their meaning is not always compositional, but is often constructional). Partially filled CIs are typically non-lexicalised items deriving from more abstract constructional schemes, namely syntactic-semantic templates through which new (syntactically and semantically similar) combinatorial constructs can be created. On the right extreme of the *continuum*, completely filled CIs are characterised by i) lexical invariability, (ii) strong syntactical cohesion, (iii) variable degrees of idiomaticity. Completely filled CIs are fully lexicalised sequences, which may derive from

productive abstract constructions which are often no longer traceable, and their semantics is often idiosyncratic.

4.3 Intensified constructions or constructional intensifiers?

Intensifiers represent a heterogeneous group even in functional terms. It has been demonstrated to a large degree that intensifiers can mitigate or strengthen the meaning of the element they modify,³⁷ and that they may even “assume different textual and pragmatic meanings on the basis of context-related factors” (Napoli and Ravetto 2017: 3). According to Grandi (2017) it is possible to distribute intensification strategies on a cline of intensifying power, from the weakest to the strongest one.³⁸ As Grandi observes, “intensification is a semantic/functional category characterised by different degrees of membership and by sub-types of intensive values” (2017: 67).

The set of coordinated CIs is extremely interesting from the functional point of view. In particular, it includes two functional categories, which differ in terms of syntactic function and distributional features: i) already *intensified* constructions, and ii) pure CI.

On the one hand, the former group includes constructional sequences, which are able to attribute an intensifying value to the linguistic element they select and co-occur with inside the construction; thus, they operate as autonomous constructional modifiers³⁹ of other lexical elements. They can be defined as *intensified* phraseological constructions, since they include an intensified meaning, as happens in the following sequences:

- (25) a. *unico e irripetibile*
 ‘unique and unrepeatable’
 b. *una stagione unica^{INTENSIFIED} e irripetibile^{INTENSIFIER}*
 a season unique and unrepeatable
 ‘a totally unique season’
 (itTenTen16, 34682682)

³⁷ *Inter alia*, cf. Bolinger (1972).

³⁸ According to Grandi’s research (2017), for example, the weakest Italian intensification tools are represented by reduplication and Wh-exclamatives, while morphological tools (e.g. prefixes as *stra-* and the suffix *-issimo*) and idiomatic expressions (e.g. *freddo cane* ‘freezing cold’, *lit.* cold dog) can be conceived of as the strongest strategies.

³⁹ For the notion of “constructional modification” cf. Piunno (2018a).

- (26) a. *puro e cristallino*
 ‘pure and crystal clear’
 b. *rendi pura*^{INTENSIFIED} *e cristallina*^{INTENSIFIER} *la mia fede*
 make pure and crystalline the my faith
 ‘make my faith extremely pure’
 (itTenTen16, 284905653)

In such cases intensifier and intensified belong to the same construction. Their function partially⁴⁰ corresponds to the one of synthetic intensified elements, such as those related to evaluative morphology and containing prefixes or suffixes (examples taken from Grandi 2017: 58-59):

- (27) a. *stra-simpatico*
 ‘very nice’ (*lit.* INTENS-nice)
 b. *rossissimo*
 ‘very red’ (*lit.* red-INTENS)

In such cases, the intensification affects the inner layer of the construction. The intensifying construct is functionally equivalent to an affixed linguistic element:

- (28) a. **A e C = b. prefix-C / C-suffix**
 aⁱ. *puro e sincero*
 ‘extremely sincere’
 bⁱ. *stra-sincero/sincerissimo*
 ‘very sincere’

It is important to bear in mind that, in this case, the semantics is fully specified *inside* the construct itself. Thus, *sincero* ‘sincere’ conveys the overall semantics that can be intensified by the constructional scheme in (28aⁱ) and through evaluative morphology in (28bⁱ).

On the other hand, pure CI are an analytical means of intensification of linguistic elements not included in the sequence. Namely, as single and independent lexical elements, they play the role of intensifiers of linguistic units which are completely external to the construction.

⁴⁰ In fact, CI also bear full semantics, thus, the functional overlap between the two entities is not complete.

- (29) a. *bello e buono*
 ‘downright’ (*lit.* beautiful and good)
- b. *è una truffa*^{INTENSIFIED} ***bella e buona***^{INTENSIFIER}
 is a fraud beautiful and good
 ‘it is a downright fraud’
 (itTenTen16, 37556920)
- (30) a. *vero e proprio*
 ‘out-and-out’ (*lit.* true and proper)
- b. ***Veri e propri***^{INTENSIFIER} *luoghi*^{INTENSIFIED} *dell’aggregazione sociale*
 true and proper places of aggregation social
 ‘Out-and-out places of social aggregation’
 (itTenTen16, 8841915)
- (31) a. *calzato e vestito*
 ‘completely’ (*lit.* shod and dressed)
- b. *ha fatto la figura dello sfigato*^{INTENSIFIED} ***calzato e vestito***^{INTENSIFIER}
 have done the figure of.the loser shod and dressed
 ‘He/She made himself/herself look like a total loser’
 (itTenTen16, 4176783449)
- (32) a. *nudo e crudo*
 ‘true’ (*lit.* nude and raw)
- b. *Un affresco*^{INTENSIFIED} ***nudo e crudo***^{INTENSIFIER} *della società*
 A fresco nude and raw of.the society
 ‘A true picture of the society’
 (itTenTen16, 453548641)

In such cases the construction only includes an intensifier, which selects the intensified elements outside the construction. It is worth noting here that the examples belonging to this group are completely filled units, thus, fully lexicalised and highly cohesive. Their single components have completely lost their original semantics, thus developing, in the context of the specific construct, a new functional value. They are *pure* intensifiers, in the sense that they can be used to modify another lexical element, thus transferring the intensifying value to it. In such cases, the intensification affects the outer layer of the construction and the intensifying construct is functionally equivalent to an affix:

(33) a. **A e C = b. prefix/suffix**aⁱ. *bello e buono*

‘downright’

bⁱ. *stra-/ -issimo*

‘very’

This presupposes a mismatch between the original sequence and the meaning/function of the construct. Such sequences have been subject to a constructional semantic *neoanalysis*⁴¹, namely a micro-step change in their meaning/function, which has resulted in the creation of “a form_{new}-meaning_{new} type” (Traugott and Trousdale 2013: 75).

5 Conclusion

This paper has been devoted to the analysis of a particular type of construction, having an intensifying function and the coordinated structure [Lexeme_A + Conjunction + Lexeme_B]. Coordinated constructional intensifying units have been distinguished into different classes, according to morpho-syntactic, semantic and functional parameters, such as: i) their syntactic pattern (i.e. the exact sequence of PoS), ii) the semantics of the lexical words in slots A and C, iii) semantic analysability, iv) the degree of lexicalisation, and v) their functional value. Thus, a *continuum* of lexicalisation has been postulated, and two macro-types of coordinated intensifying sequences have been placed at the opposite poles: *completely* and *partially filled* ones (Fig. 2). The former are fully lexicalised phraseological units, whose semantics is often not predictable (they often carry an idiomatic and - apparently - idiosyncratic meaning); it has been noted that there may be a strict relation between the degree of lexicalisation and the functional value of such units: some fully-lexicalised sequences (e.g. *bello e buono*) have been subject to semantic neoanalysis, and a new constructional intensifying value has arisen. The latter are more abstract constructions from which non-lexicalised constructs may originate; they contain empty positions, that can be generally filled with semantically related items. On the one hand, new combinatory forms can be created from the lexically-empty patterns and new abstract constructions and productive schemes can arise; in these terms, variation can be interpreted as an index of routinisation and conventio-

41 Cf. Andersen (2001), Traugott and Trousdale (2013).

nalisation. On the other hand, some constructs deriving from partially filled schemes may lexicalise and the related constructional scheme may lose its productivity.

Thus, beyond the criteria listed above, i) productivity, ii) semantic predictability, iii) lexical variability, and iv) schematicity have also been considered. Constructional schemes can consequently be put along a *continuum* of constructionalisation, having at the opposite poles non-productive configurations (to the left pole of the scale), and highly schematic and productive patterns (to the right pole).

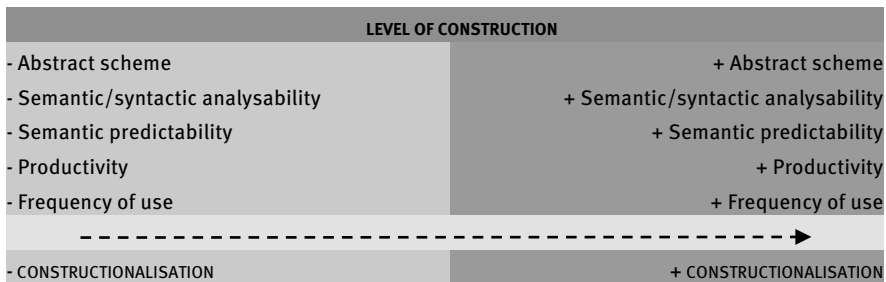


Fig. 3: *Continuum* of constructionalisation

In conclusion, even though coordinated constructional intensifying units represent only a minimal cluster of the bigger set of Italian intensifying strategies, the analysis has shown that this group of constructions contains heterogeneous entities. This group represents just one of the pieces of a more complex (and “chaotic”, cf. Grandi 2017: 58) domain.

Finally, even though only Italian has been selected for this investigation, the phenomenon is not peculiar to this language. As a matter of fact, it is frequently attested in other languages and further investigation is needed in this sense, from an interlinguistic and contrastive perspective.

References

- Andersen, Henning. 2001. Actualization and the (uni)directionality. In Henning Andersen (ed.), *Actualization: Linguistic Change in Progress*, 225–248. Amsterdam & Philadelphia: Benjamins.
- Ansamble Jean-Claude & Irène Tamba (eds.). 2013. *L'intensification*. [Special issue]. *Langue Française* 177.

- Baayen, R. Harald. 2001. *Word Frequency Distributions*. Dordrecht: Kluwer.
- Barðdal, Johanna. 2008. *Productivity. Evidence from Case and Argument Structure in Icelandic*. Amsterdam & Philadelphia: Benjamins.
- Benigni, Valentina. 2017. Una festa da paura! Mi sono divertito da morire! Gli intensificatori iperbolici dell'italiano e la loro resa in russo. *Studia de Cultura* 9 (1). 5–18.
- Benveniste, Émile. 1967. *Problèmes de linguistique générale*. Paris: Gallimard.
- Berlanda, Sara. 2013. Constructional Intensifying Adjectives in Italian. In Valia Kordoni, Carlos Ramisch & Aline Villavicencio (eds.), *Proceedings of the 9th Workshop on Multiword Expression*, 132–137. Stroudsburg: The Association for Computational Linguistics.
- Berrendonner, Alain. 2002. Les deux syntaxes. *Verbum* 24. 23–25.
- Blanche-Benveniste, Claire. 2003. Le recouvrement de la syntaxe et de la macrosyntaxe. In Antonietta Scarano (ed.), *Macro-syntaxe et pragmatique. L'analyse linguistique de l'oral*, 53–75. Roma: Bulzoni.
- Blank, Andreas. 2001. Pathways of lexicalization. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language Typology and Language Universals*, Vol. II, 1596–1608. Berlin & New York: Mouton de Gruyter.
- Bolinger, Dwight. 1972. *Degree words*. The Hague: Mouton.
- Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
- Bybee, Joan L. & James L. McClelland. 2005. Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. In Nancy A. Ritter (ed.), *The Role of Linguistics in Cognitive Science*. [Special Issue]. *The Linguistic Review* 22. 381–410.
- Carter, Ronald. 1987. *Vocabulary*. London & New York: Routledge.
- Cimaglia, Riccardo. 2011. Intensificatori. In Raffaele Simone (ed.), *Enciclopedia dell'italiano*, 666–667. Roma: Istituto dell'Enciclopedia Italiana.
- Conzett, Jane. 1997. Integrating collocation into a reading and writing course. In James Coady & Thomas Huckin (eds.), *Second language vocabulary acquisition*, 70–87. Cambridge: Cambridge University Press.
- Cruse, David A. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.
- De Mauro, Tullio (ed.). 1999. *Grande Dizionario Italiano dell'Uso (GRADIT)*. Torino: Utet.
- De Santis, Cristiana. 2011. Reduplicazione espressiva. In Raffaele Simone (ed.), *Enciclopedia dell'italiano*, 1224–1225. Roma: Istituto dell'Enciclopedia Italiana.
- Detges, Ulrich & Richard WALTEReit. 2002. Grammaticalization vs. Reanalysis: a Semantic-Pragmatic Account of Functional Change in Grammar. *Zeitschrift für Sprachwissenschaft* 21. 151–195.
- Dixon, Robert M. W. 2004. Adjective Classes in Typological Perspective. In Robert M. W. Dixon & Alexandra Y. Aikhenvald (eds.), *Adjective Classes: A Cross-Linguistic Typology*, 1–49. Oxford: Oxford University Press.
- Fillmore, Charles J., Paul Kay & Mary Catherine O'Connor. 1988. Regularity and idiomacity in grammatical constructions: the case of *let alone*. *Language* 64. 501–538.
- Givón, Talmy. 1979. *On Understanding Grammar*. New York - San Francisco & London: Academic Press.
- Goldberg, Adele E. 1995. *Constructions: a Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldberg, Adele E. 2006. *Constructions at work*. Oxford: Oxford University Press.

- Grandi, Nicola. 2017. Intensification processes in Italian. In Maria Napoli & Miriam Ravetto (eds.), *Exploring Intensification. Synchronic, diachronic and cross-linguistic perspectives*, 55–77. Amsterdam & Philadelphia: Benjamins.
- Gustaffson, Marita. 1975. *Binomial expressions in present-day English*. Turku: Turun Yliopisto.
- Halliday, Michael A.K. 1976. *System and function in language*. Oxford: Oxford University Press.
- Hardy, Heather & Timothy Montler. 1988. Imperfective gemination in Alabama. *IJAL* 54. 399–415.
- Hopper, Paul J. & Sandra A. Thompson. 1984. The discourse basis for lexical categories in Universal Grammar. *Language* 60 (4). 703–752.
- Inkelas, Sharon & Laura Downing. 2015a. What is Reduplication? Typology and Analysis, Part 1/2: The Typology of Reduplication. *Language and Linguistics Compass* 9 (12). 502–515.
- Inkelas, Sharon & Laura Downing. 2015b. What is Reduplication? Typology and Analysis, Part 2/2: The Analysis of Reduplication. *Language and Linguistics Compass* 9 (12). 516–528.
- Jakubíček, Miloš, Adam Kilgarriff, Vojtěch Kovář, Pavel Rychlý & Vít Suchomel. 2013. The Ten-Ten Corpus Family. In Andrew Hardie & Robbie Love (eds.), *7th International Corpus Linguistics Conference CL 2013*, 125–127. Lancaster: UCREL.
- Kilgarriff, Adam, Pavel Rychlý, Pavel Smrž & David Tugwell. 2004. The Sketch Engine. *Information Technology* (2004). 105–116.
- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý & Vít Suchomel. 2014. The Sketch Engine: ten years on. *Lexicography* 1. 7–36.
- Kopaczyk, Joanna & Hans Sauer. 2017. Defining and Exploring Binomials. In Joanna Kopaczyk & Hans Sauer (eds.), *Binomials in the History of English. Fixed and Flexible*, 1–23. Cambridge: Cambridge University Press.
- Labov, William. 1984. Intensity. In Deborah Schiffrin (ed.), *Meaning, Form, and Use in Context: Linguistic Applications*, 43–70. Washington: Georgetown University Press.
- Landheer, Ronald. 1989. L'importance des relations hyponymiques dans la description lexicographique. In Dieter Kremer (ed.), *Actes du XVIIIe Congrès International de Linguistique et de Philologie Romanes*, 139–151. Tübingen: Max Niemeyer Verlag.
- Lehmann, Christian. 2005. *Pleonasm and hypercharacterisation*. In Geert Booij & Jaap Van Marle (eds.), *Yearbook of Morphology*, 119–154. Dordrecht: Springer.
- Lenci, Alessandro, Francesca Masini, Malvina Nissim, Sara Castagnoli, Gianluca E. Lebani, Lucia C. Passaro & Marco S. G. Senaldi. 2017. How to harvest Word Combinations from corpora. Methods, evaluation and perspectives. *Studi e Saggi Linguistici* 55 (2). 45–68.
- Lindquist, Hans & Magnus Levin. 2018. *Corpus Linguistics and the Description of English*. Edinburgh: Edinburgh University Press.
- Lipka, Leonhard. 2002. *English Lexicology: Lexical Structure, Word Semantics and Word formation*. Tübingen: Max Niemeyer Verlag.
- López Meirama, Belén. 2020. [a todo + INF]: velocidad e intensificación en una construcción fraseológica del español. In Mellado Blanco, Carmen (ed.), *Romanica Olomucensia* 32 (1) [Special Issue]: *Nuevas aportaciones de la Gramática de Construcciones a los estudios de fraseología en las lenguas románicas*. 91–110.
- López Meirama, Belén & Carmen Mellado Blanco. 2018. Las construcciones [de X a Y] y [de X a X]: realizaciones idiomáticas y no tan idiomáticas. In Marta Díaz, Gael Vaamonde, Ana Varela, M^a Carmen Cabeza, José M. García-Miguel & Fernando Ramallo (eds.), *Actas do XIII Congreso Internacional de Lingüística Xeral*, 576–583. Vigo: Universidade de Vigo.

- Louw, Bill. 1993. Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In Mona Baker, Gill Francis & Elena Tognini Bonelli (eds.), *Text and technology. In honour of John Sinclair*, 157–176. Amsterdam - Philadelphia: Benjamins.
- Lyons, John. 1981. *Language and Linguistics. An introduction*, Cambridge: Cambridge University Press.
- Malkiel, Y. 1959. Studies in irreversible binomials. *Lingua* 8. 113–160.
- Marantz, Alec. 1982. Reduplication. *Linguistic Inquiry* 13 (3). 435–482.
- Masini, Francesca. 2006. Binomial Constructions: Inheritance, Specification and Subregularities. *Lingue e linguaggio* V (2). 207–232.
- Mel'čuk, Igor & Alain Polguère. 2007. Chapitre 1. Notions de base. In Igor Mel'čuk & Alain Polguère (eds.), *Lexique actif du français*, 13–28. Louvain-la-Neuve: De Boeck Supérieur.
- Mellado Blanco, Carmen. 2020a. Esquemas fraseológicos y construcciones fraseológicas en el continuum léxico-gramática. In Carsten Sinner, Encarnación Tabares Plasencia & Esteban T. Montoro del Arco (eds.), *Clases y categorías en la fraseología española*, 13–36. Frankfurt am Main: Peter Lang.
- Mellado Blanco, Carmen. 2020b. (No) *me importa un comino* y sus variantes diatópicas. Estudio de corpus desde la Gramática de Construcciones. *Estudios de Lingüística (ELUA)* VII. 87–109.
- Michaelis, Laura A. 2019. Constructions are Patterns and so are Fixed Expressions. In Beatrix Busse & Ruth Möhlig-Falke (eds.), *Patterns in Language and Linguistics*, 193–220. Berlin & New York: Mouton de Gruyter.
- Mollica, Fabio & Elmar Schafroth. 2018. Der Ausdruck der Intensivierung in komparativen Phrasem-Konstruktionen im Deutschen und im Italienischen: eine konstruktionsgrammatische Untersuchung. In Kathrin Steyer (ed.), *Sprachliche Verfestigung. Wortverbindungen, Muster, Phrasem-Konstruktionen*, 103–136. Tübingen: Narr.
- Moravcsik, Edith A. 1978. Reduplicative Constructions. In Joseph H. Greenberg (ed.), *Universals of Human Language*, Vol. 3: *Word Structure*, 249–289. Stanford: Stanford University Press.
- Napoli, Maria & Miriam Ravetto (eds.). 2017. *Exploring Intensification. Synchronic, diachronic and cross-linguistic perspectives*. Amsterdam & Philadelphia: Benjamins.
- Partington, A. 1998. *Patterns and meanings: Using corpora for English language research and teaching*. Amsterdam & Philadelphia: Benjamins.
- Piunno, Valentina. 2016. CombiNet. A Corpus-based Online Database of Italian Word Combinations. In Ilan Kernerman, Iztok Kosem, Simon Krek & Lars Trap-Jensen (eds.), *GLOBALEX 2016. Lexicographic resources for human language technology*, 45–51. Paris: European Language Resources Association.
- Piunno, Valentina. 2018a. *Sintagmi preposizionali con funzione aggettivale e avverbiale* München: LINCOM Studies in Romance Linguistics.
- Piunno, Valentina. 2018b. Negated Multiword Expressions. Types, properties and lexicalization degrees. In Natalia Filatkina & Sören Stumpf (eds.), *Konventionalisierung und Variation / Conventionalization and variation*, 125–147. Frankfurt am Main: Peter Lang.
- Piunno, Valentina. 2020. Le combinazioni di parole parzialmente riempite in alcune lingue romanze. Schematismo e predicibilità semantica. In Mellado Blanco, Carmen (ed.), *Romanica Olomucensia* 32 (1) [Special Issue]: *Nuevas aportaciones de la Gramática de Construcciones a los estudios de fraseología en las lenguas románicas*. 143–171.

- Rapatel, Philippe 2015. De la délexicalisation / grammaticalisation à l'intensification. *Hal archives-ouvertes.fr*. 1–13. <https://halshs.archives-ouvertes.fr/halshs-01066454/document> (accessed 31 October 2020).
- Roberts, Ian. 2010. Grammaticalization, the clausal hierarchy and semantic bleaching. In Elizabeth Closs Traugott & Graeme Trousdale (eds.), *Gradience, Gradualness, and Grammaticalization*, 45–73. Amsterdam & Philadelphia: Benjamins.
- Rosch, Eleanor. 1978. Principles of categorization. In Eleanor Rosch & Barbara B. Lloyd (eds.), *Cognition and Categorization*, 27–48. Hillsdale: Lawrence Erlbaum.
- Rubino, Carl. 2005. Reduplication: Form, function and distribution. In Bernhard Hurch (ed.), *Studies on Reduplication*, 11–30. Berlin & New York: Mouton de Gruyter.
- Schafroth, Elmar. 2020. Fraseologismi a schema fisso – basi teoriche e confronto linguistico. In Mellado Blanco, Carmen (ed.), *Romanica Olomucensia* 32 (1) [Special Issue]: *Nuevas aportaciones de la Gramática de Construcciones a los estudios de fraseología en las lenguas románicas*. 173–200.
- Simone, Raffaele. 2007. Constructions and categories in verbal and signed languages. In Paola Pietrandrea, Elena Pizzuto & Raffaele Simone (eds.), *Verbal and Signed Languages. Comparing Structures, Constructs, and methodologies*, 198–252. Berlin & New York: Mouton de Gruyter.
- Simone, Raffaele & Francesca Masini. 2014. On Light Nouns. In Raffaele Simone & Francesca Masini (eds.), *Word Classes: Nature, Typology and Representations*, 51–73. Amsterdam & Philadelphia: Benjamins.
- Simone, Raffaele & Valentina Piuino. 2017. Combinazioni di parole che costituiscono entrata. Rappresentazione lessicografica e aspetti lessicologici. *Studi e Saggi Linguistici* 55 (2). 13–44.
- Sinclair, John. 1998. The lexical item. In Edda Weigand (ed.), *Contrastive Lexical Semantics*, 1–24. Amsterdam & Philadelphia: Benjamins.
- Steyer, Kathrin. 2013. *Usuelle Wortverbindungen. Zentrale Muster des Sprachgebrauchs aus korpusanalytischer Sicht*. Tübingen: Narr.
- Steyer, Kathrin. 2018 (ed.). *Sprachliche Verfestigung. Wortverbindungen, Muster, Phrasem-Konstruktionen*. Tübingen: Narr.
- Stolz, Thomas, Cornelia Stroh & Aine Urdze. 2011. *Total reduplication: the areal linguistics of a potential universal*. Berlin: Akademie Verlag.
- Tognini Bonelli, Elena. 2001. *Corpus linguistics at work*. Amsterdam - Philadelphia: Benjamins.
- Traugott, Elizabeth Closs & Graeme Trousdale. 2013. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.
- Van Os, Charles. 1989. *Aspekte der Intensivierung im Deutschen*. Tübingen: Narr.
- Xiao, Richard & Tony McEnery. 2006. Collocation, semantic prosody and near synonymy: A cross-linguistic perspective. *Applied Linguistics* 27 (1). 103–129.
- Ziem, Alexander & Alexander Lasch. 2013. *Konstruktionsgrammatik. Konzepte und Grundlagen gebrauchsbasierter Ansätze*. Berlin & New York: Mouton de Gruyter.

Kathrin Steyer

Preposition-noun combinations of TIME in German

A pattern-based approach to minimal phraseological units

Abstract: This paper deals with a specific type of lexeme, namely binary preposition-noun combinations containing temporal references like *am Ende* [at (the) end] or *für Sekunden* [for seconds].¹ The main characteristic of these combinations is the recurrent internal zero gap. Despite the fact that the omission of the determiner can often be explained by grammatical rules, the zero gaps indicate a higher degree of lexicalization. Therefore, we interpret these expressions as minimal phraseological units with holistic meanings and functions. The corpus-driven exploration of typical context patterns (e.g. using collocation profiles and the *lexpan* slot filler analysis) shows that a) even such minimal expressions are based on semi-abstract schemes and b) temporal expressions can also fulfill modal or discursive functions, usually with fuzzy borders and overlapping structures. In the case of modalization or pragmatization one can regard such PNs as distinct lexicon entries.²

Keywords: preposition-noun combinations, temporal phraseological units, German phraseological patterns.

1 This paper is a slightly revised and abridged English version of my German publication “Für Jahre vom Tisch sein. Temporale Präposition-Nomen-Verbindungen zwischen Zeitreferenz und modal-diskursivem Gebrauch” (Steyer 2019). Special thanks to Annelen Brunner and Marcas Mac Coinnigh for reading this manuscript and for giving valuable advice concerning the correctness and comprehensibility of this English version.

2 English translations are marked by square brackets; meanings by single quotes, and patterns by angle brackets. For the English translation and equivalents (marked by ee), we used – among others – the *Oxford Idioms Dictionary* (OID 2006) and the English web corpus *en-TenTen15* in *Sketch Engine* (over 13 billion tokens, November 2019) (see *SkE*).

Kathrin Steyer, Leibniz-Institut für Deutsche Sprache, Mannheim, steyer@ids-mannheim.de

<https://doi.org/10.1515/9783110520569-007>

1 Object of study

1.1 PNs as autonomous lexical units

Binary preposition-noun combinations (PNs) are understudied in phraseology and lexicography, particularly with regard to their status as minimal holistic units, as lexemes.

PNs were and are discussed primarily from the grammatical point of view as a subtype of prepositional phrases. German grammar traditionally focused on the regularities of the omission of a determiner in front of the noun, the so called ‘zero article’ (Helbig and Buscha 2013). As central criteria for the omission of determiners, the plural use of a noun (*nach Jahren* [after years]) or its uncountable status (*mit Genugtuung* [with satisfaction]) were defined. Besides general grammar books, Kiss’ study (2011) is one of the few scholarly papers on German PNs. Kiss examined the circumstances in which it may be possible to have combinations containing a preposition plus countable noun.³

Our focus here is not on the preposition as a part of speech. We are interested in mechanisms of lexicalizing PNs, in the holistic quality of PNs that makes them phrasemes or multi-word expressions. The question of the lexicalization of word groups, the question of how they turn into more or less holistic units, has been a crucial question of phraseology from its inception (Burger et al. 2007). In light of new theories of fixedness and entrenchment this question is more relevant than ever, even for such minimal units as PNs.

With regard to PNs, it may be asked: under which circumstances does a PN become an autonomous semantic unit with a holistic quality? Which components are mandatory for the status as a multi-word expression? The following two fictional examples illustrate the main criteria:

- (1) *unter Aufsicht* [under supervision]
*Die Kinder sind oft auf dem Spielplatz, aber immer **unter Aufsicht**.*
 [The children are often at the playground but always **under supervision**.]

- (2) *unter Leitung* [under guidance]
*Ein Ausschuss **unter Leitung des Präsidenten** prüft den Plan.*
 [A committee **under guidance of the president** examines the plan.]

³ Kiss et al. (2016) also published a comprehensive handbook of meanings of German prepositions.

The PN in (1) has an autonomous status as an adverb with the meaning ‘beaufsichtigt’ [supervised]. The PN in (2) is a non-autonomous component **unter Leitung* [*under guidance] because of its mandatory NP (*des Präsidenten*) [under guidance of the president]. Other excluding criteria are cases in which PNs are parts of:

- a VP, e.g. *um Hilfe bitten*
[to ask for help]
(**um Hilfe* [*for help])
- a LP (see 1.2), e.g. <*um N {Hilfe/Geld/Erlaubnis} bitten*>
[to ask for N[help/money/permission]]
(**um N* [*for N])⁴
- a copular phrase *am Ende sein*
[to be at (the) end]
(**am Ende* [*at (the) end])
- more complex unit *von Tag zu Tag*
[from day to day]
(**von Tag *zu Tag* [*from day *to day])

Autonomous PNs can be used as simple adverbials in addition to binary word classes like binary adverbs or discourse markers, e.g.

(3) temporal

<i>an Weihnachten</i>	[at Christmas]
<i>auf Dauer</i>	[ee: for good]
<i>nach Jahren</i>	[after years]

(4) local

<i>vor Ort</i>	[ee: on site]
<i>an Bord</i>	[on board]
<i>ab Mannheim</i>	[from Mannheim]

(5) modal

<i>durch Zufall</i>	[ee: by chance]
<i>mit Genugtuung</i>	[with satisfaction]
<i>nach Belieben</i>	[at will]

⁴ The lexical unit *bitten um* belongs to the significant group of German phrasal verbs that feature noun combinations. These are explicitly excluded from our investigation.

Such PNs are – as evidenced by their high frequency – important building blocks of language use with a wide range of ambiguity and pragmatic value. Therefore, these entities often become obstacles in the effective use of a foreign language.

1.2 PNs as core of lexical patterns (LP)

Based on corpus exploration and particularly on the slot filler analysis of patterns it became clear that binary PNs – despite their high degree of lexicalization – are almost always subject to complex processes of internal or external extensions. These extensions do not arise by chance, but from recurrent use of many speakers in different communicative situations. As internal and external extensional lexical patterns they are also entrenched in the mental lexicon.

We define LPs as conventionalised, partially filled lexical schemes that are frozen by recurrent use. Recurrence is defined as the repeated appearance of similar linguistic structures in comparable contexts (Steyer 2013, 2015, 2020) (cf. “geprägte komplexe Ausdrucksmuster” by Feilke 1996: 187). Slots can be filled with specific lexical items. Such fillers have similar semantic and/or pragmatic characteristics, but do not necessarily belong to the same morpho-syntactic category. Sometimes they only have functional characteristics in common, which cannot be captured by a traditional ontology. In any case LPs are characterised by a holistic meaning and/or function on different levels of abstraction. Speakers are able to recall those schemes as lexicon entries and fill the gaps in a specific communicative situation in a functionally adequate way.⁵

It is obvious that there are strong cross-connections to Construction Grammar. Our pattern concept focuses much more on structures and interrelations of lexical items and seeks to contribute to a usage-based theory of lexis. This approach arose from the tradition of phraseology as a genuine discipline of lexicology. This does not mean that we deny the validity of syntactic analysis. Naturally, our explorations are based on syntactic structures. But the dominance of the syntactic view can induce us to overlook the complexity of lexical phenomena. Probably this is a heuristic problem of analysis: One cannot observe all phenomena with the same intensity but some of them must by necessity fade into the background (e.g. syntactical phenomena) for a much clearer observation of others (lexical structures and networks).

⁵ Some examples of pattern theories are Renouf and Sinclair (1991); Hunston and Francis (2000); Biber (2009) and Hanks (2013).

The central criterion ‘autonomy’ can also be related to semi-abstract lexical patterns where lexical components are mandatory for the holistic meaning, as in the case of the semi-abstract lexical pattern <in ADJ Zeit> [in ADJ time]. In German, one cannot reduce this scheme to the lexical core *in Zeit without loss of the autonomous status. It requires adjective fillers like *absehbarer*; *kurzer*; *nächster* [foreseeable; short; next] for the pattern meaning ‘forthcoming’. In contrast to this, the extensions of the pattern <am ADJ {anderen/unteren} Ende DET N_{LOCAL}{Gebäudes/Skala/Dorfstraße}> [at the ADJ {other/lower} end of DET N_{LOCAL} {building/scale/village street}] are optional for the core meaning (‘place, where something stops’). This distinction between mandatory lexical components and optional extensions is a central idea in the concept of ‘Lexical Patterns’ that we have explored in recent years.

In chapters 3 and 4 we use the example of temporal PNs to illustrate their dual status as autonomous entities as well as components of other lexical units and patterns. Temporal PNs are especially interesting because of the transition zones between specific time references and subjective perception of duration, often including connotative values.

2 Empirical data

This case study was performed within the project *Preposition-Noun Combinations in Context* (see *PREPCON*) that was carried out within our long-term project *Usuelle Wortverbindungen* (see *UWV*) at the Leibniz Institute for the German Language in Mannheim (IDS). The empirical basis of *PREPCON* is an exhaustive collection of German PNs and their usage patterns compiled by the *UWV* group. This inventory contains the following components:

- list of 80 German prepositions and their noun slots in immediate postposition in the tagged corpora of *DeReKo* (regarding *DeReKo* see footnote 9): all in all, nearly 80,000 PN combinations
- tables of all noun fillers ranked by frequencies plus KWICs
- documentation of KWIC lines, collocation profiles and filler tables of PNs

Based on these data we selected candidates for further linguistic studies and lexicographic representation: autonomous PNs in general and temporal PNs in particular (see 3.).

The underlying methodology (Steyer 2013, 2020) can be understood as a qualitative-quantitative corpus-driven approach that is strongly focused on the word form surfaces. For this, two corpus analyzing systems are used: the

DeReKo analysis tool *COSMAS II* (see CII) and the Concordance Tool in *Sketch Engine* (see *SkE*), more specifically the collocation analysis of the IDS (see Belica 1995) and in *Sketch Engine*. The IDS collocation analysis can be used to detect significant word pairs and multi-word expressions as well as recurrent syntagmatic context patterns. This method enables us to identify typical aspects of meaning and usage – the extension of the principle of contextualism to multi-word units. An important tool for the slot filler analysis is the *Lexical Pattern Analyzer* (see *lexpan*) that we have developed in the UWV research group. In our *PREPCON* project it was used for German, Slovakian and Spanish.⁶

Central results of our corpus-driven studies are presented on the platform *PREPCON^{online}*. The central idea is that authentic language mass data can be used not only as a basis for empirical analyses but also as a form of lexicographic information itself. *PREPCON^{online}* has three different forms of data representation: PNs explorative; PNs temporal; PNs contrastive (trilingual). The explorative module is online since 2017. It consists of a fully automated database of 80,000 German PN combinations (autonomous and as part of other constructions).⁷ One cannot only build up collections of local, temporal or modal PNs of German but also study the richness of the incredible number of idiomatic PNs. The trilingual module can be regarded as a pilot study for contrasting PN usage and lexical patterns in different languages based on our corpus-driven approach. The entries contain narrative descriptions, corpus citations, grouped snippets of collocation profiles (according the respective aspects of use) and slot filler tables with visualizations of convergences and divergences in meaning, pragmatics and lexical variance. The temporal module (also available online) is a semi-automatic database containing about 1,000 PNs that have to fulfill the two main criteria we discuss in the next chapters: status as autonomous lexical units and referring to TIME.

⁶ For this trilingual part of *PREPCON* we cooperated with Spanish colleagues (head of this group: Carmen Mellado Blanco, University of Santiago de Compostela) and Slovakian colleagues (head of this group: Peter Ďurčo; University of Ss. Cyril and Methodius Trnava).

⁷ This database was implemented by Annelen Brunner.

3 Temporal PNs

3.1 Introduction

The following note at the door of the IDS student assistant room is a good example of our object of study:

- (6) *Beim Verlassen des Zimmers **am Ende des Tages** bitte daran denken, alle drei Heizkörper auszudrehen. Sonst können wir die Erde auch direkt in die Tonne kloppen.*

[When leaving the room **at the end of the day**, remember to turn off all three radiators! Otherwise, we can throw our earth directly into the garbage bin.]

The phraseme in *die Tonne kloppen* is a slang expression for ‘have become useless’ or ‘to ruin’ or ‘to destroy’. Our focus lies on the multi-word expression *am Ende des Tages* [at (the) end of the day] that refers to TIME in the aforementioned note, that is, ‘when the working day is over’. The temporal meaning still dominates in German in contrast to the English equivalent *at the end of the day*, meaning ‘finally’ or ‘eventually’. The following is a typical example of English usage:

- (7) *At the end of the day we all know that people use their phones in many different ways these days.*
(enTenTen15, 2480594)

However, this discourse marker function has also been present in German since the late 1990s as the IDS *Dictionary of Neologisms*⁸ notes. There is a widespread assumption that the multi-word unit *am Ende des Tages* is a loan translation from English, but this is not necessarily true. This syntagma existed in German for centuries as a realization of a temporal pattern: <*am Ende* DET N_{TEMPORAL} {*Tages/Jahres/Saison*}> [at (the) end of N_{TEMPORAL} {day/year/season}]. Of course,

⁸ The *Dictionary of Neologisms* is a component of *OWID* (see *OWID*), the corpus-based Online Lexical Information System for German, located at the IDS. Other dictionaries are, for example, the dictionary *lexiko* that describes the current use of words, and the proverb dictionary that is compiled by our UWV group.

one cannot deny the influence of the English language on its growing use as a discourse marker, as evidenced by the German web corpus *deTenTen13*.

In the following discussion, we are interested in the lexical core *am Ende* [at (the) end; ee: eventually; finally], which has its own holistic meanings and functions. Therefore, one can attribute the status as a minimal phraseological unit to the PN *am Ende* as well, a multi-word expression within a multi-word expression, if you will. The article *am* is a merged form of the preposition *an* and the definite article (German dative) *dem*. Recent corpus-based studies, e.g. the contrastive exploration of Augustin (2018), show that such merged forms in German like *am* or *im* have to be interpreted as separate prepositions.

The multi-word expressions *am Ende* und *am Ende des Tages* share a “double life” as both temporal expressions and discourse markers – with subtle differences (*am Ende des Tages* preferably temporal; *am Ende* preferably discourse marker):⁹

- (8) Temporal adverbial: ‘closing point’ (When?)
- a. *Sechs Kinder erhalten **am Ende des Tages** das bronzene Abzeichen, sieben das silberne. Vier Jungs bekommen Gold.*
[Six children receive the bronze medal **at the end of the day**, seven the silver. Four boys get the golden one.]
(DeReKo: Rhein-Zeitung, 30.08.2005)
 - b. *Leider haben wir unsere Möglichkeiten am Anfang und **am Ende** nicht genutzt. Diese Niederlage wäre nicht nötig gewesen.*
[Unfortunately, we did not use our advantages at the beginning and **at (the) end**. This defeat would not have been necessary.]
(DeReKo: Niederösterreichische Nachrichten, 06.10.2009, p. 80)
- (9) Discourse marker: ‘eventually’; ‘finally’; ‘after all’
- a. *„Moderne Brillen und Helme behindern nicht in der Bewegungsfreiheit, und **am Ende des Tages** zählt nur die Sicherheit auf der Piste“, sagt Franz Saurer, Ex-Skirennläufer, Wintersport-Experte.*
[“Modern glasses and helmets don’t obstruct the freedom of movement and **at the end of the day** only safety on the ski slope is important”, says Franz Saurer, ex-ski-racer and winter sports expert.]

⁹ The German examples are taken from the “W Archive” (about 11 billion word forms, November 2019), the largest subarchive of the *German Reference Corpus (DeReKo)* at the IDS (Leibniz-Institut für Deutsche Sprache 2019). The quantitative results are checked again using the German Web corpus in *Sketch Engine* (over 16 billion word forms, November 2019) (*deTenTen13*).

(DeReKo: Niederösterreichische Nachrichten, 07.12.2010)

- b. *Ob Nacktscanner nun moralisch sauber sind oder nicht, sei dahingestellt. **Am Ende** entscheidet eben doch das Geld.*

[Whether the body scanner (lit.: naked scanner) are morally clean or not is left undecided. **At (the) end**, money decides (talks) after all.]

(DeReKo: Mannheimer Morgen, 12.03.2010, p. 2)

Furthermore, the unit *am Ende* is the core element of constructions or so-called multi-word patterns as a special type of lexical patterns. The combination *am Ende* is – among others – the core of the following LPs:

- (10) <*am Ende* DET N_{CALENDRIAL-METRIC}>

N={*Tages/Jahres/Saison*}

[N=day/year/season]

- (11) <*am Ende* DET N_{EVENT_PROCESS}>

N={*Veranstaltung/Konzerts/Krieges*}

[N=event/concert/war]

- (12) <*am* ADJ *Ende* DET N_{LOCAL}>

ADJ={*anderen/unteren*}

[ADJ=other/lower]

N={*Gebäudes/Skala/Dorfstraße*}

[N=building/scale/village street]

The filler *Tages* in (10) has a special status insofar as it can also be a component of a phraseological unit, as previously demonstrated. All other nominal fillers like *year* or *concert* solely indicate references to TIME or – in case of *building* or *village street* – SPACE. Finally, *am Ende* is also a part of several idioms:

- (13) ***am Ende der Fahnenstange*** (*sein*)

[at (the) end of the flagpole]

‘at the end of all possible options’

- (14) *Licht **am Ende** des Tunnels*

[the light at (the) end of the tunnel]

‘signs of improvement in a bad or difficult situation that has already gone on for a long time’

- (15) **am Ende mit seinem Latein sein**
 [to be at (the) end with one's Latin]
 'to be at a loss/at end'

Examples (10)–(15) are used as separate patterns and idioms with different meanings. Nevertheless, they are connected by the node *am Ende* referring to TIME and SPACE. For instance, the idioms in (13) *am Ende der Fahnenstange (sein)* and in (15) *am Ende mit seinem Latein (sein)* have comparable pragmatic characteristics: both express emphatically that something cannot go on in the same way and therefore has to be immediately finished (temporal). An English equivalent is *the end of the road/line* ('the point where sb./sth. cannot continue'; OID: 106). The idioms *am Ende der Fahnenstange* in (13) and *Licht am Ende des Tunnels* in (14) share a common characteristic as to local aspects despite their contrary connotations: You look up and see the end point of the flagpole to which the flag can be physically raised (see 13). You can see the brightness at the end of the tunnel after a long drive at a great distance (see 14). There also exists an interrelation with the pattern <*am Ende* DET N> [at (the) end of DET N]: temporal and local.

In the cooperation *PREPCON* project, we investigated such lexical networks, overlapping concepts and fuzzy borders evoked by PNs – from a mono- as well as contrastive perspective.¹⁰

3.2 The underlying concept of time words

In the second edition of *Einführung in die Zeitlinguistik* [Introduction to time linguistics]¹¹ (2007) Vater refers – among many others – to William E. Bull, a well-known hispanist and former professor at the University of California in Los Angeles, who died in 1972. Bull combined mathematics and physics with linguistics. Regarding “what time is” Bull writes:

¹⁰ See the following selected central project publications: Steyer (ed.) (2018); Steyer (2018); Tabačeková (2019); Ďurčo and Tabačeková (eds.) (2019). With regard to PNs from a contrastive point of view see also Mellado Blanco and Steyer (2018); Ďurčo (2018); Hein et al. (2018); with regard to the didactic value for teaching and learning German as a foreign language see Iglesias Iglesias (2019).

¹¹ We cannot go into the large number of publications on TIME in language. For German see the overview in Rothstein (2017); see also Comrie (1976).

For the physicist time is the fourth dimension, an objective entity characterized and defined by three attributes: (1) linearity, (2) durational infiniteness, and (3) the capability of being divided into an infinite number of segments having an infinite variety of magnitudes. In less scientific but equally practical terms, time is what is measured by clocks, the phases of the moon, the revolutions of the earth upon its axis, and, on the verbal level, by minutes, hours, days, years, etc. In terms of human emotions and subjective perception, time is what passes as we grow older and what seems to go nowhere while we wait impatiently at a traffic light. Time is what we live in and by, what you are using up in reading this page, what is long if your head is under water and what is short toward the end of a summer's vacation.

Bull (1971: 4)

His distinction between public and personal time is perfectly applicable to our exploration of temporal PNs: TIME can be interpreted by the observation of the metric periodicity of natural phenomena ('public time') and a personal or subjective estimate of duration ('personal time').¹²

Examples of public-time words of duration are:

- morning (the interval between dawn and high sun (noon))
- afternoon (the interval between high sun and sunset)
- night (the interval between sunset and sunrise)
- day (the interval between sunrise and sunset)
- month (the interval between two new moons)
- year (the interval between two passages of the sun through the same equinox) (Bull 1971: 6)

Other examples are *week*, *decade*, *minute* or *hour*. Bull describes personal-time words as follows:

Personal-time words exhibit three functions. They may be used simply as imprecise substitutes for public-time words (*It'll be done in a moment, in a while*, and so on). They may perform the same function as well as connoting a personal judgement about the length of time involved (*The Republicans were out of office for ages*). Lastly, they may express only a personal judgement about the passage of objective time (*That week was no longer than a moment*).

Bull (1971: 6)

¹² Vater (2007) refers to a comparable distinction between 'dates' and 'pseudodates' by Rescher and Urquhart (1971). As for the concept of metric units of TIME, Fillmore (2002) defines three types: a. 'Measurepent [sic!] (measurement) units' like *millennium; year; month; day; hour* or *minute*; b. 'calendar units' like *December* or *Sunday*; c. 'calendar subunits' like *week-end, Halloween; summer or night* (34–36).

Our selection of temporal PNs is based exactly on this concept. The main criterion was the temporal meaning of the noun, and not – as in grammar – the possible temporal function of the preposition. We asked: a) To what extent is the temporal meaning incorporated in the noun? and b) Does the PN as a whole express a time concept despite the noun not being temporal?

As to a) we categorize temporal nouns/nominalizations and the respective PNs in four basic groups:

1. (chrono-)metric and calendrical nouns (times of the day, days of the week, months, bank holidays)
PNs: *gegen Abend; bis Montag; an Weihnachten* [towards evening; until Monday; at Christmas]
2. nouns that refer to natural phenomena restricted by *time*
PNs: *bei Sonnenaufgang; bis Sonnenuntergang; bei Ebbe* [on sunrise; until sunset; at low tide]¹³
3. nouns that refer to BEGINNING – DURATION – ENDING etc.
PNs: *am Anfang; ohne Unterlass; seit Kurzem* [at (the) beginning; without cease; since lately]
4. nouns that have a metric component but also express a subjective perception of time
PNs: *über Nacht; seit Generationen; in Sekundenbruchteilen* [over night; for generations; in a split second]

The PNs of the first and second group can be regarded as public-time, the PNs of the third group as private-time expressions. The fourth group is especially interesting because they can also be subject to a semantic and pragmatic change from a specific reference to a time interval to an adverbial (sometimes modal) use: *Über Nacht* [over night] does not only mean an interval from evening to morning but also ‘suddenly’. *Seit Generationen* [since generations] does not only mean an interval over several generations but also ‘very long’; ‘traditionally’; ‘resistant’. *In Sekundenbruchteilen* [in a split second] does not only mean a very short part of a second but also ‘immediately’; ‘without hesitating’; ‘extremely fast’).

An example for holistic time PNs without temporal nouns in b) is the PN *auf Knopfdruck* [on pushing/at the push of the button]. The noun *Knopfdruck* is not

13 *Sunset* (at sunset) refers to an everyday concrete event that always happens at a specific time and can be regarded as a temporal noun as well as *ebb* (at ebb), whereas *sunshine* cannot be interpreted as temporal. *Sunshine* (by sunshine) is a weather phenomenon not fixed at a specific time of the day.

a temporal lexeme but a noun–noun compound (*Knopf – Druck* [button – pressure]) that means ‘physically pressing a button’.

- (16) *Der Arbeitsplatz des Fahrers erinnert mit seinen zahlreichen Bedienelementen an ein Flugzeugcockpit. Um alle Hebel, Knöpfe und Tasten zielsicher bedienen zu können, bedarf es einer ausgiebigen Unterweisung. Wichtigste Neuerung: Der Motor wird **durch Knopfdruck** – unter dem unbeweglichen Zündschlüssel – gestartet und wieder abgestellt.*

[The driver’s workplace reminds of an airplane cockpit with its numerous controls. In order to be able to operate all levers, buttons and keys in a precise manner, extensive instruction is required. The most important innovation: the engine is started and switched off again **at the push of a button** – under the immovable ignition key.]

(DeReKo: Nordkurier, 08.10.2005)

In combination with the preposition *auf* [on] a temporal meaning is realized in almost 50% of all uses (‘immediately’; figurative partial meanings are also ‘on demand’, ‘by request’) – often embedded in comparative or negative structures:

- (17) *Eikmanns Rat: „Man darf von Weihnachten nicht zu viel erwarten!“ Sonst werde allzu leicht aus jeder Mücke ein Elefant. Von heute auf morgen auf Harmonie umzuschalten, den Alltagsfrust wie **auf Knopfdruck** einfach ausblenden – das sei nahezu unmöglich: „Fest steht: Weihnachten macht nicht automatisch besinnlich!“*

[Eikmann’s advice: “Don’t expect too much from Christmas!” Otherwise, every mosquito easily becomes an elephant. Switching from one day to the next to harmony, simply hiding everyday frustration like **at the push of a button** – that is almost impossible: “One thing is certain: Christmas does not automatically make you contemplative!”]

(DeReKo: Hamburger Morgenpost, 24.12.2005, S. 2-3)

Example (17) illustrates additional pragmatic effects in the corresponding contexts, a characteristic that can be observed for almost all temporal-modal PNs. Speakers very often use them to position themselves and to express their own attitudes towards what has been said, in our example for a rejection.

Hereinafter, we discuss the semantic and pragmatic aspects of temporal PNs using the example of the minimal temporal pattern <*für* N_{CALENDARISCH_METRISCH}> [*for* N_{CALENDARISCH_METRISCH}].¹⁴

4 A case study

4.1 The pattern <*für* N_{CALENDARISCH_METRISCH}>

The pattern <*für* N_{CALENDARISCH_METRISCH}> [*for* N_{CALENDARISCH_METRISCH}] seems to be unspectacular at first glance because of the explicit metric noun fillers:

(18) <i>für Sekunden</i>	[for seconds]
<i>für Minuten</i>	[for minutes]
<i>für Stunden</i>	[for hours]
<i>für Tage</i>	[for days]
<i>für Wochen</i>	[for weeks]
<i>für Monate</i>	[for months]
<i>für Jahre</i>	[for years]
<i>für Jahrhunderte</i>	[for centuries]
<i>für Jahrtausende</i>	[for millennia]

The pattern meaning can be paraphrased as follows: ‘Something takes a specific time interval. This interval is defined by nature and/or calendrical systems.’ Despite this common semantic core, the collocation profiles of the PNs differ significantly.

The PN *für Sekunden* [for seconds] typically correlates with verbs that express the sudden onset or minimal duration of situations such as *aufblitzen*; *stillstehen*; *einnicken*; *abgelenkt sein*; *sprachlos sein*; *Totenstille herrschen*; *den Atem anhalten* [to flash; to stand still; to doze off; to be distracted; to be speechless; there was dead silence; to hold one’s breath]. Typical contexts are sudden dangerous situations such as falling asleep at the wheel or emotional reactions such as complete silence in response to unexpected events. Basically, no durative meaning is realized here, but circumstances at a certain point in time are referenced. Many co-occurrences in the context of this PN are idiomatic or met-

¹⁴ With regard to temporal PNs see also Holzinger and Mellado Blanco (2019); Mansilla (2019).

aphorical and indicate a modal adverbial use with a sense of emotion or emphasis.

The PN *für Minuten* [for minutes] expresses an – albeit very short – time period, a brief interruption:

- (19) *Status-Quo-Gitarrist Rick Parfitt kann nach einem Herzinfarkt nicht mit auf die Abschiedstournee der britischen Kult-Band. Der 67-Jährige sei bei seinem Kollaps im Juni in der Türkei **für Minuten** „praktisch tot“ gewesen und falle mindestens bis Jahresende aus, teilte der Band-Manager Simon Porter mit.*

[Status quo guitarist Rick Parfitt is unable to join the British cult band's farewell tour after a heart attack. The 67-year-old was “practically dead” **for minutes** when he collapsed in Turkey in June and will be absent at least until the end of the year, said band manager Simon Porter.]

(DeReKo: Rhein-Zeitung, 16.09.2016, p. 28)

In most cases, *Minuten* [minutes] cannot be replaced by *Sekunden* [seconds]. The game was interrupted for minutes (*The game was interrupted for seconds). But also, the usage of *für Minuten* is characterized by the expression of subjective time or emotionality.

The PN *für Stunden* [for hours] primarily focuses on the perceived duration of concrete, mostly negative events or situations such as traffic jams or disruptions by construction sites. Typical collocation partners are *gesperrt; lahmgelegt; unterbrochen; stillstehen; von der Außenwelt abgeschnitten* [blocked; incapacitated; interrupted; to stand still; to be cut off from the outside world].

The PN *für Tage* [for days] has a wide and relatively heterogeneous spectrum of contexts, many comparable to those of the PN *für Stunden* [for hours], whereby the actual passing of several days is usually described. Trends towards a modal use can be found in this PN especially in recurrent lexical extensions such as *für Tage und Monate* [for days and months]; *für Tage oder gar Wochen* [for days or even weeks]. These syntagma emphasize the unexpectedly long duration.

The PNs *für Wochen* [for weeks] and *für Monate* [for months] have a comparable range of usage scenarios with the focus on a specific time duration with a foreseeable end. In addition, the PNs are very often used to refer to situations in which someone has been removed from everyday life for a long time, with serious consequences: *ans Krankenbett gefesselt sein; außer Gefecht gesetzt; ausfallen; von der Bildfläche verschwinden; von der Familie getrennt werden* [to be tied to the sick bed; incapacitated; to fail; to disappear from the scene; to be separated from the family]. One explanation could be that only certain condi-

tions (not all) seem to be worth explicit commentary, e.g. if they involve an unexpected extended period of time. If someone was in bed for several days because of a bout of flu (and not weeks or even months), this would not stimulate speakers to choose such a complicated construction like *für* N_{KALENDARISCH_METRISCH} *ans Krankenbett gefesselt sein*. One would simply say that someone was in bed with the flu for three days.

For the PN *für Jahre* [for years] one can expect a reference to the interval of a finite number of years. Interestingly, the concrete time reference fades in most of the *Jahre*-realizations. Typically, this PN correlates with lexemes and syntagmas that belong to the semantic field ‘in custody’: *ins Gefängnis müssen; hinter Gitter bringen; inhaftiert werden* [to have to go to prison; to put behind bars; to be imprisoned]. Even if the duration is limited (at some point the detention ends and someone is released), there is still an assessment (something is perceived as particularly long). Furthermore, one can identify many idioms in the collocation profile of *für Jahre* [for years] like *vom Tisch sein; aufs Eis gelegt; in die Schublade verbannt* [to be off the table; to put on ice; banished to/put back in the drawer; ee: *put on a back burner; kicked into the long grass*], expressing that an initially relevant matter is no longer being dealt with or pursued, and often combined with criticism.

The PNs *für Jahrzehnte; für Jahrhunderte* und *für Jahrtausende* [for decades; for centuries; for millennia] relate to circumstances with long-term consequences. The PN *für Jahrzehnte* [for decades] is used for issues and their consequences that last for a long time but with an end, e.g. *hinter Gitter bringen; verschwinden; in Vergessenheit geraten; zementieren* [to put behind bars; to disappear; to be forgotten; to cement]. *Für Jahrhunderte* und *für Jahrtausende* [for centuries and for millennia] express a far-reaching, actually unforeseeable end, both retrospectively and prospectively: *unter einer Ascheschicht begraben; unbewohnbar; unter osmanischer Herrschaft; Atommüll sicher lagern; unbewohnbar werden; verseucht sein* [buried under a layer of ash; uninhabitable; under Ottoman rule; to store nuclear waste safely; to become uninhabitable; to be contaminated].

In summary, it can be stated that in this pattern the focus on a concrete interval or duration (of primarily everyday events) is typically given in the middle of the metric range: *Tage; Wochen; Monate* [days; weeks; months], while minimal units (this applies especially to *seconds*) tend to express the immediate point in time when something happened suddenly or unexpectedly. The observation that the PN *für Sekunden* [for seconds] exhibits a preference for correlation with idiomatic expressions can be explained by the metaphorical and pragmatic potential of idioms. Speakers make use of this potential to emphasise

the surprising occurrence vividly. Idiomaticity comes into play again in the case of those PNs that deal with longer periods (*für Jahre; Jahrhunderte; Jahrtausende* [years; centuries; millennia]). Here, the concrete reference to TIME is secondary to the subjective evaluation, that something felt very long, and its consequences become the central communicative task. In particular, the PN *für Jahre* is almost always used as a modal adverb comparable with the English PN *for ages*.

The assumption that there is always a scale from time reference towards a subjective perceived duration can be confirmed from a reciprocal perspective: it is revealing that the prepositions for all of the listed (chrono-)metric nouns have very different frequencies and degrees of typicality.

While the prepositions *für* [for] and *nach* [after] occur for all nouns that we have looked at above, the preposition *binnen* [within] is only relevant for the smaller time intervals. With the nouns *Jahre* [years] and *Jahrzehnte* [decades] this preposition occurs in an exceptionally low percentage. The combinations *binnen Jahrhunderte* and *binnen Jahrtausende* [within centuries and within millennia], can be completely neglected due to the minimal frequency in the four DeReKo W-archives.

The genuine temporal prepositions *seit* [since] and *vor* [before, ee: ago] have not been used in case of *Sekunden* (**seit Sekunden; *vor Sekunden* [**since seconds; *before seconds, ee: *seconds ago*]) in our corpus: The preposition *seit* [since] appears for *Minuten* (*seit Minuten*) [since minutes] for the first time, the preposition *vor* [before ee: ago] first for *Stunden* (*vor Stunden*) [before hours ee: hours ago; **before seconds/minutes, ee: *seconds/minutes ago*]. The question to what extent these restrictions are set up in the grammar of the respective preposition and/or if they are generated through processes of usage to express perceived time cannot be examined in more detail here, and must be reserved for future studies.

4.2 Extension Patterns

As already mentioned in the first section PNs are strongly lexicalized on the one hand, but one can also observe recurrent extensions (internal and external) on the other. In the case of temporal PNs, internal extensions move the purely temporal meaning to the fore but almost always connected with a more or less pragmatic potential (see also Iglesias Iglesias and Alonso Santos 2019).

For instance, the internal gap X in *<für X Sekunden>* [for X seconds]¹⁵ is not filled in nearly 20% of occurrences (ranked first regarding the filler frequencies in the X slot). This substantiates the status of *für Sekunden* as a lexeme. At the same time, one can also find a diverse number of fillers with different characteristics. The second group of frequent fillers *<für {ein paar/einige/wenige/...} Sekunden>* [for {a few/several/some...} seconds] is located in a transition zone: they both refer to the specific point in time when something unexpected happens, as well as to a specific, very short period of time. The third group are numerical fillers such as *<für {zwei/zehn/...} Sekunden>* [for {two/ten/...} seconds] that express a duration but with vagueness and tendencies towards modal use with the meaning ‘extremely short’:

(20) *Die Oper ist das Schwierigste, **für zehn Sekunden** Musik brauchst Du einen ganzen Tag.*

[Opera is the most difficult, you need a whole day **for ten seconds** of music.]

(DeReKo: Süddeutsche Zeitung, 13.12.2012, p. 14)

*Ein Sprinter hat aber sicher einen größeren Trainingsumfang **für zehn Sekunden** Wettkampf.*

[A sprinter certainly has a larger training scope **for ten seconds** of competition.]

(DeReKo: Die Rheinpfalz, 14.04.2016)

In contrast to the PN *für Sekunden* [for seconds] and its dominant zero gap, the internal slots for *<für X {Minuten/Stunden/Tage/Wochen/Monate/Jahre}>* {for X minutes/hours/days/weeks/months/years} are primarily filled by adjectives specifying the time reference in greater detail. These X fillers mainly express concrete periods as numerals and cardinal numbers or with adjectives such as *kommenden* or *nächsten* [following or next] in combination with further extensions *<für die nächsten zwei Monate/drei Jahre>* [for orther next two months/three years].

At the other end of the scale, in the case of the nouns *centuries* and *millennia*, the internal gap is again not filled in a high percentage of cases (centuries: almost 40%, millennia: almost 50%). This observation corresponds with the assumption we mentioned in the previous section that concrete time references are first and foremost located in the middle of the metric range whereas the modal-connotative use is predominantly at its ends.

15 Our slot filler analysis was based on a slot of one to four gaps.

Other filler groups constitute new, idiomatic meanings. So, the PN *für Stunden* [for hours] is expanded by adjectives like *gemütliche*; *unterhaltsame*; *schöne*; *besinnliche*; *gesellige*; ... [cozy; entertaining; beautiful; contemplative; sociable; ...]. Here, the autonomous status has changed, the fixed unit is not the PN *für Stunden* [for hours] but the collocation pattern <ADJ *Stunden*> [ADJ hours], which is determined by one of the meanings of the German word *Stunde* [hours]: ‘nice experience; event for a certain period of time’ (ee: quality time). The preposition *für* [for] is used differently syntactically, mostly as a component of an NP (nominalization), e.g. *Danke für die unterhaltsamen Stunden* [Thanks for the entertaining hours]. This is also applicable for the internal fillers of <*für* X Tage> [for X days] in which the preposition only indicates the syntactic embedding but no longer has a semantic connection with the time nouns: e.g. (*für*) *die heißen Tage/kalten Tage* [(for) the hot days/cold days] (‘summer or winter time’); (*für*) *die drei tollen Tage* [(for) the three crazy days] (‘Fastnacht; Carnival’); (*für*) *die schönsten Wochen des Jahres* [(for) the best weeks of the year] (‘vacation; holidays’). Only the frequent realization <*für den Rest* DET Tage> [for the rest of DET days] (‘for the rest of somebody’s life’) can be regarded as a real internal extension pattern of binary PN *für Tage* [for days]. In all other cases, the autonomous status is no longer given at the PN level.

These few selected examples show that internal extensions have different functions: specifying meanings and domains of the binary core, disambiguating meanings and pragmatically adding nuances as well as constituting new meanings whereby the autonomous PN status is almost always getting lost.

In the external context of temporal PNs, we could also observe recurrent context partners and patterns, such as

- Negations like *nicht für Jahre ... so weitermachen*; *einsperren*; *schließen* [not keep doing in this manner; not lock up; not close ... for years]
- adverbs/adverb phrases like *manchmal nur/höchstens für Sekunden* [sometimes only/at most for seconds]
- functional chunks (*und das für Jahre* [and that for years])

(21) „Ansonsten landen Sie mal in einem psychiatrischen Krankenhaus – und das für Jahre“, sagte er. „Ändern Sie Ihr Leben.“

[“Otherwise, you end up in a psychiatric hospital – and that for years”, he said. “Change your life.”]

(DeReKo: Nordkurier, 23.10.2010)

- idiomatic expressions (see the context partners in the examples *für Sekunden* [for seconds] and *für Jahre* [for years] in 4.1.)
- communication verbs in a broad sense, e.g. *ankündigen*; *sagen*; *warnen*; *ausführen* [to announce; to say; to warn; to point out] (e.g. *schon vor Jahren gewarnt*) [already warned years ago] with a high pragmatic potential.

The central function is to self-position a speaker to express his view on a specific current situation, often as a reaction to questions, or as a criticism of other speakers. Temporal meaning stays in the background; more implicit, but always resonating.

5 Summary

Using a type of minimal phraseological unit, namely preposition-noun pairs, we showed that lexical units (mono-lexemes and multi-word expressions) as well as lexical patterns are entrenched and interconnected in many ways – and on different levels of abstraction. If one explores language use inductively based on mass data it becomes obvious that no item is a unique entity with clear-cut forms and meanings, even strongly fixed expressions like idioms. We illustrated the autonomous status of PNs for a deeper understanding of their use as complex lexicalized entities. The connected aspects of fuzzy borders and overlapping phenomena can be demonstrated particularly well by the case of temporal PNs. Based on Bull's distinction between public and private time we showed that not only private-time expressions like *am Ende* [at (the) end] are expressions of subjective time, but also public-time expressions that seem to be metric at first glance like *für Sekunden* [for seconds] or *für Jahre* [for years]. The reference to specific points in time also fades in the background. Instead, these units tend to have a modal adverbial use with a scale of connotative meanings. All units are highly restricted by the contexts in which they are embedded. The relevance of pragmatic indicators in the cotext of temporal PNs becomes clear if one contrasts temporal PNs with local PNs. For local PNs we could not find such recurrent patterns to a comparable extent. A hypothesis is that speakers have a much more specific need to evaluate a perceived time than a perceived space. One reason could be that TIME is less comprehensible than SPACE. TIME is a central concept in our Kulturkreis, Western culture. This perhaps leads to the situation in which speakers often feel compelled to evaluate the duration of processes. Something is perceived as too long or too short, too soon or too late, expected or unexpected, adequate or inadequate regarding time processes etc.

Finally, the results of our investigation open new perspectives on pattern-based lexicon theories and on Construction Grammar. The discussed usage-based restrictions of binary temporal PNs again provide strong support for a network perspective¹⁶. The dynamic interplay between foreground and background seems to be the central principle. Sometimes frozen lexical fragments are in the foreground and become the reference point for recurrent usage. Sometimes syntactical structures are primarily activated and filled by several lexical elements. These processes can also be changed in the self-same communicative act. All exist at once but depending on communicative needs specific aspects are much more visible and discernable than others.

References

- Augustin, Hagen. 2018. *Verschmelzung von Präposition und Artikel: Eine kontrastive Analyse zum Deutschen und Italienischen*. Berlin & Boston: Mouton de Gruyter.
- Belica, Cyril. 1995. *Statistische Kollokationsanalyse und Clustering: Korpuslinguistische Analysemethode*. Mannheim: Leibniz-Institut für Deutsche Sprache. <http://www1.ids-mannheim.de/kl/projekte/methoden/ka.html> (accessed 28 August 2019).
- Biber, Douglas. 2009. A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics* 14 (3). 275–311.
- Bull, William E. 1971. *Time, tense, and the verb: A study in theoretical and applied linguistics, with particular attention to Spanish*. Berkeley, Los Angeles & London: University of California Press.
- Burger, Harald, Dimitri Dobrovol'skij, Peter Kühn & Neal R. Norrick (eds.). 2007. *Phraseology: An international handbook of contemporary research*, 28.1.–28.2. Berlin & New York: Mouton de Gruyter.
- CII. *COSMAS II. Korpusrecherche- und -analysesystem*. Mannheim: Leibniz-Institut für Deutsche Sprache. <https://www.ids-mannheim.de/cosmas2/> (accessed 28 August 2019).
- Comrie, Bernard. 1976. *Aspect: An Introduction to the Study of Verbal Aspects and Related Problems*. Cambridge: Cambridge University Press.
- Ďurčo, Peter. 2018. Faktoren der konvergenten und divergenten Äquivalenz von präpositionalen Wortverbindungen: Deutsch – Slowakisch. In Kathrin Steyer (ed.), *Sprachliche Verfertigung: Wortverbindungen, Muster, Phrasem-Konstruktionen*, 285–306. Tübingen: Narr.
- Ďurčo, Peter & Jana Tabačková (eds.). 2019. *Präposition-Nomen-Verbindungen: Korpusstudien zu Gebrauch und Musterhaftigkeit phraseologischer Minimaleinheiten*. Berlin: Logos.
- Feilke, Helmuth. 1996. *Sprache als soziale Gestalt: Ausdruck, Prägung und die Ordnung der sprachlichen Typik*. Frankfurt am Main: Suhrkamp.

16 We have been able to demonstrate these mechanisms by using very different snippets of language: lexical fields and pattern networks of idioms, of body part expressions, of the key-word GRUND [reason], of proverbs and quotations.

- Fillmore, Charles J. 2002. Mini-grammars of some time-when expressions in English. In Joan L. Bybee & Michael Noonan (eds.), *Complex Sentences in Grammar and Discourse: Essays in honor of Sandra A. Thompson*, 31–59. Amsterdam & Philadelphia: Benjamins.
- Hanks, Patrick. 2013. *Lexical Analysis: Norms and Exploitations*. Cambridge, MA & London: The MIT Press.
- Helbig, Gerhard & Joachim Buscha. 2013. *Deutsche Grammatik: Ein Handbuch für den Ausländerunterricht*. München: Klett-Langenscheidt.
- Hein, Katrin, Peter Ďurčo, Carmen Mellado Blanco & Kathrin Steyer. 2018. *Am Anfang – na začiatku – al principio*: Eine musterbasierte Fallstudie im Sprachvergleich. In Kathrin Steyer (ed.), *Sprachliche Verfestigung: Wortverbindungen, Muster, Phrasem-Konstruktionen*, 307–339. Tübingen: Narr.
- Holzinger, Herbert J. & Carmen Mellado Blanco. 2019. Die usuellen Wortverbindungen *auf Dauer* und *auf die Dauer* und deren spanische Entsprechung *a la larga*: Eine korpusbasierte Beschreibung. In Peter Ďurčo & Jana Tabačková (eds.), *Präposition-Nomen-Verbindungen: Korpusstudien zu Gebrauch und Musterhaftigkeit phraseologischer Minimaleinheiten*, 49–84. Berlin: Logos.
- Hunston, Susan & Gill Francis. 2000. *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam & Philadelphia: John Benjamins.
- Iglesias Iglesias, Nely M. 2019. Fremdsprachlicher Erwerb von Präpositionen-Nomen-Verbindungen im Kontext (*PREPCON*): Didaktisierungsvorschläge für den Fremdsprachenunterricht. In Peter Ďurčo & Jana Tabačková (eds.), *Präposition-Nomen-Verbindungen: Korpusstudien zu Gebrauch und Musterhaftigkeit phraseologischer Minimaleinheiten*, 85–106. Berlin: Logos.
- Iglesias Iglesias, Nely M. & Alejandro Alonso Santos (2019): *Viel mehr als nur zu zweit!* Korpusgesteuerte Beschreibung des Musters [für + AD] + Stunden zu zweit] In: Irene Doval & Elsa Lamas (eds.): *Germanistik im Umbruch – Linguistik, Übersetzung und DaF*, 69–76. Berlin: Frank & Timme.
- Kiss, Tibor. 2011. Bedingungen für den Wegfall eines Artikels: Distribution und Interpretation von Präposition-Nomen-Kombinationen. In Stefan Engelberg, Anke Holler & Kristel Proost (eds.), *Sprachliches Wissen zwischen Lexikon und Grammatik*, 251–283. Berlin & Boston: de Gruyter.
- Kiss, Tibor, Antje Müller, Claudia Roch, Tobias Stadtfeld, Katharina Börner & Monika Duzy. 2016. *Ein Handbuch für die Bestimmung und Annotation von Präpositionsbedeutungen im Deutschen*, 2nd edn. Bochum: Ruhr-Universität Bochum, Sprachwissenschaftliches Institut. <https://www.linguistics.ruhr-uni-bochum.de/forschung/arbeitsberichte/14.pdf> (accessed 28 August 2019).
- Leibniz-Institut für Deutsche Sprache. 2019. *Deutsches Referenzkorpus / Archiv der Korpora geschriebener Gegenwartssprache 2019-I* (Release vom 18.03.2019). Mannheim: Leibniz-Institut für Deutsche Sprache. PID: 00-04BB-AF28-761A-2901-2. <http://www.ids-mannheim.de/DeReKo> (accessed 28 August 2019).
- lexpan. *Lexical Pattern Analyzer. Ein Analysewerkzeug zur Untersuchung syntagmatischer Strukturen auf der Basis von Korpusdaten*. Developed by the project „Usuelle Wortverbindungen“. Mannheim: Leibniz-Institut für Deutsche Sprache. <http://www.ids-mannheim.de/lexpan/> (accessed 28 August 2019).
- Mansilla, Ana. 2019. Eine korpusbasierte Fallstudie im Sprachvergleich: Die usuelle Wortverbindung *de repente* und ihre Entsprechungen im Deutschen. In Peter Ďurčo & Jana Ta-

- bačeková (eds.), *Präposition-Nomen-Verbindungen: Korpusstudien zu Gebrauch und Musterhaftigkeit phraseologischer Minimaleinheiten*, 107–123. Berlin: Logos.
- Mellado Blanco, Carmen & Kathrin Steyer. 2018. Auf der Suche nach Äquivalenz: Lexikalisch geprägte Muster kontrastiv: Deutsch – Spanisch. In Kathrin Steyer (ed.), *Sprachliche Verfestigung: Wortverbindungen, Muster, Phrasem-Konstruktionen*, 265–284. Tübingen: Narr.
- OWID. *Online-Wortschatz-Informationssystem Deutsch*. Mannheim: Leibniz-Institut für Deutsche Sprache. <http://www.owid.de> (accessed 28 August 2019).
- Oxford Idioms Dictionary for learners of English (OID)*. 2006. Oxford: Oxford University Press.
- PREPCON. *Präposition-Nomen-Verbindungen im Kontext*. Website: <http://uwv.ids-mannheim.de/prepcon/> (accessed 28 August 2019).
- PREPCON^{online}. *Präposition-Nomen-Verbindungen im Kontext. Online Plattform*. Mannheim: Leibniz-Institut für Deutsche Sprache. http://uwv.ids-mannheim.de/prepcon/prepcon_online.html (accessed 28 August 2019).
- Renouf, Antoinette & John McH. Sinclair. 1991. Collocational frameworks in English. In Karin Aijmer & Bengt Altenberg (eds.), *English Corpus Linguistics*, 128–143. London: Longman.
- Rescher, Nicholas & Alasdair Urquhart. 1971. *Temporal Logic*. Wien & New York: Springer.
- Rothstein, Björn. 2017. *Tempus*, 2nd edn. Heidelberg: Winter.
- SkE. *Sketch Engine: Corpus query system*. <http://sketchengine.co.uk/> (accessed 28 August 2019).
- Steyer, Kathrin. 2013. *Usuelle Wortverbindungen: Zentrale Muster des Sprachgebrauchs aus korpusanalytischer Sicht*. Tübingen: Narr.
- Steyer, Kathrin. 2015. Patterns: Phraseology in a state of flux. In Phraseology and Dictionaries. [Special Issue]. *International Journal of Lexicography* 28 (3). 279–298.
- Steyer, Kathrin. 2018. Lexikalisch geprägte Muster: Modell, Methoden und Formen der Online-präsentation. In Kathrin Steyer (ed.), *Sprachliche Verfestigung: Wortverbindungen, Muster, Phrasem-Konstruktionen*, 227–264. Tübingen: Narr.
- Steyer, Kathrin (ed.). 2018. *Sprachliche Verfestigung: Wortverbindungen, Muster, Phrasem-Konstruktionen*. Tübingen: Narr.
- Steyer, Kathrin. 2019. *Für Jahre vom Tisch sein*: Temporale Präposition-Nomen-Verbindungen zwischen Zeitreferenz und modal-diskursivem Gebrauch. In Peter Ďurčo & Jana Tabačková (eds.), *Präposition-Nomen-Verbindungen: Korpusstudien zu Gebrauch und Musterhaftigkeit phraseologischer Minimaleinheiten*, 9–47. Berlin: Logos.
- Steyer, Kathrin. 2020. Multi-word patterns and networks: How corpus-driven approaches have changed our description of language use. In Gloria Corpas Pastor & Jean-Pierre Colson (eds.), *Computational Phraseology*, 273–295. Amsterdam & Philadelphia: Benjamins.
- Tabáčková, Jana. 2019. *Präpositionale Wortverbindungen: Eine kontrastive Fallstudie*. Berlin: Logos.
- UWV. *Usuelle Wortverbindungen*. <http://www1.ids-mannheim.de/lexik/uwv.html> (accessed 28 August 2019).
- Vater, Heinz. 2007. *Einführung in die Zeit-Linguistik*, 4th edn. Trier: Wissenschaftlicher Verlag.

Part II: Productive patterns: bi- and multilingual

Dmitrij Dobrovol'skij, Ludmila Pöppel

Russian constructions with *nu i* in parallel corpora

Abstract: The present paper describes a family of Russian constructions on the basis of parallel corpus data. We focus on the constructional pattern *nu i X* (literally: well and X). This pattern is a phrasal template with an abundance of instances. The two basic groups of *nu i N* consist of constructions expressing surprise, bewilderment or admiration, plus these with a negative nuance. Constructions in both groups have different degrees of idiomaticity ranging from full idioms to free word combinations. Another aspect of our investigation concerns the degree of language specificity of the Russian constructions under analysis. The higher the degree of language specificity, the more approximate is any way to translate a given unit into another language. Our analysis shows that the pattern *nu i X* is a language specific unit of the Russian construction, and that the degree of language specificity is more important for bilingual lexicography, teaching of foreign languages and translation studies than the degree of phraseologicality. For these contrastive purposes it is more effective to describe such units in terms of Construction Grammar. The empirical data have been collected from the corpus query system *Sketch Engine*, subcorpora of parallel texts *OPUS2*, and the *Russian National Corpus (RNC)*, subcorpora of parallel texts.

Keywords: construction, Russian, language specificity.

1 Theoretical and conceptual framework

The theoretical aim of the present paper is to test a tenet of CxG concerning the linguistic status of phraseology: there is no strict borderline between free word combinations and phraseological units; i.e. idiomaticity, fixedness and compositionality are a matter of degree. If language is conceived of as a system of con-

Note: This paper is based on work supported by the RFFI (Russian Foundation for Basic Research) under Grant 19-012-00505.

Dmitrij Dobrovol'skij, Russian Language Institute, dobrovol'skij@gmail.com
Ludmila Pöppel, Stockholm University, ludmila.poppel@slav.su.se

<https://doi.org/10.1515/9783110520569-008>

structions, then idioms, collocations and other phraseological units are special cases of constructions and do not need specific instruments for their description. This idea goes back to Fillmore et al.:

It appears to us that the machinery needed for describing the so-called minor or peripheral constructions [...] will have to be powerful enough to be generalised to more familiar structures. [...] It can be hoped that the structure-building principles of the so-called core and the machinery for building the phraseological units [...] may be of a uniform type, the former being a generate instance of the latter.

Fillmore et al. (1988: 534)

Fillmore sees one important difference between free word combinations and phraseological units formed on the basis of a pattern, which we call *constructional phrasemes*: free word combinations are generated, while phraseological units are coined:

We can distinguish two kinds of 'creativity' in language. In one case there is the ability of speakers, using existing resources in the language, to produce and understand novel expressions. In the other case, the one for which we use the term coining, a speaker uses existing patterns in the language for creating new resources. [...]

Fillmore (1997)

According to this idea, the instances of *constructional phrasemes* are coined but not generated. By *constructional phrasemes* it is meant constructions with open slots whose meaning is generally close to their lexical meaning. The filling of open slots of syntactic structure is in principle free and subject only to certain morphosyntactic, semantic, and, in rare instances, lexical restrictions. In addition to open slots, such constructions have lexically filled positions, which also makes them similar to units traditionally regarded as belonging to phraseology (Dobrovol'skij 2011). They should be considered to be phrasal templates with constrained selection properties.

The focus of our investigation is the ability of native speakers to coin new expressions according to the same pattern. More specifically, we will describe a family of Russian constructions based on the pattern *nu i X* (literally: *well and X*). This pattern is a phrasal template with an abundance of instances. Templates of this kind realise the idea of pattern in its purest form:

[...] if we take pattern to mean a recurrent configuration containing some fixed and some variable components – which is presumably the standard sense of the term – only a phrasal template would seem to qualify.

Michaelis (2019: 196)

A particular aspect of our investigation concerns how to pose the question of the degree of language specificity of the Russian constructions under analysis. The notion of language specificity and the means for its objective analysis, including the use of corpus data, has been addressed in quite a few studies (Wierzbicka 1992, 1996; Zaliznjak, Levontina, and Šmelev 2005, 2012, Zaliznjak 2015, Šmelev 2015). Of interest to us here is how to determine the degree of language specificity on the basis of parallel corpus data. According to a view repeatedly expressed in the literature, the more different ways in which a given linguistic unit can be translated into other languages and the higher its level of scattering, the higher its degree of language specificity (Dobrovol'skij and Pöppel 2017, Sitchinava 2016, Šmelev 2015).

The empirical data have been collected from the corpus query system *Sketch Engine*, subcorpora of parallel texts *OPUS2* (Russian, English, German and Swedish), the *Russian National Corpus (RNC)*, Russian-English, English-Russian, Russian-German and German-Russian, Russian-Swedish, Swedish-Russian subcorpora of parallel texts. Both the *RNC* and *Sketch Engine* parallel corpora are sentence aligned.

The *RNC* is characterised by representative and well-balanced collections of texts. The *RNC* contains subcorpora of parallel texts in 14 languages. For the present analysis we use the following subcorpora: English-Russian (over 18 million words), Russian-English (over 10 million words), German-Russian (about 4 million words), Russian-German (about 6 million words), Swedish-Russian (about 5 million words) and Russian-Swedish (over 1,7 million words). The *OPUS2* corpus in *Sketch Engine* contains parallel subcorpora in 40 languages. They are labeled according to one of the languages included, e.g. *OPUS2 Russian*, *OPUS2 English*, *OPUS2 German*, *OPUS2 Swedish*. The subcorpora do not mark the direction of the translation. All parallels are in the same corpus. The source language is not necessarily one of the languages in a language pair, it is often a third language, for example, English. The size *OPUS2 Russian* (all parallel corpora with Russian as one of the languages) is over 300 million words. *OPUS2 English* is over 1,1 billion words; *OPUS2 German* is over 125 million words; *OPUS2 Swedish* is over 120 million words. Monolingual data for the analysis were collected from the main corpus of the *RNC* (about 300 million words). Some examples were taken from parallel corpora (the *RNC* and *Sketch Engine*) with their English translations for illustration.

2 Russian constructions *nu i X (well and X) in monolingual corpora*

The element *nu i* is fixed, while X is interchangeable. Some of the constructions are fully compositional, some are fully idiomatic. Many of them are somewhere in between. An example of a fully compositional construction is *nu i* + special question: (1) and (2).

- (1) *Ну и кто ж вам поверит, ну и какое у вас есть хоть одно доказательство?*
 ‘And who will believe you, and what single proof have you got?’
 (Ф. Достоевский. Братья Карамазовы | F. Dostoevsky. The Brothers Karamazov)
- (2) – *Ну и сколько будет делаться этот отчет? – Я не знаю.*
 ‘– So how long will that report take? – I have no idea.’
 (Sketch Engine, 31499969)

At the other pole is the idiom *nu i nu* expressing surprise as in (3).

- (3) *Бухгалтер [...] сделал такой вид, как будто и самое слово «Варьете» он слышит впервые, а сам подумал: «Ну и ну!...».*
 ‘The bookkeeper [...] pretended it was the first time he had heard even the word ‘Variety’, while thinking to himself: “Oh-oh!...”’
 (М. Булгаков. Мастер и Маргарита. | M. Bulgakov. Master and Margarita)

Between these poles is a broad intermediate zone: i.e. units that are not fully compositional but are not sufficiently irregular to be considered idioms, see (4) and (5).

- (4) – *Но мне подobaется, что этот мой комкор женится на совершенно чуждом человеку. – Ну и пусть, мне бы твои заботы, – сказала Галина Терентьевна.*
 ‘But it seems to me that this corps-commander of mine is marrying an alien and unreliable element himself. – Well, let him! – said Galina Terentyevna.
 – What strange things you worry about.’
 (В. Гроссман. Жизнь и судьба. | V. Grossman. Life and fate)

- (5) **Ну и ладно**, ну и катись к своей Катерине и кланяйся ей в ножки!
 ‘Very well then, go and do your precious Katya’s bidding, and leave me alone!’
 (Р. Погодин. Мы сказали клятву. | R. Pogodin. We swore an oath)

Data from the main corpus of the RNC were previously analysed to determine the parts of speech preferred in the *nu i* X construction (Dobrovolskij, Kopotev, and Pöppel 2019). At the first stage, we gave a general characteristic of the X filler in terms of parts of speech according to the available morphological annotation. To make a list of fixed combinations, queries of the form [*nu + i + {V, N,...}*] “well + and + {V, N,...}” were created. The syntactic relationship between the elements was not explicitly specified. However, in the vast majority of cases, parts of the trigram are syntactically related. Four of the most well-known measures – pMI, t-score, Loglikelihood, and dice (Pivovarova et al. 2017) – were used to rank trigrams by their degree of stability. The lists for subsequent qualitative analysis included only those trigrams that were the most stable for all four measures (the sum of ranks). At the next stage, lists of semantically meaningful expressions were compiled for further analysis.

It was shown that the position of the variable X can be occupied by words of different parts of speech (verbs, adjectives, adverbs, particles, exclamations, nouns) and by phrases.

Noun: *nu i durak* (what a fool), *nu i molodec* (good for him), *nu i vopros* (what a question), etc, see (6).

- (6) *Моя любовь к мужу? – Ну и вопрос! – Хороший вопрос.*
 ‘My love for my husband? – What a question! – Good question!’
 (Sketch Engine, 11070690)

The most frequent constructions with nouns in the X position are *nu i dela* (how do you like that?); *nu i gady* (the swine); *nu i frukt* (you sly-boots, you); *nu i balda* (what a dunderhead), etc, as in (7).

- (7) *И необыкновенный этот старик [...] протянул нам руки, и Смуров [...] неожиданно с ним обнялся. – Ну и дела... Вот чудной!*
 ‘And the astounding old fellow [...] proffered us both hands. Smurov [...] unexpectedly embraced him. “How do you like that? There is a queer one for you!”’
 (В. Набоков. Соглядатай. | V. Nabokov. The Eye)

Verb: *ну и стреляйте* (so shoot me); *ну и льет* (boy, it's pouring); *ну и защищайте* (go ahead and protect), etc, see (8).

(8) – *Не хочешь – ну и сиди, глупая голова! – сказал Виталик и вернулся домой.*

“‘You won’t? All right, you can stay there hungry,” said Vitalik and went home in a huff.’

(Н. Носов. Карасик. | N. Nosov. The Crucian Carp)

Adjective: *ну и глупо* (that is silly); *ну и чудно* (that’s wonderful), etc, as in (9).

(9) ***Ну и прекрасно, теперь я спокоен.***

‘Well, that’s all right then; now I’m satisfied.’

(Л. Толстой. Анна Каренина | L. Tolstoy. Anna Karenina)

Adverb, particle, interjection (invariable parts of speech): *ну и довольно* (come, that’s enough); *ну и ладно* (fine); *ну и пусть* (I don’t mind/ whatever), see (10).

(10) ***Ну и довольно, прощай, что болтать-то!***

‘Come, that’s enough. Goodbye. It’s no use talking!’

(Ф. Достоевский. Братья Карамазовы | F. Dostoevsky. The Brothers Karamazov)

Quantitative methods were used to identify different parts of speech serving as frequent fillers of the *nu i X* construction. Because they are cognitively available and mentally accessible as chunks, constructions with high-frequency fillers are formulaic units. They can represent at least two different groups that we can agree to call constructions of surprise and indifference. In the first group we find constructions such as, *nu i nu* (oh, oh), *nu i žara* (it’s hot/some heat this is), *nu i denek* (what a day/some day!), *nu i šutočki* (some sense of humor/you’ve got to be kidding), *nu i povezlo* (well, we got lucky/just our luck), etc., and in the second – *nu i ladno* (fine), *nu i pust’* (well, let him/her/whatever), *nu i chren s nim* (screw it/him, to hell with it), *nu i plevat’* (to hell with it), *nu i čert s tobój* (to hell with you/screw you), etc.

It was shown in our analysis of constructions based on parts of speech with less frequent fillers that constructions expressing surprise, admiration or bewilderment dominated in some groups, whereas in others they communicated only indifference. This will be demonstrated with the example of the two constructions *nu i N* and *nu i +* an invariable part of speech.

2.1 The constructions *nu i N*

The two basic groups of *nu i N* consist of constructions expressing surprise, bewilderment or admiration plus these with a negative nuance. One interesting property of the combination of *nu i* with nouns is that they practically never express indifference, see (11)–(15).

(11) ***Ну и гараж!***

‘Call this a garage?’

(Н. Носов. Приключения Незнайки и его друзей | N. Nosov. The Adventures of Dunno and his Friends)

(12) ***Ну и голова.***

‘He has a head.’

(Л. Толстой. Воскресение | L. Tolstoy. The Awakening)

(13) ***Ну и голос!***

‘So loud!’

(К. Vonnegut. Hocus Pocus | К. Воннегут. Фокус-покус)

(14) ***«Ну и шик!»*** заметила моя вульгарная красотка, щурясь на лепной фасад.

‘“Wow! Looks swank,” remarked my vulgar darling squinting at the stucco [...]’

(V. Nabokov. Lolita | В. Набоков. Лолита)

(15) ***Ну и жара!*** Мы на экваторе?

‘It’s hot! Are we on the equator?’

(Sketch Engine, 11178690)

Often the notion of surprise is accompanied by a strongly expressed (generally negative) judgment. Sometimes the evaluative component is so strong that it more or less crowds out the element of surprise. This is especially typical of combinations of *nu i* with nouns in which the noun is itself pejorative, as in (16)–(18).

- (16) **Ну и гады**, чуть свет – уже гавкают.
 ‘The swine – hardly light yet and they’re at it again.’
 (Н. Островский. Как закалялась сталь | Nikolai Ostrovsky. How the Steel was Tempered)
- (17) *Я тебе по совести, как перед богом... а ты, тово... Ну и дура!* Возьму вот и не повезу к Павлу Иванычу!
 ‘I tell you on my conscience, before God,...and you go and... Well, you are a fool! I have a good mind not to take you to Pavel Ivanitch!’
 (А. Чехов. Горе | A. Chekhov. Sorrow)
- (18) *Она орала на тебя, когда ты бегала за ее обедом, а сама при этом знала, что будет есть в другом месте? Ну и стерва!*
 ‘She yelled at you because you ran to get her lunch – just like she asked – and then couldn’t possibly have known that she’d already eaten somewhere else? What a bitch!’
 (L. Weisberger. The Devil Wears Prada | Л. Вайсбергер. Дьявол носит Прада)

In such cases we are dealing with fully compositional word combinations. The noun is responsible for basic semantics and evaluation, and the semantic and pragmatic function of *nu i* amounts to ensuring discourse coherence by explaining the connections between the different utterances of the dialogue.

In the construction *nu i N* the position of the variable is often occupied by diminutives (N_{dim}) and less frequently by augmentatives (N_{aug}). It has already been noted (Dobrovol'skij, Kopotev, and Pöppel 2019: 13) that the construction *nu i N_{dim}* generally signifies ‘negative surprise’; consider (19) and (20).

- (19) **Ну и денек!** Кажется, за этот день я постарела на 10 лет.
 ‘What a day. In one day I’ve grown 10 years older.’
 (Sketch Engine, 17035157)
- (20) **Ну и работка**, скажу я вам.
 ‘[...] and that was my hardest job.’
 (J. Fowles. The Collector | Дж. Фаулз. Коллекционер)

We also found several examples expressing surprise or admiration (21), sometimes with a negative nuance (22).

- (21) *Разве только какой-нибудь дяденька, обтерев губы от пивной пены, воскликнет: «Ну и пиво, восторг!»*
 ‘Except maybe some old codger exclaiming as he wipes the beer foam off his lips “Now that is what I call beer, delightful!”’
 (В. Розов. Удивление перед жизнью. RNC, main corpus)
- (22) *Вот их руки, ну и лапищи, и в то же время не лишены своеобразного изящества.*
 ‘So their hands are just big paws, yet they do not lack a certain peculiar elegance.’
 (Ю. Буйда. Город палачей. RNC, main corpus)

As the construction *nu i N_{dim}*, the construction *nu i N_{aug}* can also express a negative judgment, see (23)–(24).

- (23) *Фу, ну и духотища!*
 ‘Phew! How it smells in here’
 (C. S. Lewis. The Chronicles of Narnia. The Lion, the Witch and the Wardrobe | К. Ст. Льюис. Хроники Нарнии. Лев, колдунья и платяной шкаф)
- (24) *Ну и жарщица.*
 ‘It’s hotter’n hell on the road.’
 (J. Steinbeck. The Grapes of Wrath | Дж. Стейнбек. Гроздь гнева)

2.2 The constructions *nu i* + an invariable part of speech

Unlike the construction *nu i N*, in most cases *nu i* + an invariable part of speech expresses ‘indifference on the part of the speaker toward the evaluated circumstances of the situation’ (Dobrovolskij, Kopotev, and Pöppel 2019: 16), as in (25) and (26).

- (25) *Я с тобой в ссоре! – кричал ему вдогонку Незнайка. – Ну и пожалуйста!* – отвечал Гунька. – Сам первый придешь мириться.
 “I won’t play with you any more!” Dunno cried after him. “Don’t!” called back Gunky. “You’ll be the first to come and make it up.”
 (Н. Носов. Приключения Незнайки и его друзей | N. Nosov. The Adventures of Dunno and his Friends)

(26) *Я вам выиграла ваше пари, да? Ну и прекрасно! А до меня вам никакого дела нет.*

'I've won your bet for you, haven't I? That's enough for you. I don't matter, I suppose.'

(B. Shaw. *Pygmalion* | Б. Шоу. Пигмалион)

Our analysis of the two groups (*nu i N* and *nu i* + an invariable part of speech) revealed clear semantic preferences inherent in the units of each group: surprise for *nu i N* and indifference for *nu i* + an invariable part of speech. The exceptions are the idiom *nu i nu*, whose meaning arose out of a complete semantic reinterpretation, and the phraseme *nu i dela*, which is close to it with respect to degree of reinterpretation.

The next step is to analyse constructions that are not fully compositional.

3 English, German and Swedish equivalents: corpus analysis

Four constructions were selected for the present analysis— *nu i nu* (literally ≈ well and well); *nu i dela* (literally ≈ well and things); *nu i pust'* (literally ≈ well and let) and *nu i ladno* (literally ≈ well and fine). Semantically they constitute two distinct groups – 'surprise' (*nu nu*, *nu i dela*) and 'indifference' (*nu i pust'/puskaj*, *nu i ladno*). The semantic difference correlates with a prosodic difference: surprise constructions and indifference constructions follow different prosodic patterns.

Using parallel corpora of the *RNC* and *Sketch Engine* for English, German and Swedish equivalents in both directions, we will search for translation equivalents, as none of these constructions can be translated word for word; that is, they are language specific. The analysis of translation equivalents allows us to identify possible systematic equivalents. Within each group the analysis proceeds as follows: first, *RNC* materials from and to Russian, then data from *Sketch Engine* (which do not indicate the direction of translation; the examples are often translations from English).

3.1 *Nu i nu* in parallel corpora

Tab. 1: *Nu i nu*: English equivalents

Equivalents	Number of occurrences		
	<i>RNC: Russian-English</i>	<i>RNC: English-Russian</i>	<i>Sketch Engine</i>
<i>oh-oh/ uh-oh</i>	1	1	1
<i>well and well</i>	1		
<i>(oh) well</i>		9	6
<i>well, well</i>		9	6
<i>(oh) gee</i>		7	4
<i>well, gee/gee, well</i>		1	1
<i>my, my</i>		3	1
<i>dear me</i>		2	
<i>christ</i>		2	
<i>say</i>		2	
zero equivalent		2	7
<i>wow</i>			25
<i>(oh) boy</i>			14
<i>(oh) crikey</i>			4
<i>(oh) my God</i>		1	3
<i>holy cow</i>			3
<i>great Scott</i>		1	2
<i>oh, man</i>			4
<i>come on</i>			3
<i>what</i>			3
<i>gosh</i>		1	1
<i>well, well, well</i>		1	1
<i>what an idea</i>			2
<i>golly</i>		1	1

Besides the examples in Table 1, the following ones were found in only one of the corpora:

RNC English-Russian: *yo, yo, yo; ding-dong; ah me; cripes; well, really; phew; whys; upon my word; what a to-do; by Jove; dear, dear; Tst! Tst! Tst!; oh, rats; there; d'ye tell o't; hoity-toity; how now.*

Sketch Engine: *coo; my, my, my; isn't that amazing; way to go; holy cats; just like that; brother; oh, great; oh, dear; I say, that's rich; how is that possible; goodness; oh, my; my, oh, my; say, boy; good gracious me; oh, really; indeed; fucking hell; pussy; I'll be damned; I don't know; say; oh for fuck's sake; my god; that's so lame; all right and all right; son of a bitch; watch it; well, now; look; okay; what a rush.*

The first thing that stands out here is the large number of different English parallels to the Russian construction. In all we found 74 such equivalents. The second important feature is that of these 74, 50 occur only once, which indicates significant scattering in these English parallels.

Tab. 2: *Nu i nu*: German equivalents

Equivalents	Number of occurrences		
	<i>RNC: Russian-German</i>	<i>RNC: German-Russian</i>	<i>Sketch Engine</i>
<i>pah!</i>		1	
<i>is dös a Hetz</i>		1	
<i>dausend</i>		1	
<i>nun</i>		1	
<i>oho</i>		1	
zero equivalent			3
<i>mein Gott</i>			2
<i>na, na, na</i>			1
<i>einfach so</i>			1
<i>absoluter Wahnsinn</i>			1
<i>tut mir Leid</i>			1
<i>soso</i>			1
<i>sieh mal einer an</i>			1
<i>gut, gut</i>			1
<i>wow</i>			1

The Russian-German data are considerably smaller in scope. In *Sketch Engine* we found 9 German parallels and in the *RNC* 5. Parallels from the two corpora do not coincide. Two of them occur more than once – the zero equivalent and *mein Gott*. The results exhibit tendencies similar to those observed in the English data – i.e., significant scattering.

Tab. 3: *Nu i nu*: Swedish equivalents

Equivalents	Number of occurrences		
	<i>RNC: Russian-Swedish</i>	<i>RNC: Swedish-Russian</i>	<i>Sketch Engine</i>
<i>nej, verkligen</i>	1		
<i>nå nå nå nå</i>		1	
<i>nej men</i>		1	
<i>ja, ja</i>		1	
<i>ja jag säger då det</i>		1	
<i>har ni hört, va!</i>		1	
<i>i all sin dar</i>		1	
<i>nej, minsann</i>		1	
<i>det är besynnerligt</i>		1	
<i>se så där ja</i>		1	
zero equivalent		1	10
<i>(åh) jösses</i>			3
<i>det må jag säga</i>			1
<i>herregud</i>			1
<i>tamejfan</i>			1
<i>sådär ja</i>			1
<i>ser man på</i>			1
<i>nej nej</i>			1

We found 18 Swedish parallels, the most frequent of which is the zero equivalent. This testifies to considerable scattering and to a certain non-translatibility of *nu i nu*.

3.2 *Nu i dela* in parallel corpora

Tab. 4: *Nu i dela*: English equivalents

Equivalents	Number of occurrences		
	<i>RNC: Russian-English</i>	<i>RNC: English-Russian</i>	<i>Sketch Engine</i>
<i>how do you like that</i>	1		
<i>well</i>		1	
<i>great Scott</i>		1	

Equivalents	Number of occurrences		
	<i>RNC: Russian-English</i>	<i>RNC: English-Russian</i>	<i>Sketch Engine</i>
<i>that is strange</i>		1	
<i>It's a fair do</i>		1	
<i>dear me</i>		1	
<i>well</i>		1	
<i>nasty business</i>			1
<i>I'll be damned</i>			1
<i>gee</i>			4
<i>whoa</i>			1
zero equivalent			2
<i>look here</i>			1
<i>how interesting</i>			1
<i>what do you know</i>			1
<i>I cannot believe this shit</i>			1
<i>blimey</i>			1
<i>oh, wow</i>			1
<i>damn it</i>			1
<i>holy shit</i>			1
<i>oh, my</i>			1

We found 22 English parallels, and only two equivalents occur more than once – *gee* and the zero equivalent.

Tab. 5: *Nu i dela*: German equivalents

Equivalents	Number of occurrences		
	<i>RNC: Russian-German</i>	<i>RNC: German-Russian</i>	<i>Sketch Engine</i>
<i>das ist ja reizend</i>		1	
<i>Seht mal!</i>			1
zero equivalent			2
<i>(du liebe) Scheiße</i>			2
<i>okay</i>			1

The Russian-German data are very small in scope. We found only 1 equivalent in the *RNC* and 4 in *Sketch Engine*. Two of them occur twice – *Scheiße* and the zero equivalent.

Tab. 6: *Nu i dela*: Swedish equivalents

Equivalents	Number of occurrences		
	<i>RNC: Russian-Swedish</i>	<i>RNC: Swedish-Russian</i>	<i>Sketch Engine</i>
<i>jösses</i>			1
<i>herregud</i>			1
<i>titta här</i>			1
zero equivalent			4
<i>okej</i>			1
<i>det här är allt en riktig soppa</i>			1
<i>det här var en värre historia</i>		1	
<i>det här blir just månljust</i>		1	
<i>det var det värsta</i>		1	
<i>men vad i all sin dar</i>			
<i>aldrig har jag varit med om slikt</i>		1	

Among 11 Swedish parallels found in both corpora only the zero equivalent occurs more than once (4 occurrences), which also indicates considerable scattering.

3.3 *Nu i ladno* in parallel corpora

Tab. 7: *Nu i ladno*: English equivalents

Equivalents	Number of occurrences		
	<i>RNC: Russian-English</i>	<i>RNC: English-Russian</i>	<i>Sketch Engine</i>
<i>(that's) fine</i>	1	1	15
<i>(it's/that's) okay/o.k.</i>		1	10
<i>Whatever</i>		1	9
<i>(oh) well</i>			5
<i>(it's/that's) all right</i>	1	2	5
<i>(so) all right</i>		1	3

Equivalents	Number of occurrences		
	<i>RNC: Russian-English</i>	<i>RNC: English-Russian</i>	<i>Sketch Engine</i>
zero equivalent		1	3
<i>okay then</i>			2
<i>ok. fine/fine, okay</i>			2
<i>never mind</i>		1	2
<i>that's cool</i>			1
<i>I'm through</i>			1
<i>doesn't matter</i>			1
<i>what matter</i>			1
<i>tough</i>			1
<i>forget it</i>			1
<i>(very) well then</i>	1	1	
<i>let him/her V</i>	1	1	
<i>enough</i>	1		
<i>I don't care</i>		2	
<i>anyway</i>		3	
<i>so what</i>	1		

We found a total of 22 parallels, 8 of which occur once in one corpus. The most frequent equivalents are the zero equivalent (17 occurrences), followed by (*it's/that's*) *okay/o.k.* (11 occurrences) and *whatever* (10 occurrences).

Tab. 8: *Nu i ladno*: German equivalents

Equivalents	Number of occurrences		
	<i>RNC: Russian-German</i>	<i>RNC: German-Russian</i>	<i>Sketch Engine</i>
<i>(na/ist schon) gut</i>			3
<i>fein</i>			2
zero equivalent			2
<i>okay</i>			2
<i>nun denn</i>			1
<i>alles klar</i>			1
<i>nun gut</i>		1	

We found 7 parallels, 3 of which occur once.

Tab. 9: *Nu i ladno*: Swedish equivalents

Equivalents	Number of occurrences		
	<i>RNC: Russian-Swedish</i>	<i>RNC: Swedish-Russian</i>	<i>Sketch Engine</i>
zero equivalent		7	8
<i>(det) gör inget/ ingenting</i>		2	3
<i>(helt) okej</i>		2	2
<i>varsågod</i>			1
<i>jaha</i>			1
<i>toppen</i>			1
<i>nåja</i>		1	
<i>verkligen</i>		1	
<i>tja</i>		2	
<i>nå, då så</i>	1		

In the Swedish corpora we found 10 parallels, 6 of which occur once in one corpus. The zero equivalent leads by a large margin (15 occurrences).

3.4 *Nu i pust'* in parallel corpora

Tab. 10: *Nu i pust'*: English equivalents

Equivalents	Number of occurrences		
	<i>RNC: Russian-English</i>	<i>RNC: English-Russian</i>	<i>Sketch Engine</i>
<i>let him/them/it P</i>		5	12
<i>(it) doesn't matter</i>		1	2
<i>I don't/she didn't care</i>		2	2
zero equivalent		9	2
<i>(ah) well</i>		2	
<i>well, then</i>		1	
<i>whatever</i>			1
<i>so be it</i>		1	1
<i>that's all right</i>			1
<i>so what</i>			1
<i>fine</i>			1
<i>sure</i>	1	1	

Equivalents	Number of occurrences		
	<i>RNC: Russian-English</i>	<i>RNC: English-Russian</i>	<i>Sketch Engine</i>
<i>good</i>	1		1
<i>all right</i>	1		1
<i>it will be all one to me</i>	1		1

A total of 15 equivalents were found, 9 of which occur once in one of the corpora. Most often the construction *nu i pust'* is translated with the similar English construction *let him/them/it P*. The zero equivalent is the second most frequent equivalent.

Tab. 11: *Nu i pust'*: German equivalents

Equivalents	Number of occurrences		
	<i>RNC: Russian-German</i>	<i>RNC: German-Russian</i>	<i>Sketch Engine</i>
<i>(nun) gut</i>	1		1
<i>lass/lasst sie/ihn P</i>	3		1
<i>eben</i>			1
zero equivalent	2		1
<i>sollen sie</i>	4		
<i>na, in Gottes Namen</i>	1		
<i>wenn schon</i>		1	
<i>von mir aus</i>		2	
<i>dann mag er</i>		1	
<i>das tut nichts</i>		1	

Of 10 German equivalents 5 occur once. The zero equivalent and the construction *lass/lasst sie/ihn* occur 4 times each.

Tab. 12: *Nu i pust'*: Swedish equivalents

Equivalents	Number of occurrences		
	RNC: Russian-Swedish	RNC: Swedish-Russian	Sketch Engine
<i>låt honom/ henne/ dem P</i>	2	3	3
<i>bra</i>			1
<i>än sen då</i>			1
<i>eller hur</i>			1
<i>det gör inget</i>		1	1
zero equivalent		1	1
<i>det får N gärna P</i>		2	
<i>må de göra det</i>		1	
<i>det gjorde ingenting</i>		3	
<i>det betyder ingenting</i>		1	
zero equivalent		3	
<i>spelar ingen roll</i>		1	
<i>får det vara</i>		1	

A total of 13 equivalents were found, 7 of which occur once. The construction *låt honom/ henne/ dem P* is the most frequent (8 occurrences), followed by *det gjorde ingenting* (3 occurrences) and the zero equivalent (3 occurrences).

The English, German and Swedish equivalents of constructions expressing indifference all translate them with a similar construction in the corresponding target language - *let N + P*; *lass/lasst N + P* and *låt N + P*. These equivalents are the most frequent for all of the languages we investigated. Consider (27) and (28).

(27) *Англичане могат върнута̀ся. Ну и пусть.*

‘The English might come back. Let them.’

‘Die Engländer kommen vielleicht zurück. Lasst sie.’

(Sketch Engine, 12298189)

(28) *У него положение. – И у него ваш сын. – Ну и пусть.*

‘He’s got the position. – And he’s got your son. – Let him have him then.’

‘Han har makten. – Och han har din son. Låt honom ha honom då.’

(Sketch Engine, 38291059)

The English, German and Swedish equivalents of constructions expressing indifference all translate them with a similar construction in the corresponding target language - *let* N + P; *lass/lasst* N + P and *låt* N + P. These equivalents are the most frequent for all of the languages we investigated; see (27) and (28).

The semantic basis of the constructions *let* N + P; *lass/lasst* N + P and *låt* N + P is an appeal to the interlocutor not to change anything in the current situation, to leave everything as is. These constructions are explicitly addressed to communication partners. If the situation is fraught with some negative elements and the speaker is expressing a negative attitude toward these circumstances, the sense of the constructions *let* N + P; *lass/lasst* N + P and *låt* N + P is a recommendation to ignore the situation. The pragmatic result is a characterization of the situation (particularly its negative aspects) as something insignificant that is not worthy of attention and active involvement. The idea of indifference is thereby incorporated in these constructions on the implication level. The Russian constructions *nu i ladno* and *nu i pust'* differ from these English, German and Swedish near-equivalents in that they lack explicit dialogicity; that is, they are not explicitly addressed to an interlocutor. When speakers utter *nu i ladno* or *nu i pust'*, they are stating the triviality or insignificance of a situation. Purely semantically, therefore, these constructions express the speaker's view of the situation itself, whereas *let* N + P; *lass/lasst* N + P and *låt* N + P are directed at the potential reaction of someone else to this situation.

The analysis has enabled us to determine the frequency of translation equivalents. For constructions expressing surprise, it is above all the zero equivalent that occurs in the English, German and Swedish parallels; see (29), (30) and (31).

(29) *Взгляни- ка. Ну и ну! И на кого ж ты ставил?*

'Take a look at this. Who were you betting on?'

(Sketch Engine, 28293226)

(30) *Ну и ну, так много пушек в городе, и так мало мозгов.*

'So viele Pistolen in der Stadt und so wenig Verstand.'

(Sketch Engine, 23189)

(31) *Ну и дела. Как нам с этим конкурировать?*

'Det här är allt en riktig soppa! Hur kan vi tävla mot det där?'

(Sketch Engine, 22648167)

The most frequent equivalents in the English materials are *wow*; *boy*; *well*; *well*, *well*; and *gee*. This applies above all to the idiom *nu i nu*. The most frequent parallels found in *Sketch Engine* – *wow* and *boy* – do not occur in the *RNC*; *well*; *well*, *well*; and *gee* occur in both corpora. These divergences are quite natural. *Sketch Engine* is much larger than the *RNC*, while the *RNC* is much cleaner. In addition, the texts in these corpora differ with respect to genre. The *RNC* contains almost exclusively fictional texts, whereas non-fiction dominates in *Sketch Engine*. The German and Swedish corpora are too small to allow us to identify high-frequency equivalents.

What most convincingly argues that *nu nu*, *nu i dela*, *nu i ladno* and *nu i pust'* are language-specific with respect to English, German and Swedish is a significant scattering of equivalents and the partial absence of a translation equivalent in the parallel texts.

4 Conclusions

We have analysed two groups of constructions based on the phrasal template *nu i X*. The first group denotes 'surprise', the second, 'indifference'. Within each group, constructions have different degrees of idiomaticity ranging from full idioms to free word combinations. That is, we are dealing with a gradual transition from free word combinations to phraseology. A question that arises in this connection concerns the efficacy of this sort of description. From the perspective of language comparison, translation, bilingual lexicography, language teaching, etc., what is crucial is not the degree of phraseologicity, but the degree of language specificity. This is determined not by contrasting idioms vs. non-idioms, but by any deviation from full compositionality; i.e., by the predictability of meanings. For such contrastive purposes it is more effective to describe such units in terms of Construction Grammar.

The higher the degree of language specificity, the more approximate is any way to translate a given unit into another language. This problem is significant both in theory and in practice. From a practical point of view, it is directly related to bilingual lexicography (especially if it is not a traditional dictionary, but a construction), as well as to the teaching of foreign languages and translation studies. From a theoretical point of view, an in-depth study of the phenomenon of language specificity will allow us to develop an empirical basis for a discussion about the validity of the linguistic relativity hypothesis, which has become extremely relevant again in recent years.

References

- Dobrovol'skij, Dmitrij. 2011. Phraseologie und Konstruktionsgrammatik. In Alexander Lasch & Alexander Ziem (eds.), *Konstruktionsgrammatik III. Aktuelle Fragen und Lösungsansätze*, 111–130. Tübingen: Stauffenburg.
- Dobrovol'skij, Dmitrij, Mihail Kopotev & Ludmila Pöppel. 2019. Grupa konstrukcij *nu i X*: semantika, pragmatika, sočetaemost' [The family of constructions *nu i X*: semantics, pragmatics, combinatorics]. *Scando-Slavica* 65 (1). 5–25.
- Dobrovol'skij, Dmitrij & Ludmila Pöppel. 2017. Constructions in parallel corpora: a quantitative approach. In Ruslan Mitkov (ed.), *Computational and corpus-based phraseology. Euro-phras 2017*, 41–53. London: Springer International Publishing.
- Fillmore, Charles J., Paul Kay & Mary O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language* 64 (3). 501–538.
- Fillmore, Charles. 1997. Idiomaticity. In Charles J. Fillmore and Paul Kay, *Berkeley Construction Grammar*. <http://www.icsi.berkeley.edu/~kay/bcg/lec02.html>. (accessed 10 December 2019).
- Michaelis, Laura A. 2019. Constructions are patterns and so are fixed expressions. In Beatrix Busse & Ruth Moehlig-Falke (eds.), *Patterns in language and linguistics*, 193–220. Berlin: Mouton de Gruyter.
- Pivovarova, Lidia, Daria Kormacheva & Mikhail Kopotev. 2017. Evaluation of collocation extraction methods for the Russian language. In Mikhail Kopotev, Olga Lyashevskaya & Arto Mustajoki (eds.), *Quantitative approaches to the Russian language*, 37–157. London & New York: Routledge.
- RNC. *Russian National Corpus*. <https://ruscorpora.ru/old/en/index.html> (accessed 29 August 2019)
- Sitchinava, Dmitrij. 2016. Parallel corpora as a source of defining language-specific lexical items. In Tinatin Margalitatdze & George Meladze (eds.), *Proceedings of the XVII EURALEX international congress: Lexicography and linguistic diversity*, 394–401. Tbilisi: Ivane Javakhishvili Tbilisi University Press.
- SKÉ. *Sketch Engine: Corpus query system*. <http://sketchengine.co.uk/> (accessed 28 August 2019).
- Šmelev, Aleksej. 2015. Russkie lingvospecifičeskie leksičeskie edinicy v paralel'nyh korpusach: vozmožnosti issledovanija i "podvodnye kamni" [Russian language-specific lexical units in parallel corpora: prospects of investigation and "pitfalls"]. *Computational linguistics and intellectual technologies: Proceedings of the international conference "Dialogue 2015"* [Komp'juternaja lingvistika i intellektual'nye tehnologii: Trudy meždunarodnoj konferencii "Dialog 2015"] 14 (21). 584–595. Moscow: Izdatel'stvo RGGU.
- Wierzbicka, Anna. 1992. *Semantics, culture, and cognition. Universal human concepts in culture-specific configurations*. New York & Oxford: Oxford University Press.
- Wierzbicka, Anna. 1996. *Semantics: Primes and universals*. Oxford: Oxford University Press.
- Zaliznjak, Anna A. 2015. Lingvospecifičnyje edinicy russkogo jazyka v svete kontrastivnogo korpusnogo analiza [Russian language-specific words as an object of contrastive corpus analysis]. *Computational linguistics and intellectual technologies: Proceedings of the international conference "Dialogue 2015"* [Komp'juternaja lingvistika i intellektual'nye tehnologii: Trudy meždunarodnoj konferencii "Dialog 2015"] 14 (21). 683–695. Moscow: Izdatel'stvo RGGU.

- Zalznjak, Anna A., Irina Levontina & Aleksej Šmelev. 2005. *Ključevye idei russskoj jazykovoj kartiny mira* [Key ideas of the Russian language picture of the world]. Moscow: Jazyki slavjanskoj kul'tury.
- Zalznjak, Anna A., Irina Levontina & Aleksej Šmelev. 2012. *Konstanty i peremennye russskoj jazykovoj kartiny mira* [Constants and variables of the Russian language picture of the world]. Moscow: Jazyki slavjanskoj kul'tury.

Laura Giacomini

Phraseology in technical texts

A frame-based approach to multiword term analysis and extraction

Abstract: In much the same way as in studies on general language, phraseology is becoming central to research focused on specialised discourse, especially in the wake of corpus studies concerned with the extraction of terminologically relevant word combinations such as specialised collocations, and with their lexicographic presentation (cf. Hanks 2010). This new, strong orientation is perfectly compatible with traditional concept-oriented terminological theories. At the same time, it complies with linguistic theories postulating the phraseological nature of language (cf., among others, Sinclair 2004 and Stubbs 2001). This contribution deals with the way in which multiword terms, intended as phraseological units of terminology, can be classified from a morphological and variational perspective (Section 1). We illustrate how a frame-based approach to terminology can be applied to analyse terms and variants (Section 2), and how this lays the ground for the compilation of a tagset for multiword term annotation and extraction from a specialised corpus (Section 3). Notwithstanding the focus of this contribution on the technical domain, validation tasks mentioned in Section 3.3 hint at the cross-domain feasibility of the proposed method. Section 4 finally draws general conclusions about the advantages provided by the use of frames in terminology, especially for detecting phraseological units, and presents ongoing and future work aimed at further exploring this important area of research.

Keywords: specialised phraseology, frame-based terminology, term variation, multiword term extraction, technical language.

Laura Giacomini, University of Hildesheim, Institute for Information Science and Natural Language Processing, laura.giacomini@iued.uni-heidelberg.de

<https://doi.org/10.1515/9783110520569-009>

1 Multiword terms in technical texts

1.1 Technical language(s), variation, and phraseology

The technical domain is a vast and heterogeneous field of knowledge which is constantly exposed to rapid changes affecting people, processes, and products. Related to these intrinsic characteristics is the fact that many technical sub-fields have a relatively low level of standardisation. This is particularly true of technical or technological domains with a more recent history, or even of those which interface with non-technical disciplines. While, for instance, the electro-technical domain has a highly standardised vocabulary, the domain of thermal insulation generally shows limited terminological standardisation despite its interdisciplinary nature (it is closely related to several specialised fields such as chemistry, physics, materials sciences, construction engineering, and environmental studies). This inevitably leads to the proliferation of terminological variants, especially for what concerns multiword terms.

Most relevant literature on term variation, ranging from studies on terminology to investigations into corpus and computational linguistics, has dealt with the description of diasystemic variants, i.e. with synonymous terms belonging to parallel, mutually exclusive language varieties. The most typical example of this involves geographical variants of the kind *Mineralwolle* (Germany/Austria) vs. *Tellwolle* (Austria) in the domain of thermal insulation, or *reembolso* (Spain) vs. *rescate* (Peru) in the domain of finance and banking.¹ Another example would be variation promoted by corporate language, with different terms employed by different companies to refer to one and the same product type, for instance *Panorama-Glasdach* (BMW) vs. *Panoramadach* (VW) in the automobile industry.

However, the focus of the present study lies in a complementary, equally relevant, and yet quite underestimated phenomenon, namely synonymous variation within the same language variety, which can often be found even within the same text for a range of reasons (Freixa 2006, 2013). So, for instance, in German we encounter *technische Hydromechanik* as a synonym of *Hydraulik*, and *PUR-Platte* as a synonym of *Platte aus PUR*, in both cases with no diasystemic distinction. Synonymy is considered here in a broad sense, also including the case of merely contextual synonyms.

¹ The latter example is taken from the bilingual comparisons in Caro Cedillo (2004).

We are convinced that, from a translational and terminographic point of view, this phenomenon is even more important for terminology than the diasystemic one, since it is exactly the co-existence of synonymous terms within the same language variety which poses significant problems in translation and in the compilation of terminographic resources like, for instance, termbases. Moreover, this kind of non-diasystemic variation comparatively affects multiword terms more than single-word terms, which explains its direct connection with the phraseological component of terminology (cf. Gläser 1994, Hanks 2010).

These observations seem to largely to apply not only to the technical language(s), but also to many, if not all other specialised languages, as further examples in this contribution will show.

1.2 Aim of the study

The main objective of the study was to implement a method for a frame-based analysis and extraction of multiword terms and variants from a corpus of technical texts. At the same time, the study aimed at designing a model for data representation in a terminology database and for data presentation in a lexicographic resource supporting text production. In general, these tasks were performed by considering a multilingual perspective, combining results concerning the actual working language, German, with observations on data in English and Italian. A multi-domain perspective was also taken into account: the overall model specifically developed for the technical domain under investigation, namely the domain of thermal insulation products, was later validated through application to further technical and non-technical fields of knowledge (Giacomini 2019b; cf. also Section 3.3).

1.3 Multiword term types

We will concentrate on multiword terms (MWT) as a subtype of phraseological expressions. Burger (2015) describes, among special classes of phrasemes, *phraseologische Termini* (*phraseological terms*) such as *in Konkurs gehen*. In this study we extend this class to include other syntactic patterns, for instance A+N and N+PP, as well as longer word combinations. As explained in Giacomini (2019b), the distinction between collocations and multiword terms sometimes found in the literature on terminology has not been employed, in favour of a

more lexicographic-friendly classification. At the same time, compounds with a specialised meaning have also been included as a subset of multiword terms. This choice is due to the multilingual, also translational approach we adopt towards terminology, which constantly highlights equivalences between compounds and word combinations across different languages. Table 1 shows some examples of multiword terms in German, Spanish and English belonging to the two classes *complex terms* and *specialised phrasemes* and to different domains.

Tab. 1: Examples for complex terms and specialised phrasemes as MWT classes in different domains and different languages

MWT	Examples
Complex terms	<i>Luftkühlung</i> (DE technical language) <i>fibroepitelial</i> (ES medical language) <i>low-density</i> (EN chemical language)
Specialised phrasemes	<i>integrierter Schaltkreis</i> (DE technical language) <i>células alveolares</i> (ES medical language) <i>insulating layer</i> (EN technical language)

1.4 Multiword term variant types

For the purpose of analysing multiword terms and their variants, we introduced a variation typology consisting of three main types, namely morphological, syntactic and orthographical, which are displayed in Table 2 together with the corresponding variation criteria and examples related to the technical domain of thermal insulation.

Tab. 2: Types of multiword term variants with corresponding variation criteria and examples related to thermal insulation.

Variation type	Variation criteria	Examples
MV Morphological variation (partial / total)	Changes in lexical morphemes	<i>Dämmplatte - Isolierplatte</i>
SV Syntactic variation	Changes in the part of speech, order of compound elements, syntagmatic structure	<i>dämmen – Dämmung</i> <i>Glasschaum – Schaumglas</i> <i>Baumaßnahme – bauliche Maß-</i>

- Data modelling (specification of frames and frame elements, as well as of the relations between domain ontology and frame; syntactic and semantic analysis of variant clusters by using frame elements),
- Frame-based corpus annotation, and
- Frame-based MWT and variant extraction.

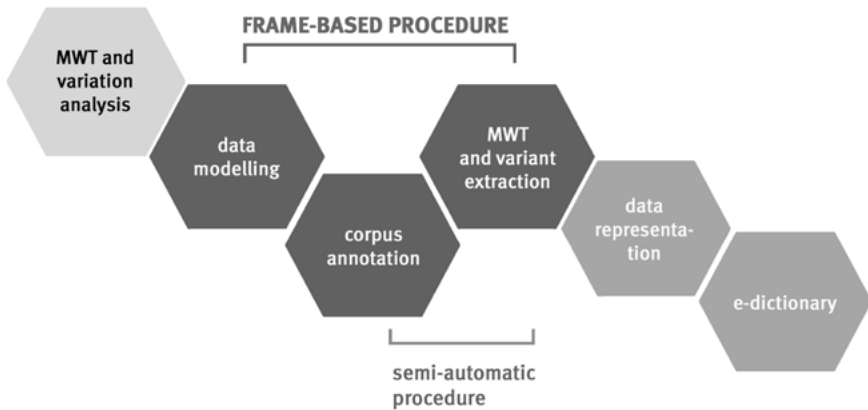


Fig. 1: Frame-based modules (in dark grey) as part of the general model structure, with indication of the semi-automatic part of the procedure.

The model structure is finally completed by data representation in a technical termbase and data presentation in a technical e-dictionary.

The next two chapters will focus on the description of the core frame-based modules.

2 A frame-based approach to multiword term analysis

The present study takes Frame-Based Terminology (Faber 2015, 2012) as a starting point and adapts it to the needs of term variation description. Frames are understood as cognitive models which cover particular word meanings and argument structures. For what concerns technical artefacts such as thermal insulation products, a set of frames can be identified which are particularly

useful for covering the entire life cycle of a product.² In this contribution, the example of the frame FUNCTIONALITY will be discussed. Frame elements in the sense of Frame Semantics (Fillmore 1985, 2008) have been identified for the frame FUNCTIONALITY, e.g. MATERIAL, MATERIAL CLASS, DELIVERY FORM, TARGET, PACKAGING, TOOL, or APPLICATION TECHNIQUE. Frame elements are understood as the necessary semantic components of a given frame. As such, they are typically found in technical texts dealing with the functionality of thermal insulation products, in which they can appear in different constellations depending on the specific topic of the text. The following two texts exemplify possible distributions of frame elements:

- a) Liegt die Folie(PACKAGING), bringt(+ auf)(APPLICATION TECHNIQUE) man die Dämmung(PRODUCT) auf. Das können **Dämm(GOAL)stoffe(MATERIAL) in Form(FORM) von Matten(FORM) oder Platten(FORM)** sein. Für Fugendichte sorgt man, indem man die Dämm(GOAL)matten(FORM) oder -platten(FORM) in zwei Lagen übereinander aufschichtet(APPLICATION TECHNIQUE), wobei man die Fugen versetzt. Wärmebrücken(PROPERTY) werden so vermieden. Auch Schütt(APPLICATION TECHNIQUE)dämm(GOAL)stoffe(MATERIAL) sind als Dämm(GOAL)material(MATERIAL) zur Dachboden(TARGET)dämmung(GOAL) einsetzbar(PROJECT). Als Dämm(GOAL)materialien(MATERIAL) eignen sich Dämm(GOAL)stoffe(MATERIAL) wie Steinwolle(MATERIAL) oder organische(MATERIAL ORIGIN) Fasern(FORM).
(www.energie-experten.org)
- b) Mineralische(MATERIAL ORIGIN) Bau(PROJECT)stoffe(MATERIAL) bestehen aus anorganischen(MATERIAL ORIGIN) Stoffen(MATERIAL). Auch hier können synthetische(MATERIAL ORIGIN) Stoffe(MATERIAL) enthalten sein. Dämm(GOAL)stoffe(MATERIAL) aus Mineralien(MATERIAL ORIGIN) bieten einen hohen Brandschutz(PROPERTY), unterstützen bei der Feuchtigkeits(PROPERTY)regulierung (PROPERTY) im Haus(SYSTEM) und liefern guten Wärmeschutz(PROPERTY). Zu den mineralischen(MATERIAL ORIGIN) Dämm(GOAL)stoffen(MATERIAL) zählen Bläh(MATERIAL PRODUCTION TECHNIQUE)ton(MATERIAL), Glas(MATERIAL)wolle(MATERIAL), Mineral(MATERIAL)schaum(MATERIAL CLASS), Kalzium(MATERIAL)silikat(MATERIAL), Perlit(MATERIAL), Schaum(MATERIAL CLASS)glas(MATERIAL) und Steinwolle(MATERIAL).
(www.obi.de)

² For a detailed discussion about the notion of *technical artefact*, see Giacomini (2019a). For frames as means of representation for the stages in the life cycle of a product, see Giacomini (2019b).

2.1 Syntactic and semantic analysis of multiword terms

As previously stated, synonymous variation implies the existence of variant clusters in which semantic similarity can be traced by means of frame elements. Clusters can be analysed by annotating each multiword term with syntactic and semantic tags. Syntactic tags correspond to the part of speech of a multiword term component, whereas semantic tags correspond to frame elements. Table 3 shows an example of the frame-based analysis of a variant cluster including six multiword terms.

Tab. 3: Analysis of multiword term variants from a syntactic (PoS) and frame-based semantic perspective. Vertical slashes are used to separate the components of a compound

MWT variants	Syntactic and semantic analysis
<i>Platte aus extrudiertem Polystyrol</i>	N _{FORM} aus V _{MAT_TECH} N _{MAT}
<i>Dämmplatte aus extrudiertem Polystyrol</i>	V _{GOAL} N _{FORM} aus V _{MAT_TECH} N _{MAT}
<i>Polystyrol-Extruderschäum-Dämmplatte</i>	N _{MAT} -V _{MAT_TECH} N _{MAT_CLASS} -V _{GOAL} N _{FORM}
<i>Polystyrol-Dämmstoffplatte aus Extruderschäumstoff</i>	N _{MAT} -V _{GOAL} N _{MAT} N _{FORM} aus V _{MAT_TECH} N _{MAT_CLASS}
<i>extrudierte Polystyrol-Hartschaumplatte</i>	V _{MAT_TECH} N _{MAT} -N _{MAT_CLASS} N _{FORM}
<i>XPS-Platte</i>	(V _{MAT_TECH} N _{MAT} -N _{MAT_CLASS})-N _{FORM}

A few observations can be made concerning the results of this analysis.

1. The relation between frame elements and parts of speech is mostly unambiguous, as shown by the multiword terms in Table 3. This has to do with the way in which specific classes of entities (Lyons 1977) correspond semantically to certain word classes. For example, MATERIAL TECHNIQUE, indicating the technique by which a thermal insulation material is produced, can be expected to be associated with verbal forms, as in *extrudiert* or *Extruder-*. On the contrary, FORM is expected to be associated with nominal forms. For this reason, we decided to focus on parts of speech rather than on syntactic patterns (e.g. *Platte aus extrudiertem Polystyrol* = N PP), since syntactic patterns prove to be very variable in texts and display no direct relation to frame elements.
2. Compounds can usually be mapped onto more than one frame element since each frame element is typically linked to single words (stems) rather than to stem combinations.

3. Some frame elements might not be lexicalised from one or the other variant, but they would still be cognitively relevant and always implied by the context. This is, for instance, the case of GOAL in the comparison between the first and the second term in Table 3.
4. Interestingly, abbreviated forms of terms, especially acronyms and initialisms such as *XPS*³, can also be decomposed in semantic constituents and analysed by means of frame elements. They are, to all effects, variants of their extended forms.

This procedure has two objectives. On the one hand, it supports validation of the available set of frame elements, possibly helping detect missing elements or even superfluous ones. On the other hand, this kind of syntactic and semantic notation will later play a role in the microstructure of the technical dictionary.

2.2 Frames and domain ontology

The designed model is not only based on the description of relevant frames but also on the relation between frames and a domain ontology. These two model components serve the purpose of substantiating the link between concepts and term meanings. If frames are necessary to pinpoint the terminological meaning of multiword terms and their variants, they can only be identified and verified against an existing ontology, the aim of which is to collect, classify, and put in relation with each other all concepts of a given specialised domain. A detailed description of the ontology of thermal insulation, including the building and representation methods employed (cf. Giacomini 2019b), is beyond the scope of this contribution. However, it is important to note that the connections between concepts and frame elements can take up various configurations, depending on the term(s) they are linked to. The way in which domain ontology and frame model interface has immediate consequences on frame-based extraction of multiword terms from a corpus (Section 3).

³ For the sake of simplicity, *XPS* is here considered to be an initialism, even though *X* is not the initial letter of the term *extruded*.

3 A frame-based approach to multiword term extraction

Multiword term extraction was performed on a 5-million-word corpus of German technical and semi-technical texts dealing with thermal insulation products. The *Insulation Corpus* was specifically compiled by the author for this project and is presently located at Hildesheim University, Institute for Information Science and Natural Language Processing. It covers domain-typical genres such as specialised magazines, handbooks, product descriptions and data sheets.

The procedure for frame-based multiword term extraction involves four major steps, illustrated in Figure 2: corpus preprocessing, term extraction, frame-based corpus annotation and frame-based term extraction.

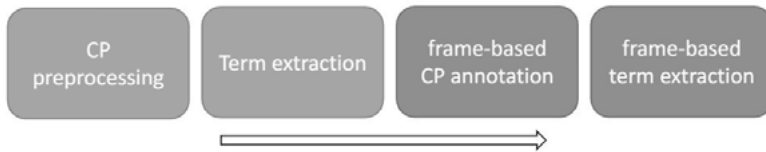


Fig. 2: Steps in the preparation and execution of MWT extraction. In the last two steps, the frame-based method comes into play.

The focus of the following sections will be on frame-based corpus annotation and frame-based term extraction.

3.1 Frame-based corpus annotation

Corpus preprocessing includes POS-tagging with the RFT tagger (Schmid & Laws 2008) and Corpus Workbench encoding (Evert & Hardie 2011). On this basis, automatic term extraction is carried out (Schäfer et al. 2015) to obtain a list of nominal term candidates. Term candidates, which are both single-word terms and multiword terms, are initially used to identify terminological strings (words or word stems) relevant to the selected frame FUNCTIONALITY.

The final purpose of this task is to compile a seed lexicon of strings-tags associations, with tags indicating frame elements. Each frame element is linked to a set of seeds. For instance,

- (1) MATERIAL > *baumwoll, glas, holz, cellulose, ...*

- (2) MATERIAL ORIGIN > *natur, pflanz, herkunft, ...*
- (3) MATERIAL PRODUCTION TECHNIQUE > *bläh, back, ...*

These sets are usually unique, since they uniquely refer to a specific frame element of a specific frame. However, a few exceptions are given, for example whenever polysemy cannot be avoided even under the relatively strict above-mentioned conditions. This happens for instance with the polysemous *-dämm-*, which has been attributed to different frame elements, namely GOAL, RESULT, and PRODUCT.

As previously pointed out, strings can match word parts and sometimes entire words. The choice of selecting word stems as the core semantic components of terms has to do with the requirements of frame-based annotation and the fact that compound constituents, as shown in Section 2.1, are mostly associated with different frame elements. *Blähperlite*, for instance, can be decomposed in *bläh*(MATERIAL PRODUCTION TECHNIQUE) and *perlite*(MATERIAL). For the purpose of annotating this term in the corpus, we would ideally need a good compound splitting tool able to detect the lemmatised forms *blähen* and *Perlite*. This would give us the opportunity to create a tagset with lemmas and frame elements. Unfortunately, experiments in compound splitting conducted at the beginning of the study returned no satisfactory results, which finally led to the option of using pre-defined word stems.

The tagset with seeds and frame elements was then employed to automatically annotate all corpus tokens containing the defined seed strings. 869,158 tokens were found matching these strings, whereas 162,462 tokens were also attributed an alternative annotation. This procedure was reiterated a few times, each time optimizing the original tagset by adding, modifying or re-ordering the seed strings, and adding a couple of new frame elements. The outcome of this step is a corpus in which all terms matching one or more seeds from the existing tagset are annotated with the corresponding frame element.

3.2 Frame-based multiword term extraction

After annotating the corpus with relevant frame-based information, two specific goals can be formulated for what concerns automatic MWT extraction:

- Identifying multiword terms with different syntactic distributions (*variant shapes*) of the same frame elements, e.g. paraphrases of compounds (see Table 3 for examples of this kind). This goal has been pursued by using the initial set of annotated seed strings.

- Detecting new relevant seed strings starting by the available ones.

3.2.1 Identifying different variant shapes

In our hypothesis, the phraseological nature of multiword terms is correlated to the nature of their constituents more than to their possible syntactic structures. The presence of different degrees of familiarity for the expert or semi-expert user does not impair this hypothesis, which also explains the tendency to term variation. Terms such as *Holzfaserdämmplatten*, *Dämmplatten aus Holzfasern* and *aus Holzfasern hergestellten Dämmplatten* can be thought of as members of a terminological continuum in which familiarity may decrease but not totally disappear. The phraseological continuum would even expand to cover combinations with a different head, such as *Dämmung mit Holzfaserplatten*, *Wärmedämmung mit Holzfaserplatten* and *dämmen mit Holzfaserplatten*. In general, this phenomenon seems to apply more to specialised language than to general language, in which syntactic structures strongly influence meaning, compositionality and idiomaticity of word combinations.

This conception of multiword terms as typical and, at the same time, multifaceted phraseological units of specialised languages justifies the need for the identification of different variant shapes of multiword terms in the corpus. This task begins with the analysis of the constituent elements of annotated compounds. Given the set of strings within a compound, we compute all its possible variant shapes, e.g.

Vakuumisolationspaneel
 ‘vacuum insulated panel’
 s={*vakuum, isolation, paneel*}

with four different variant shapes:
 s_v1 = {{*vakuum*}, {*isolation*}, {*paneel*}}
 s_v2 = {{*vakuum, isolation*}, {*paneel*}}
 s_v3 = {{*vakuum*}, {*isolation, paneel*}}
 s_v4 = {{*vakuum, paneel*}, {*isolation*}}

Within each variant shape (s_v...) we indicate the way in which the elements of a compound are grouped together in sets indicated by curly brackets, or non-grouped. For instance, s_v1 refers to all corpus instances in which the strings *vakuum*, *isolation* and *paneel* are separated from each other, whereas s_v3 refers

to corpus instances in which *vakuum* is separated from the compound made up of *isolation* and *paneel*. Experiments carried out beyond sentence boundaries did not produce good results, and for this reason we restricted the span of variation by only considering single sentences. However, permutations of sets within a variant shape are allowed, which opens up the possibility of finding the same set in just a different order. According to sv_3, for example, the term *Isolations-paneel* can be preceded or followed by the string *vakuum*.

Variant shapes are computed for each frame-based-annotated compound in the corpus. Multiword term variants corresponding to the different shapes are then extracted and manually validated.

3.2.2 Detecting and semantically delimiting new variants

Variant extraction described so far strictly depends on the tagset of seed strings and associated frame elements presented in Section 3.1. This tagset, in turn, depends on manual annotation of a relatively small sample of technical texts from the corpus. For this reason, new variants need to be detected which possibly contain new seed strings in order to raise the quantity of variation data in the termbase and to enlarge the tagset by inserting new seed. The latter goal enables the activation of an iterative process for the progressive detection of more (and more refined) data by starting from a frame-based corpus annotation.

The employed method for detecting further variants consists of leaving a string unspecified when computing variant shapes. Given the example of

Vakuumisolationspaneel
 ‘vacuum insulated panel’
 $s=\{\textit{vakuum}, \textit{isolation}, \textit{paneel}\},$

we decide to leave out a string and only indicate the frame element tag to which it is linked. *Paneel*, for instance, is a string associated with the frame element FORM, so the new set of strings will be

$s_1=\{\textit{vakuum}, \textit{isolation}, \text{FORM}\}.$

One string at a time is left unspecified, in order to proof all possible elements of a compound. This method allows us to find multiword terms with FORM corresponding to already available seed strings, for instance *platte* (*Vakuumisolationenplatte*). However, this would just confirm the results obtained by the steps described in Section 3.2.1. The main advantage of this procedure lies in the possibility of detecting new seed strings in a given position of a multiword term and to tentatively attribute them to pre-defined frame elements, thus producing a new frame element-string pair candidate.

The term extraction methods presented so far have the major drawback of producing a high number of candidates, and manual validation turns out to be extremely time-consuming. The main problem encountered when analysing these results is that many candidates appear to be well-formed but

- a. a frame element-string attribution is not correct, or
- b. a multiword term is not relevant to the frame FUNCTIONALITY.

A solution to issue (a) is not yet available. Rather we exploited this kind of information to gradually adjust and expand the existing tagset.

Issue (b) also became central during the course of the study. A number of multiword terms could be validated as technical terms as such but had to be rejected as not being related to the selected frame. These terms had been extracted on the basis of our frame-based procedure since they were combinations of correct seed strings.

For instance, the compound *Perimeterkleber* is the combination of the strings *perimeter* and *kleb*, both available in the original tagset and relevant for terms such as *Perimeterwand* or *Klebeteknik*. However, *Perimeterkleber* is only marginally related to the functionality of thermal insulation products and therefore cannot be validated. A similar example is given by the compounds *Naturschutzbund*, *Naturschutzgebiete*, *naturschutzgesetzlichen*, *Bundesnaturschutzgesetz*, which include the two relevant seed strings *natur* (like in *Naturmaterial*) and *schutz* (like in *Brandschutz*). These terms could not be validated either.

A step towards the solution of this issue was made by applying ontological constraints to frame elements combinations. The architecture of the domain ontology mentioned in Section 2.2 is based on three macrocategories, from which all classes and instances in the ontology depend, namely MATERIAL, FORM, and FUNCTION. The ontology was built with the description of the functionality of thermal insulation products in mind (i.e. it is intended as an excerpt from an overall ontology of thermal insulation). For this reason, the frame FUNCTIONALITY, which is the frame under observation, is also strictly depend-

ent on the concepts covered by these macrocategories. Terms denoting thermal insulation products are typically of the kind *Polystyrolämmplatte*, *Schafwoll-Dämmmatte*, *wood fibre insulation board*, *fiberglass insulation loose fill*, or *lastra isolante in lana di roccia*, to mention just a few examples in German, English and Italian. These terms include word stems related to the three frame elements MATERIAL, FORM, and GOAL (which is one of the possible frame elements linked to the FUNCTION macrocategory), as shown in Table 4.

Tab. 4: Typical frame element composition of terms denoting thermal insulation products in German, English and Italian

	MATERIAL	GOAL	FORM
<i>Polystyrolämmplatte</i>	<i>polystyrol</i>	<i>dämm</i>	<i>platte</i>
<i>Schafwoll-Dämmmatte</i>	<i>schafwoll</i>	<i>dämm</i>	<i>matte</i>
<i>wood fibre insulation board</i>	<i>wood fibre</i>	<i>insulation</i>	<i>board</i>
<i>fiberglass insulation loose fill</i>	<i>fiberglass</i>	<i>insulation</i>	<i>loose fill</i>
<i>lastra isolante in lana di roccia</i>	<i>lana di roccia</i>	<i>isolante</i>	<i>lastra</i>

We observed that core⁴ frame elements could be subdivided in what we called *primary* and *secondary* frame elements, depending on their semantic proximity to the three abovementioned frame elements MATERIAL, FORM, and GOAL, which are the closest to the ontological macrocategories. Secondary frame elements are connected to the primary ones in virtue of conceptual-semantic dependencies. PACKAGING, for instance, depends on FORM, whereas PROPERTY (intended as a physical or chemical property) depends on MATERIAL. A hierarchy of frame elements was thus created in which combinations of frame elements had different ‘weights’ in terms of their relevance to the frame FUNCTIONALITY. Not all possible combinations, in fact, identify relevant multiword terms. The criteria for selection can be summarised as follows:

- primary frame elements can combine with each other;
- a secondary frame element can combine with its specific primary frame element (a secondary frame element can also combine with a further primary frame element via the superordinate primary frame element);
- a secondary frame element cannot combine with another secondary frame element.

⁴ The terms *core* and *non-core* are understood in the sense of Fillmore (2008).

As a consequence, all multiword term candidates not complying with these rules have to be rejected. Going back to the previous examples of *Perimeterkleber*, we have the following, non-conforming frame element combination:

target/*perimeter*, application technique/*kleb*,

in which TARGET and APPLICATION TECHNIQUE are two secondary frame elements related to GOAL. The same applies to *Naturschutzbund* and the morphologically similar terms, since they contain a combination of two secondary frame elements related to MATERIAL:

material origin/*natur*, property/*schutz*.

This method allows for the computation of all possible frame element combinations and the exclusion of the non-relevant ones during multiword term extraction. The final output is semantically more fine-grained and quantitatively more manageable from the perspective of manual candidate validation and representation in a terminographic or lexicographic resource.

3.3 Validation

Validation of the method in its various aspects has been carried out during and after the end of the study on different technical and non-technical domains. In this section, a brief chronological overview of the main applications will be provided.

- Domain of semiconductor diodes: application of the ontology-oriented constraints to frame element combinations (Giacomini 2019b).
- Domain of DIY power tools: application of the frame-based corpus annotation method (Giacomini 2019b).
- Domain of environment: application of the frame-based approach to terminology description and of the variation typology.
- Domain of graph theory (ongoing work): application of the variation typology (Kruse & Giacomini 2019).
- Domain of immunology (ongoing work): application of the frame-based approach to terminology description and of the variation typology.

Moreover, the model for variation classification will soon be applied to multiple domains in the context of a research project for the rebuilding of an existing

terminology database. Tests performed so far prove the developed variant typology as well as the method for frame-based annotation and extraction to be suitable for application to multiword terms of different specialised fields.

Evaluation of the model for frame-based variant extraction was also performed by comparing our results with those obtained by using *TermSuite* (Daille 2017) and the tools developed in the context of the *T&O* (Terminology and Ontology) project⁵ on the *Insulation Corpus*. Multiword term variants retrieved by applying frame element permutations and ontological restrictions, namely a total of over 1 million n-ary combinations, constituted the gold standard for the comparison. Results of the evaluation procedure (cf. Giacomini 2019b for a detailed description) show that our model for frame-based extraction is capable of detecting a significantly higher number of multiword terms than similar devices. Most importantly, extracted terms are grouped in semantically homogeneous variant clusters thanks to their association with frame element tags.

4 Conclusions

This contribution argues in favour of a phraseology-oriented treatment of multiword terms in specialised language, which draws on findings of previous studies (cf., for instance, Aguado de Cea 2007). This view of terminology is in line with the notion of non-diasystemic variation as the co-existence of contextual synonyms which can be described according to their orthographical, morphological and syntactic features. Against this background, a frame-based approach to terminology offers several advantages. Frames provide a useful semantic interface between terminology and ontology by representing typical situations involving a specific domain entity. Semantic annotation by means of frame element tags complements syntactic annotation and supports the detection of term variation, especially for what concerns multiword terms. On the grounds of frame-based annotation, a procedure for multiword term extraction can be established that

- identifies all possible variant shapes of annotated seed strings,
- discovers new relevant terminological items by leaving multiword term constituents unspecified for a certain frame element, and
- restricts the range of multiword terms in a given frame by applying a set of ontological constraints.

⁵ A joint project of IMS Stuttgart and Robert Bosch GmbH.

Data obtained by this method are suitable for representation in a lexicographic information tool such as an electronic specialised dictionary.

Ongoing work mentioned in Section 3.3 will be paired in the future with further investigations concerning the application of the model to multiple frames within the same domain. Research focusing on this topic would possibly unlock the potential of frame-based models for the analysis of larger, multi-thematic corpora.

References

- Aguado de Cea, Guadalupe. 2007. A multiperspective approach to specialized phraseology: Internet as a reference corpus for phraseology. In Santiago Posteguillo, María José Esteve and María Luisa Gea-Valor (eds.), *The Texture of Internet: Netlinguistics in progress*, 182–209. Newcastle: Cambridge Scholars Publishing.
- Burger, Harald. 2015. *Phraseologie: eine Einführung am Beispiel des Deutschen*. Berlin: Erich Schmidt Verlag.
- Caro Cedillo, Ana. 2004. *Fachsprachliche Kollokationen: Ein übersetzungsorientiertes Datenbankmodell Deutsch-Spanisch*. Tübingen: Gunter Narr Verlag.
- Daille, Beatrice. 2003. Conceptual structuring through term variations. In *Proceedings of the ACL 2003 workshop on Multiword expressions: analysis, acquisition and treatment*, 9–16.
- Daille, Beatrice. 2017. *Term Variation in Specialised Corpora: Characterisation, automatic discovery and applications*. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Evert, Stefan & Andrew Hardie. 2011. Twenty-first century corpus workbench: Updating a query architecture for the new millennium. In *Proceedings of the Corpus Linguistics Conference 2011. University of Birmingham, England*. <https://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2011/Paper-153.pdf> (accessed 15 April 2020).
- Faber, Pamela (ed.). 2012. *A cognitive linguistics view of terminology and specialized language*. Berlin: de Gruyter.
- Faber, Pamela. 2015. Frames as a framework for terminology. In Hendrik J. Kockaert & Frieda Steurs (eds.), *Handbook of terminology*, 14–33. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Fillmore, Charles J. 1985. Frames and the semantics of understanding. *Quaderni di semantica* 6(2). 222–254.
- Fillmore, Charles J. 2008. Frame semantics. In Dirk Geeraerts (ed.), *Cognitive linguistics: Basic readings*, 373–400. Berlin: de Gruyter.
- Freixa, Judit. 2006. Causes of denominative variation in terminology. A typology proposal. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 12(1). 51–77.
- Freixa, Judit. 2013. Otra vez sobre las causas de la variación denominativa. *Debate Terminológico* 9. 38–46.

- Giacomini, Laura. 2017. An Ontology-terminology Model for Designing Technical e-dictionaries: Formalisation and Presentation of Variational Data. In *Electronic lexicography in the 21st century: Proceedings of eLex 2017, Leiden (NL)*, 110–123.
- Giacomini, Laura. 2018. Frame-based Lexicography: Presenting Multiword Terms in a Technical E-dictionary. In *Proceedings of the XVIII EURALEX International Congress, 2018*, 309–318.
- Giacomini, Laura. 2019a. Functional Representation of Technical Artefacts in Ontology-Terminology Models. In *Proceedings of the 2nd Conference on Language, Data and Knowledge (LDK 2019), Leipzig (DE)*, 5:1–5:6.
- Giacomini, Laura. 2019b. *Ontology – Frame – Terminology. A method for extracting and modelling variants of technical terms*. Habilitation Thesis, University of Hildesheim.
- Gläser, Rosemarie. 1994. Relations between phraseology and terminology with special reference to English. *Alfa: Actes de langue française et de linguistique* 7–8. 41–60.
- Hanks, Patrick. 2010. Terminology, phraseology, and lexicography. In *Proceedings of the XIV Euralex International Congress, Leeuwarden (NL)*, 6–10.
- Kruse, Theresa & Laura Giacomini. 2019. Planning a domain-specific electronic dictionary for the mathematical field of graph theory: definitional patterns and term variation. In *Electronic lexicography in the 21st century: Proceedings of eLex 2019, Sintra (PT)*, 676–693.
- Lyons, John. 1977. *Semantics*. Cambridge: Cambridge University Press.
- Schäfer, Johannes, Ina Rösiger, Ulrich Heid & Michael Dorna. 2015. Evaluating noise reduction strategies for terminology extraction. In *Proceedings of the 11th International Conference on Terminology and Artificial Intelligence, Granada, Spain*, 123–131.
- Schmid, Helmut & Florian Laws. 2008. Estimation of conditional probabilities with decision trees and an application to fine-grained POS tagging. In *Proceedings of the 22nd International Conference on Computational Linguistics*, 777–784. Manchester, United Kingdom: Association for Computational Linguistics.
- Sinclair, John. 2004. The search for units of meaning. In John Sinclair & Ronald Carter (eds.), *Trust the Text. Language, corpus and discourse*, 24–48. London: Routledge.
- Stubbs, Michael. 2001. On inference theories and code theories: Corpus evidence for semantic schemas. *Text* 21(3). 437–465.

Martina Häcker

Cross-language transfer of formulae

The case of English letters

Abstract: In England before the Norman Conquest letters were written in both English and Latin, while in the Middle English period the medium of letter writing was Latin and French up to the end of the fourteenth century (Häcker 2011). Around 1400 French was replaced by English in all letters, except for communication involving the clergy, in which Latin continued to be used. This raises the question of how far French and Latin epistolary traditions influenced English letter writing, in particular whether and to what extent English letter formulae are loan translations.

To answer these questions a database of letters written in England in both Middle English and French was investigated. The analysis shows that letters contain a high number of formulae and that these correspond closely in English and French letters; in other words, the English formulae are calques of French ones. The evidence regarding the availability and content of letter writing manuals suggests the following scenario for the acquisition of letter writing formulae. It appears that both French and Latin letter composition was taught by tutors with the help of manuals, while in the case of English, it is likely that after a period of *ad hoc* translations by bilinguals (most likely tutors), the formulae became conventionalized and spread quickly. The competent use of the formulae by women indicates that formulae were also acquired via oral input outside the classroom.

Keywords: acquisition of letter formulae, language transfer, medieval French, Middle English.

Martina Häcker, University of Siegen, martina.haecker@uni-siegen.de

<https://doi.org/10.1515/9783110520569-010>

1 Introduction

1.1 Language contact in medieval Britain

Various languages were in contact in the medieval period in the British Isles: Latin and Celtic languages, Latin and Old English, Old English and Old Norse, and English and varieties of French. Of these, only Latin and French, the two contact languages that were extensively used in writing, left a major impact on English vocabulary, and it is the contact with French that received most scholarly attention, first in studies on loanwords (e.g. Serjeantson 1935, Suggett 1946), and more recently in studies on bi- and multilingualism and code-switching, as documented *inter alia* in studies by Short (1980), Rothwell (2001), Nurmi and Pahta (2004), Ingham (2012), and the contributions to edited volumes such as *Multilingualism in Later Medieval England* (Trotter [ed.] 2000), *Multilingualism in medieval Britain (c. 1066-1520): Sources and analysis* (Putter and Jefferson [eds.] 2012), and *Code-Switching in Early English* (Schendl and Wright [eds.] 2011). The sources used in these studies were predominantly documentary and literary texts, but there are also two studies on code-switching in letters (Schendl 2002, Nurmi and Pahta 2004), which include switches between English and French, although these are much less frequent than the ones between English and Latin. Nurmi and Pahta (2004) point out that many switches they found could be interpreted as borrowing, in particular in the case of Latin legal terminology. By contrast, there has been much less research on English calques of French phraseology, with the exception of Prins (1952), whose study investigated both borrowing and calquing.

1.2 Conventions in letter writing

Letters contain both conventional and creative, discursive, parts. The opening and closing sections tend to be conventional and formulaic, while the body of the letter is predominantly non-formulaic and creative; that is, it will usually contain phrasing relevant to a specific situation. The degree to which a letter contains creative parts depends on the purpose of writing, on the relationship with the addressee and on cultural conventions. Cultural conventions may differ considerably from one society to another. For example, in business correspondence in modern Western societies it is uncommon to ask about the addressee's health and that of his or her family. In other societies such as that of China and Japan, for example, failing to do so may be considered impolite (Bo-

iarsky 1992, King n. d.). But even in Western societies we have different conventions. In present-day Britain, the U.S.A., the German-speaking countries and the Scandinavian countries the valediction is short, while in French-speaking countries it often is more elaborate, with forms such as *Veillez accepter Madame, Monsieur mes plus respectueuses salutations*. Cultural conventions also change over time, whether as the result of changes within a language community or through contact with other communities with different conventions.

Medieval letters belong to a different culture with different conventions from that of present-day Western societies. The purpose of this paper is to investigate the conventions of Middle English letters with a main focus on the use of formulaic language and on the input of different traditions and cultures on Middle English letter writing. In the period 1100–1400, it was common practice for members of the clergy to write their letters in Latin and for members of the nobility and gentry to write letters in Latin or French, while before 1100 there had been a tradition of writing letters in the vernacular. This means that there was a change in the medium of letter writing from Old English to a variety of French. This raises the question of whether the shift to French also entailed a change in epistolary tradition.

1.3 Formulaic language and letter formulae

Formulaic language is a neutral term for a concept that has been described by some fifty different labels such as multi-word unit, fixed expression, formulaic sequence and prefab (Wray 2002: 8) and includes a wide range of different types (inter alia collocations, idioms, proverbs, sayings, but also complex prepositions and phrasal verbs) which differ in length and degree of fixedness. There are three approaches to formulaic language: (i) a phraseological approach, with semantic non-compositionality as its defining criterion (Cowie 1998), (ii) a frequency-based approach (Sinclair 1991, Stubbs 2004, Biber 2009) and (iii) a psychological approach, whose defining criterion is holistic storage (Wray 2002 and 2008). In contrast to the other two approaches, Wray's definition of formulaic language includes single-word units. For an investigation of formulaic language used in letter writing, the phraseological approach is too restrictive, but both frequency and holistic storage are relevant: formulae are recognized as such because they recur in different letters and holistic storage explains multiple use of identical formulae across letters and people.

The term *formula* is used in this study for recurrent linguistic units with specific discourse functions. Two levels are distinguished: the functional level

denoting the discourse function (e.g. salutation), and the realisation level, that is, the actual formula (e.g. *Ryght reverent and worshyful Syr*) or formulae if there is more than one.

Although the term *formula* indicates fixedness, it is important to bear in mind that formulae can be completely fixed or contain slots for variable and optional items, and that they show social and diachronic variation. Some letter formulae are completely fixed, while others show extensions (mostly modification by adjectives and adverbs) or replacement by near synonyms. In addition, formulae can be transferred from one language to another, and this transfer can either be a direct loan or a loan translation ('calquing'). It will be shown below that calquing plays a major role in the creation of Middle English letter formulae.

Research on historical letter formulae is limited and mostly restricted to address formulae (Nevalainen & Raumolin-Brunberg 1995, Raumolin-Brunberg 1996, Nevala 2004, Nevala 2007, Tiisala 2007, and Häcker 2019); in addition there are a few studies which include observations on other formulae (Elspaß 2012 and Laitinen and Nordlund 2012 for the nineteenth century, and Sánchez Roura 2000 and Häcker 2011 for Middle English). Sánchez Roura analyzed the *Cely letters* with respect to intra- and cross-writer variation from the perspective of politeness theory, while the focus of Häcker 2011 is the impact of French lexis and phraseology on Middle English as documented in three family correspondences. Apart from my limited study, there is no other study on the transfer of letter formulae, although Davis (1965) points out some correspondences between formulae used by Chaucer in *Troilus and Criseyde* and his French source.

2 Database and Methodology

The database used for this investigation consists of four Middle English family correspondences, those of the Celys (Hanham [ed.] 1975), Pastons (Davis [ed.] 1971 and 1976), Stonors (Carpenter [ed.] 1996) and Plumpton's (Kirby [ed.] 1996), as well as Tanqueray's (1916) edition of French letters written in England.¹ The Celys were a family of London wool merchants, with one employee based in Calais, who dealt with the wool sale on the Continent. The Pastons, Stonors and

¹ As the difference between Anglo-Norman and French is of no consequence for the content of this study, no distinction between the two varieties will be made and the term *French* will be used for both.

Plumpton were landed gentry in Norfolk, Oxfordshire and Yorkshire respectively. All of the family correspondences as well as the letters in Tanqueray's edition are predominantly business letters, as purely private letters were extremely rare in the Middle Ages. Tanqueray's compilation spans the years 1265–1399, and consists of a total of 164 previously unedited letters from the following sources: (i) the 58 volumes of “ancient correspondence” in the Public Record Office, which are addressed to the king or chancellor, (ii) Patent Rolls, (iii) Close Rolls, (iv) Chancery Files, (v) the Coucher Books, (vi) Cathedral registers, and chapter libraries (Tanqueray 1916: vii-x). The *Stonor letters* cover the period 1290–1483, the *Paston letters* the period 1425–1495, the *Plumpton letters* the period 1433–1502, and the *Cely letters* the period 1472–1488. While the *Stonor letters* contain 32 French letters and 223 English ones, the letters from the other three family correspondences are predominantly in English, with only two in French and one in Dutch (written by continental correspondents) in the 250 *Cely letters* and a French endorsement in a letter by William Paston I (no. 4, dated 1426) in the 838 Paston letters.² In addition, Latin texts can be found in all four of the family correspondences, mostly legal documents, but also a few letters by members of the clergy, which will not be analyzed, as the focus of this study is on English and French letters. The letters were analyzed manually for recurrent phrases or sentences, as (i) only close reading allows us to identify unknown formulae, (ii) few of the texts are available electronically, and (iii) due to the wide range of spelling variations the texts are unsuitable for electronic searches. It needs to be born in mind that the survival of letters is a matter of historical coincidence; letter collections mostly survived as a result of legal disputes, which means that many others are likely to have disappeared without a trace.

3 Overview of formulae used in Middle English letters

Before turning to the origins of formulae, I will provide a brief overview of the most important formulae I have identified in Middle English letters. These can be illustrated by examples from the late fifteenth-century *Cely letters*, the correspondence of a London family of wool merchants. It contains letters between fami-

² In addition to letters, the English material also includes some accounts and memoranda; these are not included in the figures.

ly members, that is, mostly Richard Cely the Elder and his sons George Cely and Richard Cely the Younger, between the Celys and their employees, and between the Celys and other wool merchants. In the quotations from the letters, abbreviations have been silently expanded and only the letter numbers of Hanham's (1975) edition are provided for reference. All examples are from letters dating between 1476 and 1482.

Regularly recurring formulae are found at the beginning of letters (salutation formula, recommendation formula, welfare formula, health formula, acknowledgement formula, and astonishment formula)³. They introduce the purpose for writing the letter (information formula and request formula), and they are also found in the final part of the letter (closing formula, benediction formula, haste formula, place and date formula, and subscription formula). Not all of these occur in all letters; in particular, the acknowledgement formula and the astonishment formula tend to be alternatives, and there is also a low co-occurrence of the information formula and the request formula. The most frequent formulae are salutation, subscription and recommendation, which are found in all writers except members of the Pastons born before 1450 and Richard Cely the Elder, whose letters usually have subscriptions but may lack salutation and recommendation.

Formulae which show little variation are the following:

Recommendation formula

- (1) *After dew recommendaschon I louly recommend me vnto yowre masterschypp, etc*⁴.

(William Cely to George Cely, no. 163)

³ The labels introduced here are mine. There is some overlap with Sánchez Roura (2000), although her categories are generally broader than mine and she only includes formulae that can be treated under the heading of *captatio benevolentiae*. In the following list of correspondences my term is followed by Sánchez's. Recommendation formula = commendation; welfare formula and health formula = health matters; acknowledgement formula = acknowledgement of receipt; astonishment formula: not listed by Sánchez; information formula = introduction of the exposition; request formula: not listed by Sánchez; closing formula = end of news; haste formula = in haste; benediction formula = pious valediction (also under "offer of services"); place and date formula: not listed by Sánchez; endorsement formula: not listed by Sánchez; salutation formula = address; subscription formula: not listed by Sánchez; offer of services: only listed by Sánchez; gratitude: only listed by Sánchez.

⁴ The formula used here is a combination of two formulae: *after all due recommendation and I recommend me to you*.

Welfare formula

- (2) *Desyring to here of your prosperous welfare, which Jhesu preserue to th'accomplysshment of your hertys desire*
(Draft in name of Richard Cely, no. 16)

Health formula

- (3) *At the making of thys owr father and mother wher in good heyll, thankyd be God*
(Richard Cely the Younger to George Cely, no. 19)

Acknowledgement formula

- (4) *I haue resayuyd a letter from you wryt at Caleys the xxij of Auguste, the weche I haue wyll understand*
(Richard Cely the elder to George Cely, no. 98)

Astonishment formula

- (5) *I marwhell grettely*
'I wonder greatly'
(Richard Cely the Younger to Joyce Parmentier, no. 126)

Information formula

- (6) *Plessed yow to wete that*
'Please it you to know that'
(William Maryon to George Cely, no. 130)

Request formula

- (7) *And I pray yow*
(Richard Cely the Younger to George Cely, no. 117)

Closing formula

- (8) *No more to you at thys tyme*
(Richard Cely the Younger to George Cely, no. 74)

Haste formula

- (9) *Wryt ... in grete haste*
(Richard Cely the Elder to George Cely, no. 11)

Benediction formula

- (10) *The Trenyte haue yow in hys blessed kepyng*
(William Maryon to Richard Cely the Younger, no. 40)

Place and date formula

- (11) *Wrytyn at London the iiijth day of Juyn*
(Richard Cely the Younger to George Cely, no. 117)

Endorsement formula

- (12) *To the ryght worschipful George Cele, merchand of the Stapull at Cales be þys delyuerd*
(John Spencer to George Cely, no. 11)

The next two types of formulae, that is, salutation and subscription, are more variable, as a range of different address forms and epithets is used, depending on the relationship between writer and addressee.⁵ These can be viewed as semi-fixed with open slots for address terms and modifiers appropriate for the addressee, as well as showing diachronic variation. Thus *I greet you weel*, which is a structurally different formula from the dominant vocative one, is clearly restricted to the older generation, and only found in Richard Cely the Elder's letters to his sons (but also in early Paston and Stonor letters).

⁵ *Writer* will be used as a cover term for the composers of letters, irrespective of whether a letter was a holograph or dictated to a scribe.

Salutation formulae

- (13) *Ryght ruerent syr and my specyall ffrende*
(William Maryon to George or Richard Cely the Younger, no. 39)
- (14) *Ryght rewerent and whorshipfful ffadyr*
(George Cely to Richard Cely the Elder, no. 41)
- (15) *I gret you wyll*
'I greet you well'
(Richard Cely the Elder to George Cely, no. 90)
- (16) *Ryght worschypfull syr*
(William Cely to George Cely, no. 128)
- (17) *Ryught enterly whelbelouyd brother*
(Richard Cely the Younger to George Cely, no. 83)
- (18) *Brother George*
(William Dalton to George Cely, no. 138)
- (19) *Whelbelouyd*
(Richard Cely the Younger to Joyce Parmenter, no. 126)

Subscription formulae

- (20) *By youre owne, Wylliam Maryon*
(William Maryon to George or Richard Cely the Younger, no. 39)
- (21) *Per yowr son, George Cely*
(George Cely to Richard Cely the Elder, no. 41)
- (22) *Per Richard Cely*
(Richard Cely the Elder to George Cely, no. 90)
- (23) *Per Rychard Cely the zeungar*
(Richard Cely the Younger to Joyce Parmenter, no. 126)

- (24) *Per yowr seruaunte, Wylliam Cely*
(William Cely to George Cely, no. 128)
- (25) *Per your brother, Rycharde Cely*
(Richard Cely the Younger to George Cely, no. 83)
- (26) *Your B[rother] W. Dalton*
(William Dalton to George Cely, no. 138)
- (27) *Your verey ffrynd Nycholas Knyveton*
(Nicholas Knyveton to George Cely, no. 159)
- (28) *Yours, Wylliam Adam*
(Wylliam Adam to George Cely, no. 166)

The variation shows that, when writing to parents or brothers, sons and brothers use a kinship term in their salutation, that is, *father* or *brother*, and a subscription consisting of *per* followed by kinship term and name. By contrast, Richard Cely the Elder uses neither a kinship term nor a name in the salutation in letters to his sons, and he uses the subscription *Per Richard Cely*. Employees use *sir* in the salutation and *servant* followed by name in the subscription. Both employees and sons mostly use the epithet *right worshipful* in their salutations, which may be expanded to *right reverent and worshipful*. Fellow merchants use the salutation *brother* followed by name (mostly first name) and most frequently the subscription *your brother* followed by name. Nicholas Knyveton's subscription *Your verey frynd Nicholas Knyveton* indicates a more personal relationship to George Cely, as does Richard Cely the Younger's salutation *Whelbelouyde* in his letter to Joyce Parmenter (for the use of *wellbeloved*, see Häcker 2019: 107–113). A special case is William Maryon, who was a fellow merchant but also employed by the Celys, as well as being a close family friend. This double relationship is reflected in the salutation *Ryght reuerent syr and my specyall ffrende* in his letter to George Cely, and in the subscription *By youre owne Wylliam Maryon*.

The same formulae are also found in the *Paston letters* (Davis [ed.] 1971 and 1976), the *Stonor letters* (Carpenter [ed.] 1996), and the *Plumpton letters* (Kirby [ed.] 1996). With the exception of the salutation and the subscription formulae, the formulae are fairly stable across ranks and across time.

4 Formulae in pre-Conquest letters

Anglo-Saxon Writs (Harmer [ed.] 1989 [1952]) contains documents that we would today classify as letters.⁶ These allow us to point out some characteristic differences from contemporary Latin letters. The salutation of these Old English letters (or writs in Harmer's terminology) consists of the following elements: Sender [name title-NOM] *grets* addressee [name title-OBL] *humbly/like a friend*. From the salutation the text moves straight to the purpose of writing:

- (29) *Ælfric gret eadmodlice Æðelwerd ealdorman and ic secge þe, leof ...*
 'Ælfric sends humble greetings to ealdorman Æðelwerd, and I tell you [sg] beloved ...'⁷
 (Harmer [ed.] 1989: 23)
- (30) *Ælfric abbod gret Sigefyrð freondlice. Me is gesæd ...*
 'Abbot Ælfric sends friendly greetings to Sigeferth. I am told ...'
 (Harmer [ed.] 1989: 23)

By contrast, contemporary Latin letters contain epithets for both the sender and the addressee and in (31) a different greeting formula:

- (31) *Ælfric humilis servulus Christi, honorabili et amando Archiepiscopo Sigerico perpetuum sospitatem optat in Domino.*
 'Ælfric, a humble servant of Christ, wishes the honorable and loveable archbishop Sigeric everlasting happiness in God.'
 (Harmer [ed.] 1989: 21)

The style in this letter is similar to Latin letters written on the Continent; the same grammatical construction is used in a letter written by Charlemagne to Offa, King of Mercia:

- (32) *Karolus gratia Dei rex Francorum et Langobardoum et patricius Romanorum viro venerando et fratri karissimo Offæ regi Merciorum salutem.*

⁶ Harmer ([ed.] 1989: 1) defines the term *writ* as follows: "The 'Anglo-Saxon writ' was a letter on administrative business to which a seal was appended, and the protocol (or opening clauses) of which named the sender of the letter and the person or persons to whom it was addressed, and contained a greeting."

⁷ All translations of Old English, French and Latin text extracts are mine.

‘Charles by God’s grace king of the Franks and Langobards and patrician of the Romans [sends] greetings to his dearest brother Offa, king of the Mer-cians.’

(Harmer [ed.] 1989: 26)

The structure of the formula in both examples is much more elaborate, but otherwise identical, to the opening formula of classical Latin, which consists of the following structure:

(33) Sender-NOM addressee-DAT *salutem dicit*.

Cicero Caio suo salutem dicit.

‘Cicero sends [his] Caius greetings.’

In this formula the verb *dicit* is often omitted in classical as well as in medieval Latin. The *perpetuum sospitatem optat* in (31) is a replacement of the much more frequent *salutem dicit*.

Pre-Conquest Old English letters also contain benediction formulae, but in these the formulations vary. Harmer lists three different ones:

(34) *God eow (ealle) gehealde.*

‘God keep you all.’

(Harmer [ed.] 1989: 70)

(35) *God sy eower elra freond.*

‘God be the friend of you all.’

(Harmer [ed.] 1989: 71)

(36) *God be mid ihu/eow.*

‘God be with you.’

(Harmer [ed.] 1989: 71)

The information conveyed in an Old English letter is typically introduced by the following formula:

(37) *Ic cyðe eow/ðe/inc.*

‘I make known to you-PL/you-SG/you two.’

(Harmer [ed.] 1989: 65)

The admittedly limited body of Old English letters shows that only a few simple formulae were used and that the elaborate epithets characteristic of contemporary Latin letters written in England were not employed in vernacular letters.

5 Formulae in post-Conquest letters and their origin

A comparison of the Middle English letters from the family correspondences of the Pastons, Stonors, Celys and Plumpton shows an almost complete break with the Old English formulae discussed above. There is only one formula that survives into the fifteenth century. This is the salutation: [sender] *greet*s [addressee] ADV (see examples [29] and [30] above). A modified version of this formula *I greet the/you well* is used by Richard Cely the Elder and the older generation of the Pastons, and in a single instance by Richard Cely the Younger in a letter to an employee (no. 120).⁸ In the earlier *Stonor letters* it is preceded, with one exception, by salutations containing the epithet *wellbeloved*, which is used for close personal relationships in letters to people of a lower social rank or younger family members (Häcker 2019: 107–113).⁹ In the information formula and the benediction formula, the wording of pre-Conquest formulae differs from post-Conquest ones.

This dramatic change in the formulae used is part of a larger change in epistolary style, which was caused by a number of socio-cultural changes. Constable (1976: 2–24) relates it to a shift from the personal letter of antiquity to political and ecclesiastical letters (see also Lanham 1975), while Murphy (1974: 201–202) sees the cause for the change in (i) the emergence of a feudal society, which entailed a greater need for written communication, and (ii) a decrease in literacy, which meant that scribes were required. The new epistolary style was modelled on the *ars dictaminis* ('the art of formulation') and the rhetorical structure consisting of *salutatio* ('salutation'), *exordium/captatio benevolentiae/proverbium* ('opening'/'ensuring the addressee's goodwill'/'proverb'), *narratio* ('report'), *petitio* ('request') and *conclusio* ('closing'). The medieval *ars*

⁸ Agnes Paston, her sons John Paston I and William Paston II, and John Paston I's wife Margaret Paston (Davis 1971, no. 18, dated not after 1449, no. 62, dated 1462, no. 106, and no. 198, dated 1466).

⁹ Carpenter ([ed.] 1996) nos. 46, dated 1424, no. 112, dated 1470, no. 120, dated c.1472, no. 148, dated ?1475, and no. 193, dated 1475.

dictaminis originated with Alberic of Monte Cassino (floruit 1057–1087), and gradually spread through all of Christian Europe, reaching England with Peter of Blois in the 1180s (Murphy 1974: 202–214). Yet Alberic's *Breviarium de dictamine* (Bognini [ed.] 2008) is unlikely to have been widely used in teaching. This would have required a more practical manual such as Albert of Samaria's *Praecepta dictaminum* (Schmale [ed.] 1961), dated 1111–1118, which provided thirty-four model letters.

Unlike Old English letters, Middle English letters show the impact of the *ars dictaminis* in their range of salutations, which reflect the social status of writer and addressee. The rich inventory of Middle English formulae, as documented above, shows a close correspondence to French ones. The same range of formulae is documented in the later letters of Tanqueray's collection and the French letters of the Stonor correspondence, all of which predate the Middle English letters, but also in French letters from the Continent contemporary to the Middle English ones (*Cely letters*, no. 54, Clare to her lover George Cely, and no. 62, a business letter by Waterin Tabary to George Cely), which shows that the French formulae were used on the Continent as well. Corresponding pairs exist for all the formulae presented above, that is, recommendation formula, welfare formula, health formula, acknowledgement formula, astonishment formula, information formula, request formula, closing formula, haste formula, benediction formula, place and date formula, and endorsement formula, as well as salutation formula and subscription formula, but in the case of the latter two not all variants are identical. The endorsement formula is different in the French letters of the Stonor correspondence (it does not contain a verb), but a corresponding formula is documented in the endorsement of an early letter by William Paston I (no. 4, 1424) and in a French letter of the Cely correspondence (no. 62, 1479). Moreover, a Latin equivalent is documented in the Stonor correspondence (no. 57, dated “?before 1450”).

In the list below the basic formulae are italicized, to distinguish them from optional additions.

Recommendation formula

(38) a. *Jeo me recomans a vous si humblement come je sai et puis.*

(John Waltham, bishop of Salisbury to Richard II, Tanqueray, no. 161 [1388-91])

b. *I recomaunde me vnto yow as hartely as I can or may.*

(John Dalton to George Cely, no. 141, 1481/2)

Welfare formula

- (39) a. *en desirant souverainement d'oier bonx novelx de vostre tres noble seignurie, que pri a nostre tres doulx seignur Jhesu Crist qu'il le maynteigne pour sa grand misericorde.*
(Maud, nurse of Philippa, daughter of Jean de Gand, Tanquerey no. 163 [1389-99])
- b. *desyryng to here of your prosperous welfare, which Jhesu preserue to th'accomplysshment of your hert desire.*
(Draft in name of Richard Cely no. 16, 1477/8)

Health formula

- (40) a. *Et si de mon estat vous plese assavoir, al departier du cestes jestoie en sancte du corps, le mercie dieu.*
(Henry Dounham to Edmund Stonor, no. 28 [c. 1380])
- b. *And yf it lyke you ser to h[e]r [o]f my [h]elt[he], at the] making of thys sympyll letter I was in good helthe of bode, blessyd be J[hesu].*
(Margery Cely to George Cely, no. 222, [1484])

Acknowledgement formula

- (41) a. *Entendaunt, treshonoree sire, qe jai recu voz lettres a moi directez, ..., lesqueux jai pleinement entenduz.*
(John Stoke to Edmund Stonor, no. 36 [c. 1380])
- b. *I haue resayuyd a letter from the, wryte at Caleys the xxiiij day of Jun, the weche I haue wyll understand.*
(Richard Cely the Elder to Robert Cely, no. 2, 1474)

Astonishment formula

- (42) a. *Moi mervoile grauntement.*
(Johan de Nouwers to Edmund Stonor, no. 14, 1378)
- b. *I marwhell grettely.*
(Richard Cely the Younger to Joyce Parmenter, no. 126, 1481)

Information formula

- (43) a. *Vous plese a savoir.*
(unknown sender to Edmund Stonor, no. 27 [c. 1380])
b. *Pleased [Please it] yow to wete.*
(William Maryon to Richard Cely the Younger, no. 40 [1478])

Request formula

- (44) a. *Vous priouns cherement qils ...*
(Gilbert Talbert to Edmund Stonor, no. 22 [1378])
b. *And I pray yow.*
(Richard Cely the Younger to George Cely, no. 117, 1481)

Closing formula

- (45) a. *Autre chose quant a present.*
(Nicholas Cowley to Edmund Stonor, no. 8 [c. 1365])
b. *No more to you at thys tyme.*
(Richard Cely the Younger to George Cely, no. 74, 1479)

Benediction formula

- (46) a. *Dieu vous ayt en sa sainte garde.*
(Waterin Tabary to George Cely, no. 62 [before 12 October 1479])
b. *Jhesu haue you in hys keppying.*
(Margery Cely to George Cely, no. 222 [1484])

Date and place formula

- (47) a. *Escrit a Merlawe le xiiijme jour de Novembre.*
(John de Welton to Edmund Stonor, no. 10 [1377])
b. *Wretyn atte Wodefforde, the thorsdaye next after Seynt Luke daye.*
(Thomas Porchet to ?Thomas Stonor, no. 57 [before 1450])

Haste formula

- (48) a. *Esript a Dorchestre cest Lundy en graunt hast.*
(Nicholas Cowley to Edmund Stonor, no. 8, c. 1365)

- b. *Y-wrytyn atte London on Seynte Barnebe y Evyn yn all haste.*
(John Yeme to Thomas Stonor, no. 81 [?1466])

Endorsement formula

- (49) a. *A mez treshonnurés meistes William Worstede, John Longham, et Meistre Piers Shelton soit donné.*
(William Paston, no. 4, 1426)
- b. *To the ryght worschipful George Cele, merchand of the Stapull at Cales be bys delyuerd.*
(John Spencer, no. 11 [?1476])

Salutation formula

- (50) a. *A mon treshonore et tresreverent syr.*
(Nicholas Cowley to Sir Edmund Stonor, no. 8, c. 1365)¹⁰
- b. *Ryght reverent and worshyppful Syr.*
(Oliver Wittonstall to Thomas Stonor, no. 107, before 1470)
- (51) a. *Trescher et tresffiabile amy.*
(William of Wykeham to Edmund Stonor, no. 12, 1378)
- b. *Right trusty and entierly welbeloved frend.*
(Alys, Lady Sudeley to Thomas Stonor, no. 53, 1431)

Subscription formula

- (52) a. *Vostre Johan Stoke de Brisuyt.*
(John Stoke to Edmund Stonor, no. 36, c. 1380)
- b. *Yours, Wylliam Adam.*
(Wylliam Adam to George Cely, no. 166 [1482])
- (53) a. *le vostre Chapelayn Gregori, parsonne de Bourton.*
(Gregory, parson of Bourton, to Edmund Stonor, no. 26, c. 1380)
- b. *Your B[rother] W. Dalton.*
(William Dalton to George Cely, no. 138)

10 The preposition *à* in the salutation by Nicholas Cowley is a relic from the customary opening that named both sender and addressee, in which the Latin dative was rendered by a prepositional phrase in French.

Both the salutation and subscription of the French letters show a similar kind of variation to those in the Middle English letters, depending on the relationship between sender and addressee. Due to restrictions of space not all epithets can be provided, but the examples above show that the English epithets are likewise calques of French ones, although the order of conjoined epithets may be reversed (for more details on French epithets, see Häcker 2019: 104–107).¹¹

The fourteenth century saw a gradual shift from the Latinate opening of the letter, in which the sender appears in the salutation, to a separate subscription at the end of the letter, with both systems co-existing for some time. There are two early examples of a subscription documented in Tanqueray's edition, both dating from 1339, which precede other examples by some forty years. One is a letter by the Dean of York to unknown recipients, addressed in the salutation as *Tres chers amis* which has the following subscription:

- (54) Par William la Zouache, Dean d'Everwyke. 19 Decembre 1339.
 'By William la Zouache, Dean of York.'
 (Tanqueray, no. 146)

In the Stonor correspondence the earliest French letter with a subscription dates from the 1370s. The variation we find in the subscription of French letters can be described as follows:

- (i) *Par* followed by name and title
- (ii) *Par* followed by name
- (iii) *Par* followed by title
- (iv) Name
- (v) *Le/De vostre* followed by name
- (vi) *Vostre* followed by name
- (vii) *Trestout de vostre* followed by name

The origin of the phrases with *par* and *de* is clearly the construction *escrit par/de* 'written by', with an insertion of place and date between the preposition and the verb. The prepositions were, however, frequently retained in letters in which the place and date formula was omitted. *Per* in the English versions is a retention of the preposition *per*, which was used in Latin letters, but it was also

¹¹ Joined epithets are first documented in Latin; to my knowledge, the earliest documented example is *dilectissimo socio et precordiali amico* 'most trusty companion and dearest friend', which is found in the salutation of a sample letter in Albert of Samaria's *Praecepta dictaminum* (Haskins 1929: 175, letter 3).

used as a loanword in the *Cely letters* by writers who are unlikely to have had any knowledge of Latin.

The more restricted set of subscriptions may be partly due to the composition of the collections; the French letters of the Stonor collection and Tanqueray's edition contain fewer letters between family members, fewer letters from employees to masters, and no letters between fellow members of the same guild. But it is clear that the choice between the different types of subscription was also based on social relationships: in all subscriptions containing *le vostre* or *de vostre* the addressee is of a higher rank than the writer. This is clear from the salutation formula of the same letters, which is *Trecher(e) sire* or *treshonore sire*, or a combination of both epithets.

An additional factor that accounts for the more limited set of French salutations and subscriptions is the fact that early French letters were modelled on the Latin opening of letters, in which the names of both addressee and sender appear at the beginning of the letter, and which had no subscription. For example:

(55) *A nos tres noble prince e seignur Edward, Deu grace, roy de Engleterre, seignur d'Irlande e duc d'Aquitaigne, frere Jan, le prestre de Caunterbire, saluz en grant reverence.*

'To our very noble prince and lord, Edward, by the grace of God king of England, Lord of Irlande and Duke of Aquitaine brother John, the priest [that is, archbishop] at Canterbury [gives] greetings and great reverence.'

(John Peckham to King Edward I, Tanqueray, no. 23, 1280)

(56) *William de Valence, seignor de Penbroc, a sa chere compaigne et amie, saluz.*

'William of Valence, lord of Pembroke, to his dear companion and friend [gives] greetings.'

(William de Valence to his wife, Tanqueray, no. 2, 1267)

It appears that the order of the constituents in these salutations depends on formality and the social status of writer and addressee. The classical Latin order is used in informal letters, while that in which the addressee precedes the sender is used in letters in which the addressee is of a higher rank than the sender. In example (55) the difference in status is also emphasized by the writer modestly referring to himself as "priest at Canterbury" rather than "archbishop of Canterbury".

There was no straightforward transition from the Latin letter opening to the salutation which only contains the addressee. Rather it appears that the first simplification of the Latin system consisted of a reduction in which only the

object of the Latin construction is retained. In this type the salutation consists of *saluz* on its own, or a phrasal combination in which *saluz* is conjoined with a second noun such as *amistes*, *honurs*, or *reverence(s)*, that is, ‘friendship’, ‘honour’ or ‘reverence’. The earliest example of this, dating from 1323, is a letter by John de Hampton, sheriff of Gloucester, to John de Bromeshull, bailiff of Giftesgate, where the salutation consists only of *Saluz*, after which the purpose of the letter is introduced without any introductory phrase (Tanquerey, no. 130). A more elaborate version of this type appears in a letter to the Bishop of Ely:

(57) *Honors et totes reverences come a son tres honorable seignur.*

‘Honour and all the reverence as [is due] to his most worshipful lord’
(John Walewyn to John Hotham, bishop of Ely, Tanquerey no. 122)

This suggests that when the Latin construction of the letter opening, which could only be translated by prepositional phrases into French, became rather clumsy with multiple epithets and titles, it was not clear which elements of the construction should be discarded. In the period from the 1320s to the 1350s we find letters with salutations in which only the original object of the Latin construction was retained. Thus during this period three letter openings coexisted: (i) the traditional Latin format, consisting of [sender] à [addressee] *saluz*, as in *William de Valence, seignor de Penbroc, a sa chere compaigne et amie, saluz* (= [56] above), (ii) one in which only the original accusative object of the construction was retained: *Salutz* (the prior of Christ Church to Robert of Ely, Tanquerey no. 130, dated 1331), and (iii) one in which only the addressee was retained but without the preposition: *Cher sire et ami* (Hugh de Despenser to John d’Offord, Chancellor, Tanquerey no. 150, dated 1345–1349).¹² It is the last of these, which developed into the modern salutation, with a shift of possessives from the third person to the first, which reflects a reinterpretation of the original indirect object as vocative. The variation is not only well documented in Tanquerey’s edition, but also in the salutations in French letter writing manuals such as Sampson’s (see below).¹³

¹² Example (50) above is one of the rare cases in which *à* is retained.

¹³ Option (ii) has survived as an oral greeting in French (spelt *salut*).

6 The acquisition of letter formulae

6.1 Letter writing manuals and the teaching of formulae

The first documented letter writing manual in English that has come down to us, William Fullwood's *Enimie of Idlennesse*, first published in 1568, postdates the first Middle English letters by almost two hundred years.¹⁴ This raises the question how letter writers acquired a comparatively homogeneous system of English formulae, with major variations only in the salutations and subscriptions. The fact that the Middle English formulae are literal translations of French ones, that is, calques¹⁵, suggests that it might be profitable to take a closer look at what was available in terms of manuals in the two main languages of letter writing in this period, that is, French and Latin. There is indeed a considerable number of French and Latin manuals dating from the fourteenth century and earlier, with the earliest Latin treatise produced in England dating from 1207 (Robertson 1942: 9).

The greatest number of fourteenth-century treatises appeared under the name of Thomas Sampson (edited in Richardson 1942: 329–434). Sampson's manuals exist in various versions and even more manuscripts. I will restrict myself to an overview of those edited by Richardson. Some of the manuals, or formularies, as Richardson calls them, contained only French model letters (in manuscripts dating from c. 1355, c. 1365, c. 1371, c. 1383, and from c. 1385), some only in Latin (in manuscripts dating from c. 1365, c. 1371, c. 1377, c. 1391, c. 1410–1420, and c. 1435), while some contain letters in French and Latin (manuscripts dating from c. 1380, and c. 1410–1415). Sampson's work was copied by others, as can be seen in the prologue to his third treatise, in which he states that the new manual was necessitated by the fact that various people had circulated faulty copies of his earlier ones, mangling inter alia the orthography of his French and Latin texts (Richardson 1942: 334). The content of Sampson's teaching can be gleaned from the letters. In Richardson's no. 51 the disciplines taught to a pupil are specified as *modum scribendi, dictandi et computandi*, that is, 'writing, composition, and accounting'. On the basis of text internal evidence, Richardson describes Sampson's pupils as follows:

¹⁴ This is a century later than the first documented vernacular letter writing manuals in German (Rockinger 1863 [1961]: v).

¹⁵ The chronology clearly shows that the English ones are calques of French ones and not vice-versa.

What profession these boys had in view is shown by another of Sampson's prefaces, where he says that a treatise of his on conveyancing has been composed for the instruction of youths intending to enter the employment of lords and nobles and other men of worship, that they may render themselves more able and serviceable and, by the grace of Almighty God, promote the interest of their masters.

Richardson (1942: 336)

As Richardson (1942: 335) tells us, the learning of letter writing consisted of copying letters and learning them by heart.

In the following I present the content of one of the longer formularies (Richardson 1942: 417–430), entitled *Modus compositionis litterarum*, dated 1410–1415.¹⁶ This treatise provides headings for some, but not all, letters. It contains many letters which are given in both Latin and French versions, but in many cases the Latin version appears to be solely included for didactic reasons, as it would be unnatural to write letters to relatives in Latin. By contrast, the absence of a French version in the case of a fictitious letter from the pope to the king matches contemporary conventions, as members of the clergy employed Latin rather than French in their letters. It is also clear that Sampson's Oxford pupils would never have served as secretaries to the pope or king. In addition to stylistic practice, the inclusion of this letter may therefore possibly have served the purpose of emphasizing Sampson's importance as a teacher. That his treatises were not only teaching material, but also a means of advertising his services, is particularly noticeable in Sampson's treatise from c. 1396, which contains inter alia a Latin letter, entitled by Richardson 'A father writes to his son instructing him to place himself under Thomas Sampson' (Richardson no. 80). Some of the Latin manuals (specifically those dating from c. 1381 and that dating from c. 1383), contain predominantly letters dealing with university matters,

16 The treatise contains the following letters, which are given with Richardson's numbers; the headings in square brackets are Richardsons, while the information on the language in parentheses is mine:

83A *De patre filii ad eius magistrum causa informacionis* (L), 83B *Eadem in gallicis* (F), 84A *De filio ad patrem* (L), 85A [an alternative form to the preceding one] (L), 86A [another form, asking for money] (L), 84B [French version of no. 84A], 85B [French version of no. 85A], 86B [French version of no. 86A], 87A *De filio ad parentes* (L), 87B *Eadem in gallicis* (F), 88A *De filio ad matrem* (L), 88B *Eadem in gallice* (F), 89 *Littera de papa ad regem* (L), 90 *Responcio* (L), 91A *De puero ad auunculum* (L), 91B *Eadem in gallicis* (F), 92A *De scolare ad eius magistrum, consanguineum vel specialem amicum* (L), 92B *Eadem in gallicis* (F), 93A *De scolare ad germanum* (L), 93B *Eadem in gallicis* (F), 94A *Responcio* (L), 94B *Eadem in gallicis* (F), 95A *De scolare ad consanguineum* (L), 95B *In gallicis* (F).

while the treatise on *dictamen*, dating from 1435, deals mainly with ecclesiastical matters. This shows that different versions existed for different audiences.

In the following I will take a closer look at one of the sample letters, a letter from a son to his father, a type that was rare in the French letters of the Stonor correspondence and absent from Tanqueray's edition. I have chosen a letter from Sampson's manual dated c. 1383, which belongs to the same period as most of the French letters of the Stonor correspondence, but have omitted the non-formulaic part in which the purpose of the letter, a request for money, is stated, and have added a label in square brackets after each formula:

(58) *Treshonure sire et piere* [salutation formula]. *Je me recomanc a vostre tresreverente paternite en tant come je suy digne en totes reverences et honurs* [recommendation formula], *desirant souverainement vostre benisonne et qe vous soiez en saintee, qe pry a Dieu, par sa mercy, q'il la vous voille longement granter* [welfare formula]. *Et touchant mon estat, treshonure piere, savoir vous pleise q'a l'escrire du cest lettre jeu en saintee de corps, la mercy Dieu,* [health formula]...

Autre chose, treshonure piere, ne vous sai escrire a present [closing formula], *mais qe me voillez recomandre a ma treshonure dame et miere et saluer de par moy mez freres, soers et touz autres amys* [greeting to family]. *Et luy souverain piere Dieu mesmes vous octroie plusours jours et benurez* [benediction formula]. *Escript* [date and place formula]

'Right reverend Sir and father. I recommend myself as much as I am worthy to your right reverend fatherhood in all reverence and homage, desiring most your blessing and that you be in good health, which I pray God by his mercy will grant you for a long time. Regarding my well-being, most reverend father, if it please you to know that at the writing of this letter I am in good health of body, thanked be God, ...

Other matters, right reverend father, I do not know to write [i. e. there is nothing else to write] at the moment, but that you would recommend me to my right reverend lady my mother and give greetings from me to my brothers and sisters and all other friends. And God himself may give my sovereign father a long life and happiness. Written'

(Richardson, no. 59, dated 1383)

The letter lacks a subscription formula, which is evidently assumed to be known, and the date and place formula is truncated. In addition to the formulae described above, the letter contains greetings to family members and friends,

which appear to consist of two synonymous formulae, in which the intended recipient of greetings are inserted:¹⁷

(59) *me voillez recomandre a ma treshonure dame et miere.*

‘would you recommend me to my very worshipful dame and mother.’

(60) [voillez] *saluer de par moy mez freres, soers et touz autres amys.*

‘would you greet on my part my brothers, sisters and all other friends.’

Such greetings are rare in the Middle English correspondences, and greetings with conjoined predicates (as in the French sample letter) do not occur at all. The following greeting is one of the rare Middle English ones:

(61) *No more to yow at thys tyme, but þe Holy Gost haue yow in hys kepyng. Wretyn at Hadley þe Saterdag after Seynt John ys Day. And I be-seche yow hertyly recomande me to my master Alblaste.*

(R. Dollay to John Paston I, vol. 2, no. 497, 1454)

I besech you is customarily employed in requests. Its use in greetings derives from the context in which family members were originally mentioned in letters, which is the benediction. This can be seen in Margery Brews’s letter to her future husband John Paston III:

(62) *I recommend me vnto yowe full hertely, desyryng to here of yowr welefare, wherch I beseche Almyghty God long for to preserve vn-to hys plesure ...*

(Margery Brews to John Paston III, no. 425, February 1477)

The origin of the greeting formula employed by Dollay is thus a combination of elements from the recommendation formula and the benediction formula. This is evident from another greeting:

(63) *I mekely beseche your good fadyrhod that thys my bylle may recomaund me unto my good modyr yn my most umbyl vyse, mekely besechyng my good modyr of hir dayly blessing.*

(William Stonor to his father, Thomas Stonor, no. 127 [1473])

¹⁷ Due to their infrequency, these were not included in the list of formulae above.

The English greeting formulae, which differ structurally from the French one, show the strategies that were used if no literal translation of a French formula was readily available: the combination and modification of known formulae.

6.2 Translation and conventionalization of formulae

The transfer of letter formulae from French to English must have taken place via translations by bilinguals who were well versed in French letter formulae, that is, most likely private teachers such as Thomas Sampson or, possibly, their pupils. The lack of any surviving English manuals from the fifteenth century suggests that, if such manuals existed at all, their number must have been low, and certainly much lower than that of the French and Latin ones (see Robertson 1942 for an overview of the available formularies). This suggests that there was less need for them than for the French and Latin manuals. In a predominantly oral society, in which rote learning was dominant in the classroom, the memorization of a limited number of letter formulae in the pupils' mother tongue would have been an easy task, much less demanding than other subjects. But these formulae were also used by women, who dictated their own letters. One might speculate that in their letters the formulae were inserted by scribes. But the content of the letters of Margaret Paston and Margery Brews (later wife of John Paston III), to their husband and fiancé respectively clearly indicate that they were authored by the women themselves (Häcker 2011). This argument is supported by Margaret Paston's documented grasp of legal terminology, which she also must have acquired orally (Spedding 2008). If we accept the women as authors of their own letters, the acquisition of the formulae must have proceeded orally outside the classroom, as the education of women below the rank of nobility in the fourteenth century did not include reading and writing. This may seem unlikely from the perspective of present-day Western culture, but in households where letters were dictated and read aloud, females - and also males - were able to pick up recurring formulae orally. The acquisition of letter formulae is no different from the acquisition of formulae of every-day language that present-day as well as medieval children acquire(d) effortlessly from oral input, such as politeness formulae, greetings, songs, or rhymes. The conventionalization of the English letter formulae must have happened quickly and gone hand in hand with the shift to English as the language of the classroom,

which took place around 1400.¹⁸ As soon as a sufficient number of scribes wrote letters in English and employed the English formulae, their use must have spread fast among those who could write and those who could only use them in dictation. This does not mean that all the *ad hoc* translations of French formulae by different individuals were completely identical; this would be highly implausible. Rather more likely is a short period of competing formulae, followed by a conventionalization process, in which one form became dominant while others dropped out of use. In some cases French formulae continued in use after the shift to English. A case in point is *soit donné*, literally ‘be given’, which co-existed with *be delivered*, and its abbreviation *be dd.* in the family correspondences. By those who had no knowledge of French the phrase was borrowed holistically and in some cases divided into a sequence of English words:

(64) *To my inteirly beluffyd brother Jorg Cely, merchant at the Staple of Calles, so it don.*

(John Dalton to George Cely, no. 125, 1481)

The use of this loanphrase is on a par with the use of *per*, which is used as a Latin loanword in the Cely correspondence. The division of one French word into two English ones shows that the French words were not understood, but simply used because they were seen in letters and the meaning of the whole phrase was clear. The nonsensical division is comparable to the kind of folk etymology that turned *asparagus* into *sparrowgrass*.

18 This date is suggested by the use of French and English by William Paston I in his letters, which strongly indicates that he was taught in French and Latin rather than English. The evidence of William Paston I thus contradicts Trevisa’s well-known statement on the language shift in the classroom, which reads: *þys manere [i.e. teaching in French] was moche y-used tofore the furste moreyn and ys setþthe somdel ychaunged. For Iohan Cornwall, a mayster of gramere, chayngede þe lore in gramerscole and construccion of Freynsch into Englysch; and Richard Pencrych lurnede þat manere techyng of hym, and oper men of Pencrych, so þat now, þe 3er of oure Lord a þousond þre hondred foure score and fyue, of the secunde kyng Richard after þe Conquest nyne, in al þe gramerscoles of Engelond childern leueþ Frensch, and construeþ and lurneþ an Englysch.* (quoted from Barber 1993: 143)

‘This custom was much in use before the first plague [that is, the Black Death of 1349], and since then has somewhat changed. For John of Cornwall, a licensed teacher of grammar, changed the teaching in grammar school and the construing from French into English; and Richard Pencrich learnt that method of teaching from him, and other men from Pencrich, so that now, in the year of Our Lord 1385, in the ninth year of King Richard II, in all the grammar schools of England children are abandoning French, and are construing and learning in English.’ [Translation Barber’s]

7 Conclusion

The analysis of four family correspondences and a compilation of French letters written in England between the second half of the thirteenth century and the late fourteenth century shows a continuation between letters written in French and letters written in Middle English, which is in contrast to the almost complete break with the Old English letter writing conventions. The only Old English formula that survived is the salutation *I gret the/you well*, which drops, however, out of use in the fifteenth century. In contrast to Old English letters, Middle English ones contain numerous formulae which correspond to French ones that are not only documented in letters from the fourteenth and fifteenth centuries in England and on the Continent, but are also found in French letter writing manuals, which suggests that manuals were used to teach them.

The change from French to English as the medium of letter writing in late medieval England in all probability predates that of the first manual in English by some one hundred and fifty years, as there is no English manual documented before Fullwood's (1568) *Enimie of Idlenesse*. This suggests that French manuals were also used for teaching English letter writing and that the French formulae were translated by tutors in an *ad hoc* fashion. The homogeneity of the formulae used in English letters suggests a rapid conventionalization and spread. The fact that the same formulae were also used by illiterate authors of letters indicates that the formulae were not only taught but also passed on orally in households where letters were regularly dictated and read out aloud.

References

- Alberic of Montecassino [Alberico di Montecassino]. 2008. *Breviarium de dictamine*. Ed. Filippo Bognini. Florence: SISMELE.
- Albert of Samaria [Adalbertus Samaritanus]. 1961. *Praecepta dictaminum*. Ed. Franz-Josef Schmale. Weimar: Böhlau.
- Barber, Charles. 1993. *The English language: A historical introduction*. Cambridge: Cambridge University Press.
- Biber, Douglas. 2009. A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics* 14. 275–311.
- Boiarsky, Carolyn. 1992. Strategies for successful international communication: Using cultural conventions to inform your documents. in *Crossing frontiers: International conference on professional communication*, Santa Fe: IEED 1992. online version IEEE Xplore. DOI: 10.1109/IPCC.1992.673100: <https://ieeexplore.ieee.org/document/673100> (accessed 8 August 2020).

- Carpenter, Christine (ed.). 1996. *Kingford's Stonor letters and papers (1290-1483)*. Cambridge: Cambridge University Press.
- Constable, Giles (1976). *Letters and letter collections*. Turnhout: Brepols.
- Cowie, Anthony Paul. 1998. Introduction. In: Anthony Paul Cowie, *Phraseology: Theory, analysis, and applications*, 1–20. Oxford: Clarendon Press.
- Davis, Norman. 1965. The *Litera Troili* and English letters. *The Review of English Studies* 16. 233–244.
- Davis, Norman (ed.). 1971–1976. *Paston letters and papers of the fifteenth century*. 2 vols. Oxford: Clarendon Press.
- Elspaß, Stephan. 2012. Between linguistic creativity and formulaic restriction: Cross-linguistic perspectives on nineteenth-century lower class writers' private letters. In: Marina Dossena & Gabriella Del Lungo Camiciotti (eds.), *Letter writing in late modern Europe*, 45–64. Amsterdam & Philadelphia: Benjamins.
- Fulwood, William. 1568. *Enimie of idleness*. London: Bynneman.
- Häcker, Martina. 2011. French-English linguistic and cultural contact in medieval England: The evidence of letters. *Arbeiten aus Anglistik und Amerikanistik* 36. 133–160.
- Häcker, Martina. 2019. Kinship or friendship? The word *cousin* as a term of address for non-relatives in Middle English. *Journal of Historical Pragmatics* 20. 96–131.
- Hanham, Alison (ed.). 1975. *The Cely letters: 1472–1488*. Oxford: Oxford University Press.
- Harmer, Florence Elizabeth (ed.). 1989 [1952]. *Anglo-Saxon writs*. Manchester: Manchester University Press; 2nd edn. Stamford: Paul Watkins.
- Haskins, Charles Homer. 1929. *Studies in medieval culture*. New York: Ungar.
- Ingham, Richard. 2012. Middle English and Anglo-Norman in contact. *Bulletin de l'Association des Médiévistes Anglistes de l'Enseignement Supérieur* 81. 1–23.
- King, Tamyia. n. d. *Japanese business letter etiquette*. <https://smallbusiness.chron.com/japanese-business-letter-etiquette-61225.html> (accessed 8 August 2020).
- Kirby, Joan (ed.). 1996. *The Plumpton letters and papers*. Cambridge: Cambridge University Press.
- Laitinen, Lea & Nordlund Taru. 2012. Performing identities and interaction through epistolary formulae. In Marina Dossena & Gabriella Del Lungo Camiciotti (eds.), *Letter writing in late modern Europe*, 65–88. Amsterdam & Philadelphia: Benjamins.
- Lanham, Carol Dana (1975). *Salutatio formulas in Latin letters to 1200: Syntax, style, and theory*. München: Arbo-Gesellschaft.
- Murphy, James J. 1974. *Rhetoric in the Middle Ages: A history of rhetorical theory from Saint Augustine to the Renaissance*. Berkeley: University of California Press.
- Nevala, Minna. 2004. *Address in early English correspondence: Its forms and socio-pragmatic functions*. Helsinki: Société Néophilologique.
- Nevala, Minna. 2007. Inside and out: Address forms in 17th and 18th-century letters. In: Terttu Nevalainen and Sanna-Kaisa Tanskanen (eds.), *Letter writing*, 89–113. Amsterdam & Philadelphia: Benjamins.
- Nevalainen, Terttu & Helena Raumolin-Brunberg. 1995. Constraints on politeness: The pragmatics of address formulae in early English correspondence. In: Andreas H. Jucker (ed.), *Historical pragmatics*, 541–601. Amsterdam & Philadelphia: Benjamins.
- Nurmi, Arja & Päivi Pahta. 2004. Social stratification and patterns of code-switching in early English letters. *Multilingua* 23. 417–456.
- Prins, Anton Adriaan. 1952. *French influence in English phrasing*. Leiden: Leiden University Press.

- Putter, Ad & Judith Jefferson (eds). 2012. *Multilingualism in medieval Britain (c. 1066–1520): Sources and analysis*. Turnhout: Brepols.
- Raumolin-Brunberg, Helena. 1996. Forms of address in early English correspondence. In: Terttu Nevalainen & Helena Raumolin-Brunberg (eds.), *Sociolinguistics and language history: Studies based on the Corpus of early English correspondence*, 167–181. Amsterdam: Rodopi.
- Richardson, Henry Gerald. 1942. Letters of the Oxford dictatores. In: Herbert Edward Salters, William Abel Pantin & Henry Gerald Richardson (eds.), *Formularies which bear on the history of Oxford c. 1204–1420*, 329–450. Oxford: Oxford University Press.
- Robertson, Jean. 1942. *The art of letter writing: An essay on the handbooks published in England during the sixteenth and seventeenth centuries*. Liverpool: Liverpool University Press.
- Rockinger, Ludwig. 1961 [1863]. *Briefsteller und Formelbücher des elften bis vierzehnten Jahrhunderts*, vol. 1. Reprint. New York: Franklin.
- Rothwell, William. 2001. English and French in England after 1362. *English Studies* 82. 539–559.
- Sánchez Roura, Teresa. 2000. Convention vs. choice in securing the good-will of the reader: the Cely Letters. *SELIM [Journal of the Spanish Society of Middle English Language and Literature]* 10. 123–143.
- Schendl, Herbert. 2002. Code-choice and code-switching in some early fifteenth-century letters. In: Peter J. Lucas & Angela M. Lucas (eds.), *Middle English from tongue to text. Selected papers from the third international conference on Middle English: Language and text, held at Dublin, Ireland, 1–4 July 1999*, 247–262. Frankfurt am Main: Lang.
- Schendl, Herbert & Laura Wright (eds.). 2011. *Code-switching in early English*. Berlin & Boston: De Gruyter Mouton.
- Serjeantson, Mary S. 1935. *A history of foreign words in English*. London: Routledge & Kegan Paul.
- Short, Ian. 1980. Bilingualism in Anglo-Norman England. *Romance Philology* 33. 467–479.
- Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Spedding, Alison. 2008. *I shall send word in writing: Lexical choices and legal acumen in the letters of Margaret Paston*. *Medium Ævum* 77. 241–259.
- Stubbs, Michael. 2004. A quantitative approach to collocations. In: David John Allerton, Nadja Nesselhauf & Paul Skandera (eds.), *Phraseological units: Basic concepts and their applications*, 107–119. Basel: Schwabe.
- Suggett, Helen. 1946. The use of French in the later Middle Ages. The Alexander Prize Essay. *Transactions of the Royal Historical Society*. 4th ser. 28. 61–83.
- Tanqueray, Frédéric Joseph. 1916. *Recueil de lettres anglo-françaises (1265–1399)*. Paris: Champion.
- Tiisala, Seija. 2007. Power and politeness: Languages and salutation formulae in correspondence between Sweden and the German Hanse. In: Terttu Nevalainen & Sanna-Kaisa Taniskanen (eds.), *Letter writing*, 13–26. Amsterdam & Philadelphia: Benjamins.
- Trotter, David (ed.). 2000. *Multilingualism in later medieval Britain*. Woodbridge: Boydell & Brewer.
- Wray, Alison. 2002. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Wray, Alison. 2008. *Formulaic language: Pushing the boundaries*. Oxford: Oxford University Press.

Zuriñe Sanz-Villar

German-into-Basque translation of verbal patterns

Analysing trainee translators' outputs

Abstract: Authors such as Leiva Rojo (2013), Valero Cuadra (2015), and Albaladejo Martínez (2015) have analysed translations of phraseological units (PUs) in the field of translator training from varied perspectives. Whilst this area has not been explored in depth, it nonetheless arouses great interest (González Rey 2014). This paper aims to contribute to the field by providing an initial analysis of PU translations in texts translated from German into Basque by students of translation and interpreting at the University of the Basque Country (UPV/EHU). To this end, a small learner translation corpus (LTC) has been created. We pay special attention to PUs that are interesting from a translational point of view. This is either because they are challenging from a translation trainee's perspective or because they are interesting in terms of analysing the possible interference that the source language, German, as well as an intermediary language, Spanish (Sanz-Villar 2018), may exert on the translations. To summarise, the goals in this first attempt to analyse student translations of PUs in the language combination German-Basque are to create a small corpus and to manually extract phraseologisms, to analyse and discuss the translation of some of those PUs, and to think about the implications these findings may have in the translation classroom.

Keywords: phraseology, translation, learner translation corpus, German, Basque.

Note: The research group TRALIMA/ITZULIK (GIU 16/48, GIU 19/067) from the University of the Basque Country, UPV/EHU and the Basque Government consolidated research group (IT1209-19).

Zuriñe Sanz-Villar, University of the Basque Country, zurine.sanz@ehu.eus

<https://doi.org/10.1515/9783110520569-011>

1 Defining concepts

The terminology used by researchers when naming and defining their object of study varies depending on the approach or the way multi-word units (MWU) are analysed. The large range of approaches greatly increases the variety of terms used in the field of phraseology and it also influences what should be included within the field or excluded from it.

From a phraseological or linguistic point of view, some features (such as the distance between units making up the phraseologism) remain in the background, while others (the non-compositional character of the unit or the syntactic relationship between words) take on special relevance. According to Ebeling and Hasselgård, “[t]he *phraseological approach* refers to the Eastern European tradition, where the degree of non-compositionality, or idiom status, of multi-word units is one of the main concerns” (2015: 207).

Within frequency-based or statistical approaches, the distance (windows span) between elements making up a unit, as well as the frequency of co-occurrence, are of great importance. Here, idiomaticity is not regarded as essential and even word combinations that do not necessarily constitute a semantic unit are regarded as an object of study (Ebeling and Hasselgård 2015: 208). The following citation further explains this point:

In the new frequency-based approach to phraseology initiated by Sinclair (1987), the term *collocation* is used differently: It refers to statistically significant word co-occurrences, that is, lexical items occurring within a certain distance of the search item “with a greater frequency the law of averages would lead you to expect” (p. 70). Collocations are extracted with the help of statistical methods (such as MI) rather than linguistic criteria, and they may therefore include not only collocations in the traditional sense but also combinations such as *drink coffee* that would be categorized as free combinations in traditional phraseology.

Paquot and Granger (2012: 136)

Another broad approach to phraseology stores combinations of at least two words as well as “linguistic material of different size[s]” (Colson 2017: 17). According to Wray (2002: 9), a formulaic sequence is defined as “a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar”.

The constructionist approach has changed the way PUs are viewed. Within this perspective, all language sequences are constructions. As a result, the more traditional PUs are placed alongside very abstract constructions, such as the

passive: “According to this approach, PUs are in no fundamental way different from ordinary words o[r] from syntactic constructions” (Colson 2017: 17).

Gries suggests a definition of phraseologism that aims to unite the different approaches to phraseology:

A phraseologism is defined as the co-occurrence of a form or lemma of a lexical item and one or more additional linguistic elements of various kinds which functions as one semantic unit in a clause or sentence and whose frequency of co-occurrence is larger than expected on the basis of chance.

Gries (2008: 6)

This definition defends a broad conception of phraseology, while simultaneously delimiting the object of study. In order to define more rigorously the term ‘phraseologism’, Gries (2008) identifies six parameters or criteria: nature of elements, number of elements, frequency of occurrences, distance between elements, lexical and syntactic flexibility, semantic unity, and non-compositionality.

Ebeling and Ebeling (2013) use the parameters established by Gries to identify what they call ‘patterns’. These are “defined as a recurrent sequence of orthographic words that function as a semantic unit” (2013: 50). They insist on putting the focus on lexical rather than grammatical patterns (such as *back and forth, in so far as*).

The current paper represents a first attempt to identify patterns, phraseologisms, MWUs, or PUs (we will be using these terms interchangeably) which are interesting from a translational perspective. Therefore, the corpus is limited in size and phraseologisms are manually extracted. To identify PUs, we will be looking for “recurrent sequences of orthographic words that function as a semantic unit” (Ebeling and Ebeling 2013: 50), which may or may not be continuous. As presented in Ebeling and Ebeling, we can look at the translation counterparts to determine if a PU constitutes a semantic unit or not. “If a pattern in one language is repeatedly translated into a simplex in the other language, it may be inferred that we are dealing with a semantic unit, i.e. a pattern. Moreover, if we find that two patterns are regularly translated by each other, this is a strong indication that they mean the same and have semantic unity” (Ebeling and Ebeling 2013: 64). However, when translating productive multi-word patterns it is common to find a variety of outputs depending on how they have been combined with other words or how the slots have been filled.

In the translation classroom, it is important to enhance awareness of the aforementioned word combinations as recurrent and productive patterns of the source language. For this purpose, the use of corpora can be useful, because

(depending on corpus size) the patterns and their corresponding translations are often more extensively represented than in bilingual dictionaries. If students are aware of patterns, they will also develop critical thinking towards other tools, such as machine translation (MT), for which “[m]ultiword expressions may also be tricky and, as a consequence of the individual translation of the lexemes, the overall meaning is often lost” (Cerasani 2016: 281).

2 Translation of phraseological units in the translation classroom

Although Gries claims that “[n]owadays, the issues of identifying and classifying phraseologisms as well as integrating them into theoretical research and practical application has a much more profound influence on researchers and their agendas in many different sub-disciplines of linguistics as well as in language learning, acquisition, and teaching, natural language processing, etc.” (2008: 3), in the field of translation studies, research on the didactics of PUs in the translation classroom has been only very little.

As González Rey (2014) asserts, the teaching of PUs as well as the translation of PUs, are two fields that are attracting increasing interest among foreign language teachers and translators (2014: 9). The author adds that the peculiarities these types of units have in their own languages, their idiosyncrasies, make them especially difficult for language learners and translation students. In this paper, we want to bring the two fields (didactics and translation studies) together to analyse how students translate PUs in the translation classroom and see if there are, beyond the aforementioned idiosyncrasies, other factors that influence the translation process and product.

As indicated by Sardelli (2014: 200), the translation of phraseological units has already been analysed from different perspectives with different language combinations being taken into account. The didactics of phraseology in the second or foreign language classroom has also received scholarly attention. However, the branch of phraseology under analysis here, in which didactics, translation and phraseological competence come together, remains underexplored.

Serrano Lucas’ quote serves to confirm Sardelli’s statement:

La didáctica de la Fraseología representa un ámbito de investigación poco explorado hasta la fecha, y en concreto desde los estudios de traducción. [...] [P]ersiste cierta escasez de

obras de referencia y de materiales didácticos que traten específicamente el proceso de enseñanza-aprendizaje de estas unidades. Además, los que abordan este aspecto lo hacen desde la perspectiva de la enseñanza de segundas lenguas¹.

Serrano Lucas (2010: 198)

From a translation studies perspective, Leiva Rojo in his (2013) paper underlines the importance of phraseology in translation evaluation and advocates for the inclusion of phraseology in the assessment and review of translated texts, both in educative and professional contexts. Albaladejo Martínez (2015) suggests a didactic unit that focuses on the translation of specialised phraseological units (more specifically, collocations) by translation students in their very first year. Valero Cuadra (2015) uses journalistic texts related to economy and law to introduce translation and interpreting students to the field of phraseology. She focuses on the translation of collocations and how these are identified and understood, how equivalents in the target language which fit the target text are searched for and agreed upon.

3 German-into-Basque learner translation corpus

A learner translation corpus (LTC) is a parallel corpus consisting of translations made by students. As mentioned by Granger and Lefer, “[t]he idea of compiling corpora of student translations is not new. Among the forerunners were Uzar and Walinski (2001), Bowker and Bennison (2002), Floren (2006) and Kubler (2007). More recent projects include Štěpankova (2014), Kutuzov and Kunilovskaya (2014) and Wurm (2016)” (2018: 72).

In the framework of the Mellange (Multilingual eLearning in Language Engineering) project, a LTC was created with several source languages and student translations in different target languages. Although the project came to an end in 2007, the error annotation scheme (and other resources) can still be consulted on its website.² A new initiative, MUST, emerged in 2016. This involves 36

¹ The didactics of phraseology represents a field of research little explored to date, and concretely from translation studies [...] there are still few reference works and didactic materials that deal specifically with the teaching-learning process of such units. What is more, those who approach this topic do so from the point of view of second language teaching.

² <http://mellange.eila.univ-paris-diderot.fr/> (accessed 15 April 2020).

research teams from 16 different countries that collect and annotate students' translations under the same conditions and using the same tools.³

The present German-into-Basque LTC is intended to be an ad hoc corpus which will be created for specific purposes. The goal is not to obtain statistics based on a large amount of data, but rather to perform a qualitative analysis to:

- observe what mistakes are made and why. As indicated by Espunya, “errors identified by the course instructors are data worthy of systematic analysis as they inform us about the difficulties encountered by the trainee translators” (2013: 130);
- to see if the source language or a third language influences the trainee translators' output. Despite being translations from German into Basque, it has been observed (Sanz-Villar 2018) that Spanish is very present in German-into-Basque translating, either explicitly (due to the lack of direct linguistic resources between German and Basque) or indirectly through the translators' linguistic knowledge of Spanish;
- to examine the extent that MT systems have an influence when translating the patterns under analysis.

The collected translations are from students in the second term of their third year. German is their second foreign language and the command they have of the language varies from student to student. Most of the trainees had no prior knowledge of German before commencing translation and interpreting studies; i.e. some of the students' knowledge of the language is what they have acquired over a two-and-a-half year university career. Others have benefited from an Erasmus stay in a German speaking country. With regard to Basque (the target language), for some students it is their first language and the language they use in everyday life; for others, it is their second language and they are mostly accustomed to speaking Spanish on a daily basis.

The STs are non-specialised, journalistic texts extracted from online newspapers such as *Der Standard*, *Focus* and *Stern*. As can be observed in Table 1, translations from four students in a range of academic years (2014-15; 2015-16; 2017-18) have been collected per ST, with the exception of one text (T2 and T2_MT from Table 1) for which there are eight target texts (TTs) from students in

3 Together with other colleagues from the English and German Philology, Translation and Interpreting Department from the University of the Basque Country, we were recently admitted to the project and we started collecting data in the academic year 2019–2020.

two different academic years.⁴ As will be illustrated later, this is interesting due to “machine translation” variable as only students from 2017–18 were allowed to use MT systems. In total, the corpus is made up of 24 different texts consisting of 6,838 word tokens.

Tab. 1: Features of the corpus

Code ⁵	Academic year	Word tokens				
		STs	TT1	TT2	TT3	TT4
T1	2014-15	355	321	346	327	311
T2	2015-16	310	305	293	278	303
T2_MT	2017-18	264	243	269	248	258
T3_MT	2017-18	268	245	259	241	254
T4_MT	2017-18	287	273	281	287	276

All texts were translated in class, under exam conditions; i.e., students had two hours to complete the task of translating the text, with a computer, and they were allowed to use any resource they considered useful (MT systems were only allowed from 2017–18 onwards, as previously mentioned). Although DeepL is currently the most widespread MT tool amongst students, it had just been released at that time. Since it was relatively unknown, most of the students with access to MT technology used Google Translate.

It is interesting to point out the peculiarities of using such automatic tools with minority languages such as Basque. Due to the bad output obtained from MT systems when translating directly from German into Basque, students first use the tool to translate the texts from German into a familiar language (Spanish in most cases). To what extent the Spanish intermediary version is present in the translation process depends very much on the student. While some completely neglect the German source text and instead take the Spanish version as their ST to then produce the Basque translation, others keep both versions to hand.

⁴ The ST for T2 and T2_MT is the same, except the latter is one paragraph shorter. For the total word count, only the 310 of T2 were taken into account.

⁵ These codes will be used in the examples in Section 5. The addition of the abbreviation MT in the code means that students were allowed to use machine translation tools.

4 Methodology

The first step was to compile the corpus using a program called TAligner 3.0 (<http://www.ehu.eus/tralima/taligner.html>). This was developed within the TRALIMA/ITZULIK research group at the University of the Basque Country by the computer technician Iñaki Albisua. The main advantages of this program are its ability to combine all the steps necessary to create and query a corpus, and that there is no limit to the number of texts that can be aligned.

The program interface in Figure 1 clearly shows the necessary steps: *limpiar* (clean), *etiquetar* (tag), *alinear* (align), *consultar corpus* (query the corpus).



Fig. 1: Interface of TAligner 3.0

The texts are firstly cleaned. This means that all unnecessary blank spaces, extra tabs and lines, and other formatting issues are resolved. After cleaning, the texts are tagged at the sentence and paragraph level. This means that during alignment of STs and TTs (the subsequent step) the program can match sentences in the source and target texts that share the same tag. The user must make manual adjustments in order for the texts to be properly aligned. This is done using the options seen in the context menu in Figure 2: *combinar* (combine cells), *insertar blanco* (insert an empty cell), *dividir* (split a cell), *editar* (edit a cell), *eliminar* (delete a cell).



Fig. 2: Aligning the texts in TALigner 3.0

Once the sentences are aligned, they are uploaded to the corpus using the option *añadir a corpus* (add to the corpus). Here, the previously manually selected German PUs under analyses are extracted, together with their Basque counterparts. To this end, the search interface includes several options. One can search for specific words, parts of words or word combinations. Depending on the metadata included for each text, the program allows search parameters to be restricted or widened. For instance, a query can be restricted to texts translated by a specific translator. An example of a query can be observed in Figure 3. Here, the searched for German word *setzen* is highlighted with the Basque counterparts shown in the adjacent columns. The previous and subsequent sentences are also given, as had been requested using the option *antes y después* (before and after).

As explained in Section 1, the focus is on the translation of phraseological units and areas that may cause trouble to students in this field. Thus, it was decided to manually extract all PUs from the German STs: i.e. all “recurrent sequences of orthographic words that function as a semantic unit” (Ebeling and Ebeling 2013: 50), which may or may not be continuous.

The screenshot shows the 'Consultar corpus' interface with search filters and results. The search term 'setzen' is entered in the 'Texto a buscar 1' field. The results table is as follows:

Code	Text	Code	Text	Code	Text
201411KL (20)	Register Michael Krehbiel*	201411AZ (20)	Michael Krehbiel zuzendatutik honela da:	20141137 (20)	Honakoa da Michael Krehbiel zuzendatutik
201411KL (21)	Wir lieben nicht an den Monitoren, sondern setzen darauf, die Zuschauer schnell in einen Schauspieler zu ziehen, in dem sie sich wiederfinden.	201411AZ (21)	*Antzezirik ez dira pastalen inaututa geratu. Ikusleak deusarako grian murgil daitezte, zirkitorek duten edonon ikuslan ezaizten dira.	20141137 (21)	*Ez gara denbora gutxian pastalei so egoten, esenta bazuk, azkar irakurten esentura eramateko, non berriaz topo egiten di
201411KL (22)	Allen andere wäre verflucht bald langweilig, gibt er zu bedecken und fligt hinaus.	201411AZ (22)	Horrez gain, agian beritako gutia nahiko supergaria da*	20141137 (22)	Gaiozteriko gutia supergari kamara izan da beharbada* e da, eta honakoa gertatu da

Fig. 3: Querying the word *setzen* in the corpus

A wide variety of patterns were manually extracted from the STs: non-idiomatic or semi-idiomatic collocation-like patterns (*Sorge haben, heftige Debatte, eine Rolle spielen...*), binomials (*immer wieder*), polylexical conjunctions and prepositions (*ohne dass, mit Blick auf...*), or quite a few verbal multi-word patterns (*setzen auf, gehen um, liegen an, ankommen auf, gehören zu, zählen zu*, and so on). This paper will focus on these multi-word verbs consisting of a verb and a preposition as they are proportionally well represented in this small corpus and, most importantly, are interesting from a translational viewpoint (as will be demonstrated in Section 5). This topic is widely discussed within English literature, as is discussed by Chen in the following paragraph:

Among the many different types of multi-word expressions, multi-word verbs have been a popular topic due to their high frequency in English and the semantic and syntactic complexities. Multi-word verbs are complex verbs that consist of two or more lexical items (a verb followed continuously or discontinuously by another or other lexical items) and yet function to some extent like “a single verb lexically or syntactically” (Quirk, Greenbaum, Leech and Svartvik, 1985, p. 1150). They can be further divided into such sub-categories (Quirk et al., 1985) as phrasal verbs (a verb followed by an adverbial particle, *look up, turn out*), prepositional verbs (a verb followed by a preposition, *deal with, see through*), phrasal-prepositional verbs (a verb followed by an adverb and a preposition, *catch up with, look up to*), and verb-adjective combinations (a verb followed by an adjective, *put...straight, lay low*).

Chen (2017: 480)

The author points out that these MWE represent a challenge “for non-native users of English because they are semantically non-compositional, very often polysemous, and syntactically more flexible than other types of phraseological units (e.g. variation of particle positions and pronoun or noun insertions are allowed in phrasal verbs)” (Chen 2017: 481).

When identifying phrasemes in an L2 learner corpus of Italian, Konecny et al. (2016) decided to include in the analysis a subcategory called “syntagmatic verbs”. These are defined as “verbs consisting of two or more words [...] in which one of the words is usually an adverb indicating the (concrete or abstract) direction of an action” (2016: 537). Both syntagmatic verbs with compositional meanings (such as *pensarci sopra*, ‘to think about it’) and with idiomatic meanings (as in *passar su*, ‘to not care about something’) were identified, although the latter were rarely used in the corpus (2016: 538).

German prepositional verbs⁶ share some of the features mentioned by these authors, but have not been greatly studied within phraseology. Verbal patterns were selected as a subject of study because they are interesting from a translational perspective. This study observes to what extent the linguistic features of the analysed verbal patterns influence outputs produced by translation trainees. The goal here is not to analyse the schemes created by word combinations of a verb and a preposition (such as *halten* and *für*, which can be found in our corpus), how the slots are filled, or to find equivalences in the target language.

5 Translation analysis

Within the 12 analysed verbal MWUs, different tendencies were observed depending on the patterns’ characteristics, the type of tools used during the translation and the influence of similar patterns in other languages used by the students. The following paragraphs attempt to account for these findings.

5.1 Unproblematic verbal patterns

There are some verbal MWUs (such as *pochen auf*, *zählen zu*, *reagiren auf*, *anpassen an*) that seem largely unproblematic when looking at translations produced by trainee translators; i.e. they caused few mistakes in the outputs produced by the students. There are a variety of explanations for this; the verbal

⁶ Eisenberg (2004) identifies several verbal complements and one of them is represented by the *Präpositionalgruppen* (PrGr) or prepositional groups: “Typisch ist, daß ein Verb eine ganz bestimmte Präposition mit einem bestimmten Kasus regiert, wobei ihre (meist lokale) lexikalische Bedeutung weitgehend verlorengeht” [It is typical that a verb requires a very specific preposition with a specific case, whereby its (mostly local) lexical meaning is largely lost] (Eisenberg 2004: 62).

pattern is not polysemic, it is found in the consulted bilingual dictionaries (and is, thus, identified as a MWU with a semantic unit), the verbal pattern is widespread and it belongs to the students' vocabulary, the correct MT output helped the students, or a combination of factors effected the output produced by the trainee translators.

In the specific case presented in Table 2⁷, students found the verbal pattern *reagieren auf*, 'to react to', in bilingual dictionaries. Most of the students (TT1, TT2 and TT3) decided to use the Basque verb *erantzun*, 'to answer, respond', as an equivalent, although there is a formally more similar option both in Spanish (*reaccionar*) and in Basque (*erreakzionatu*). This is probably because the MT system's output uses the verb *responder*, 'to answer, respond', in Spanish.

However, the meaning of the sentence changes in TT1 and TT3, with respect to the ST. The problem in TT1 may be that the translation is too close to the MT's proposal.⁸ In TT3 the problem lies in misidentifying what the pronoun *er* refers to, as well as the meaning of the word *Leserbriefe*.

Tab. 2: Four student translations from the academic year 2017–18

T3_MT	TT1	TT2	TT3	TT4
Der Stuttgarter	Stuttgart-eko	Thienemann	Stuttgarteko	Thienemann
Thienemann-	Thienemann	Stuttgarteko	Thienemann	editorial
Verlag will diese	argitaletxeak	argitaletxeak	argitaletxeak	stuttgartarrak
diskriminierenden	Sorgin Txikiaren	Sorgin txikia	Sorgin Txikiaren	Sorgin Txikia
Begriffe in der	edizio berrian hitz	liburuaren	hurrengo berrar-	liburuan agertzen
nächsten Neu-	diskriminatzaileak	hurrengo argi-	gitalpenean	diren izendapen
ausgabe der	ezabatu nahi ditu.	talpenetatik	izendapen	diskriminatzaileak
kleinen Hexe	Horretarako,	kontzeptu	baztertzailak	argitaratuko den
herausnehmen.	editoreak	baztertzailak	ezabatuko ditu.	edizio berritik
Damit reagiert der	eskutitzei	ezabatuko ditu.	Hala argitaletxe-	kendu nahi ditu.
Verlag auch auf	erantzungo die,	Argitaletxeak	ak irakurle batek	Editorialaren
Leserbriefe, die er	zeinak azken	horrela	joandako ur-	erabakia
in den ver-	urteetan behin eta	erantzungo die	teetan behin eta	irakurleek azken

⁷ The first column of the tables contains the ST in German, and the following columns correspond to different Basque translations produced by students. The tables are further divided into two rows: The first includes the German ST and corresponding Basque translations; The second row contains the author's English translations of these German and Basque texts.

⁸ "Así, el editor responde a las cartas al editor, que ha recibido en los últimos años una y otra vez." (Google Translate). [Thus, the editor responds to letters to the editor, which he has repeatedly received in recent years.]

T3_MT	TT1	TT2	TT3	TT4
gangenen Jahren immer wieder mal bekommen hat.	berriro jaso dituen.	azken urteetan etengabe jasotako irakurleen gutunei.	berriro jasotako eskutitzari erantzungo dio.	urteetan behin eta berriro idatzi dituzten gutunen ondorioz izan da,
‘The Thienemann publishing house in Stuttgart wants to remove these discriminating words in the next edition of the Little Witch. This is in reaction to letters repeatedly received from readers in recent years.’	‘The Thienemann publishing house in Stuttgart wants to remove the discriminating words in the new edition of the Little Witch. For that, the editor will answer the letters which he has repeatedly received in recent years.’	‘The Thiennemann publishing house in Stuttgart will remove discriminating words in the next editions of the Little Witch. This is how the publishing house answers letters repeatedly received from readers in recent years.’	‘The Thienemann publishing house in Stuttgart will remove discriminating denominations in the next reissue of the Little Witch. This way, the publishing house answers a letter that a reader repeatedly received in recent years.’	‘The Thienemann publishing house in Stuttgart wants to remove discriminating denominations in new edition that will be published. The publishing house’s decision is due to the letters repeatedly written by readers in recent years.’

Be that as it may, the verbal pattern’s translation may have been influenced by the MT output, and thus the choices of students with access to MT systems are in general very similar to each other.⁹ An exception is TT4, which deviates from the structure of the original text and attempts to capture the meaning of the sentence. This text was only translated by students from the academic year 2017–18 (i.e. students with access to MT technology). In this case, it would also have been interesting to see how students without access to MT technology would have managed.

5.2 Problematic verbal patterns among students not using MT technology

The fact that students are unsure about their command of a foreign language forces them to rely too much on resources such as dictionaries or MT systems.

⁹ The fact that outputs from students using MT technology are very similar to each other is not an isolated case within this corpus.

Present students often use indirect tools (e.g. first German-into-Spanish followed by Spanish-into-Basque dictionaries). This makes it more difficult for them to view the text as a whole and they instead operate on a word-to-word level.

The verbal multi-word *beitragen zu*, ‘contribute to’, is found up in the dictionary and the Spanish equivalents are given as *ayudar* and *contribuir*. Probably due to the influence of these equivalents, some of the students (TT3¹⁰ and TT4 in Table 3) chose the direct equivalent word in Basque, *lagundu*. TT1 and TT2, on the other hand, reformulated the sentence and deviated more from the ST, thus obtaining a more natural output.

Tab. 3: Four student translations from the academic year 2015–16

T2	TT1	TT2	TT3	TT4
Das habe auch dazu beigetragen , dass in bestimmten Situationen selbst Akademiker in einer Art “Ghettoslang” kommunizierten.	Horrek eragin du baita akad- emikoek ere, egoera jakin batzuetan, “Ghettoko hizkera”ren antzeko zerbait erabiltzea.	Horrek ere zerikusia izan du unibertsitate bezalako inguru formalagoetan ere kale-hizkera erabiltzen hasi izanarekin.	Horrek egoera jakin batzuetan “ghettoko argo- tean” modu akademikoago batean komu- nikatzeari la- gundu zion.	Horretan lagundu duen beste faktore bat zera da: uni- bertsitateko kide- ak egoera jaki- netan arte ghettotan komu- nikatu izana.
‘This has also contributed to the fact that in certain situations even scholars communicate using a kind of “ghetto slang”.’	‘This has caused that even scholars, in certain situations, use something similar to a “ghetto language”.’	‘This has also to do with the fact that also in more formal situations, such as the university, they have started using street language.’	‘This has helped the fact that in certain situations they communicated in “ghetto slang” in a more academic way.’	‘Another factor that has helped in that is the following: university members having communicated in certain situations in ghetto art.’

¹⁰ In TT3 the problem is not just the verbal pattern, but the meaning of the whole sentence. TT3 suggests that in “ghetto slang” they communicated in a more academic way, while the ST says that even scholars communicate using a kind of “ghetto slang”. In TT4 the meaning of the ST is also badly reflected, since the German word *Art* (‘sort, type, kind’) is mistranslated with the formally similar but semantically different word *arte* (‘art’), a false friend. This same mistake has also been observed in other texts and with different students.

Among students that did not use an MT tool, the reason for incorrect rendering of the meaning of the verbal pattern may have been the polysemic character of the verb or misidentification of the verb and the preposition functioning as a semantic unit. This happened, for instance, with the multi-word unit *setzen auf*, meaning here ‘to focus on’.

Tab. 4: Four student translations from the academic year 2014–15

T1	TT1	TT2	TT3	TT4
Wir kleben nicht an den Monitoren, sondern setzen darauf , die Zuschauer schnell in einen Sehnsuchtsraum zu ziehen, in dem sie sich wiederfinden.	Antzezleak ez dira pantailari itsatsita geratu. lkusleak de-sirazko gelan murgil daitezten, aurkitzen duten edozein lekutan esetzen dira .	Ez gara denbora guztian pantailei so egoten, eserita baizik, ikuslea azkar irrikaren eremura eramateko, non berriz topo egiten duten.	Ez gaude pantailari pegatuta, baizik eta bertan arreta jarrita , ikuslea azkar grinez beteriko leku batera eramateko, non berriz aurkituko duen bere burua.	Ez gara pantailetara itsatsi, apustu egin dugu ikuslea irrika egoera batera azkar eramateko, non berriro aurkituko den.
‘Instead of being glued to the monitors, we focus on quickly pulling the audience into a place of desire where they can find themselves.’	‘The actors were not glued to the monitor. In order for the audience to immerse themselves in the room of desire, they sit anywhere they find.’	‘We are not all the time looking at the monitors, but sitting, to quickly bring the audience to the place of desire, where they meet again.’	‘We are not glued to the monitor, but paying attention there, to quickly bring the audience to a place full of passion, where they will find themselves again.’	‘We were not glued to the monitors, we bet to quickly bring the audience to a state of desire, where they can find themselves again.’

TT1 and TT2 focused on the meaning of the verb *setzen*, ‘to sit’, and so give the sentence a different meaning with respect to the ST. In TT3 the problem also lies with the verbal pattern, but it is unclear how the student came up with that solution. On the contrary, TT4 identified the verbal multi-word unit, but translated word for word the Spanish unit *apostar por*, ‘to opt for’, instead of using more typical patterns in the target language (for instance, *-en alde egin*).

In other instances the use of extra words to expand the verbal pattern may have caused confusion among students which resulted in a deviation of the ST's meaning. This is the case of the verbal pattern *gehen um*, 'to have to do with', from Table 5, which is expanded with the adverb *vielmehr*, 'rather'.

Tab. 5: Four student translations from the academic year 2014–15

T1	TT1	TT2	TT3	TT4
Ich hoffe, dass man die E-Mails schnell vergisst, denn es geht vielmehr um die Abhängigkeiten zweier Menschen von einander”.	“Espero dut jendea e-mailetak azkar ahaztea, batak bestearekiko duen dependentziak gehiago eskaintzen baitu ” gehitu zuen.	“Espero dut e-mailak azkar ahaztuko ditugula, askoz ere gehiago baitugu taularen gainean, bi pertsonak elkarrekiko duten mendekotasuna tarteko”.	“E-mailak azkar ahaztuko direla espero dut, askoz ere garrantzitsuagoa baita bi pertsonaien elkarrekiko mendekotasuna”.	“Espero dut jendeak e-mailak azkar ahaztea, izan ere antzezlanak gehiago ari da bi pertsonen elkarrekiko duten mendekotasunaz”.
‘I hope that the e-mails are quickly forgotten, as it has much more to do with the dependency of two people on each other.’	‘I hope that people quickly forget the e-mails, because the dependency of two people on each other has more to offer.’	‘I hope that we will quickly forget the e-mails, because we’ve got a lot more on the stage, the dependency of two people on each other, among other things.’	‘I hope that the e-mails are quickly forgotten, because it is much more important the dependency of the two characters on each other.’	‘I hope that people quickly forget the e-mails, because the theatre play is more about the dependency of two people on each other.’

TT1 and TT2 do not fully capture the meaning of the ST, probably, as previously mentioned, due to the adverb that serves to expand the verbal pattern. TT3 sticks more strictly to the ST and captures the proper meaning. In the last case (TT4), there is a case of interference from the intermediary language, since the student copies the structure of a Spanish pattern (*tratar de*). This is the second time the same student (see TT4 in Table 4) makes a word for word translation from a Spanish pattern. Thus, such analyses also serve to account for individual tendencies.

5.3 MT influence

In Guerberof's (2012: i) experiment, "translators have higher productivity and quality when using machine-translated output than when translating on their own". In the present study, texts (T2_MT and T2) translated by students from two different academic years (the first with access to MT systems and the latter without) may help to illustrate the positive and negative influence of these tools on student translators.

Tab. 6: A verbal pattern translated by students from different academic years

T2 and T2_MT	TT1	TT2	TT3	TT4
Vor allem letzteres hält die Berlinern ¹¹ durchaus für wahrscheinlich.	Berlineko ikertzailearentzat bi egoera horien artean gertagarriena azken egoera liteke. (T2)	Hizkuntzalari berlindarraren iritziz azkeneko bide honek ditu aukera gehien. (T2)	Egoera hori bereziki probablea da Berlinen geratzea. (T2_MT)	Azken hori berlindarren artean ematea oso posiblea da. (T2_MT)
'The researcher from Berlin considers the latter to be highly probable.'	'For the researcher from Berlin among the two situations the most probable would be the latter.'	'According to the linguist from Berlin, the last path is the one with more options.'	'This situation is more probable to happen in Berlin.'	'It is very probable that the latter happens among people from Berlin.'

Table 6 shows a clear example of a negative MT influence. All four students¹² from T2 (which were not allowed to use an MT system) managed to understand the pattern *für (wahrscheinlich) halten* (to consider to be probable) and render the meaning of the source sentence. On the contrary, the four students with access to MT tools (represented by TT3 and TT4 in Table 6) relied too much upon them. Google Translate (the MT system used by most the students at that time) fails to capture the meaning of the source text properly and the same error is observable through the different students' translations. In addition, in TT4 there is another case of lexical interference from Spanish. The use of the verb

¹¹ There was an error in the ST. It should be *Berlinerin* and not *Berlinern*, which makes reference to a female researcher mentioned in the article.

¹² This table shows translations of two students from each academic year, instead of the eight present in the corpus. This selection was made as such due to a lack of space.

eman in this context is considered to be an incorrect calque in Basque. As expected, this also demonstrates that interference from a language different to the ST (Spanish in this case) is not an individual occurrence.¹³

As illustrated in Table 7, there are also cases in which MT technology positively influenced the output. In some cases, students that used MT tools had less comprehension problems than those who did not use them. However, the risk of becoming stuck within the structure of the Spanish translation given by the MT system is high.

Tab. 7: A case in which the MT system may have helped the students to understand the meaning of the verbal pattern

T2 and T2_MT	TT1	TT2	TT3	TT4
“Wenn es stark auf Kürze ankommt , ist die Wahrscheinlichkeit groß, dass solche Strukturen eine Rolle spielen”, erklärt er.	Zuzendariak azaltzen jarraitzen duenez, “laburdurei dagokienez , oso litekeena da, egitura horien atzean arrazoi bat izatea, (T2)	“Labur idatzi behar denean , jendeak zenbait egitura erabiltzeko probabilitate handia dago”, azaldu du. (T2)	“Laburtasuna garrantzitsua denean , aukera handiagoa dago horrelako egiturek estrukturek paper bat jokatzeko” argitu du berak (T2_MT)	“Laburtzeak garrantzia duenean , horrelako egiturek rol bat jokatzeko duten probabilitatea handia da” argitzen du zuzendariak (T2_MT)
‘When it comes to being short, there is a high probability that such structures play a role’, he explains.’	‘As the director continues to explain, “as for the abbreviations, it is very probable that there is a reason behind those forms,’	‘When it comes to writing briefly, it is very probable that people use these types of forms.’	‘When it comes to brevity, it is more probable that such structures play a role.’	‘When it is important to write shortly, there is a high probability that such structures play a role.’

In the case of two students from T2, the meaning of the verbal pattern *ankommen auf* (something matters or is important) is captured incorrectly in their translations. This is exemplified in TT1 from Table 7. TT2, on the other hand, not

¹³ There are other examples of the negative influence that an MT system can have on translations. In cases where the ST is ambiguous, this ambiguity is maintained in the MT output, causing comprehension errors in the students’ translations.

having access to MT systems naturally expressed the meaning of the ST. TT3 and TT4, with access to MT tools, express the meaning of the German counterpart, but in TT3 the translation is too close to the Spanish MT translation given at that time (*Cuando se trata de la brevedad*, when it comes to brevity) and the result is unnatural in Basque.

6 Conclusions

The scarcity of research into the didactics of phraseology within translation studies creates opportunities, but also leads to uncertainty for the researcher. This uncertainty reveals itself when questioning which types of patterns are interesting, the methodology, the implications of the results in the translation classroom, etc. From a translational perspective, new tendencies regarding the broader definition of PUs are engaging. Students are trained to pay special attention to English phrasal verbs or idioms in general in any language, but not to German verbal multi-word units that may function as a semantic unit. The patterns I have analysed here (which from a more traditional viewpoint would not have been categorised as phraseologisms) can be interesting from the perspective of the translation trainee. Thus, it seems necessary to design didactic units where patterns are understood as “recurrent sequences of orthographic words that function as a semantic unit” (Ebeling and Ebeling 2013: 50). I agree with Espunya when she argues that “[b]ecause accurate knowledge of the lexicon is a fundamental factor for the success of students in translation exercises, teachers of translation courses pay close attention to their students’ shortcomings in order to design tasks to address them” (Espunya 2013: 130).

The corpus, and, thus, the analysed sample, has its limitations in terms of size. It would have been possible to include more STs, but the priority was to collect translations that were performed under the same circumstances (exam conditions in this case). Additionally, I have only included translations from students who filled out a consent form and agreed to their translations being used to create corpora for academic purposes. Regardless of these limitations, this corpus meets the goals of the study, which was to conduct a qualitative analysis to observe general trends and then analyse them at a deeper level within the framework of a larger project.

It has been observed that a variety of factors play an important role during the translation process. These include the linguistic features of the analysed patterns, source or third language interference, and the use of MT technology. The verbal patterns under analysis are sometimes polysemous (see Table 4), so

their syntactic flexibility allows for word insertions (see Table 5). As exemplified in the translation analysis, the characteristics mentioned may lead students to incorrectly understand the ST and therefore produce a mistranslation.

As shown in some examples (e.g. Tables 4 and 5), interference from the Spanish language is observable in student translations of verbal patterns. This is not particular to one individual student. Additionally, the use of MT systems may enhance this type of interference. The fear of deviating too much from the STs may lead students to stick too rigidly to the output produced by the MT.

With the development of MT systems, translation methods have changed enormously among trainee translators. With a language combination that includes a minority language, this indirectness has adopted new dimensions. Now, some students use German-into-Spanish MT outputs with some taking the Spanish translation as the ST. Others go further still and use Spanish-into-Basque MT systems and post-edit this final output. Since MT technology plays an important role in the translation process, it is necessary to conduct more systematic research on the use and influence of MT tools.

Although MT outputs have improved considerably, specific training is needed in order for students to be able to use these tools more critically. Sycz-Opoń and Gałuskina analysed trainee translations made using MT tools. They wanted to discover, among other things, how critical trainees are towards the outputs provided by MT tools. They concluded that “[t]he participants had serious problems with a critical evaluation of the MT output. They corrected lexical items that did not require correction (11% unnecessary corrections of the French MT output and 14% of English), yet they did not provide correction where it was necessary (20% omitted corrections for both French and English)” (Sycz-Opoń and Gałuskina 2017: 207).

Apart from the internal characteristics of patterns, language interference and MT systems, other extratextual factors may have influenced the product of translation. Features particular to individual students (e.g. competence in L1 and L3) were mentioned when describing corpus data in Section 4. I think it will be important for future studies to collect more detailed individual metadata and to consider these when analysing results.

References

- Albaladejo Martínez, Juan Antonio. 2015. Fraseología especializada en Traducción General. In Pedro Mogorrón Huerta & Fernando Navarro Domínguez (eds.), *Fraseología, didáctica y traducción*, 273–290. Bern: Peter Lang.
- Cerasani, Giovanni. 2016. On the correctness of machine translation: an English-Italian case study. In Tradumàtica Research Group (ed.), *Translators and Machine Translation: Book of presentations*, 280–300. Bellaterra: Department of Translation, Interpreting and East Asian Studies. http://ddd.uab.cat/pub/llobres/2016/168388/tramactra_a2016.pdf (accessed 2 May 2020).
- Chen, Meilin. 2017. Phraseology in English as an Academic Lingua Franca: A Corpus-Based Study of Prepositional Verbs in Writing by Chinese, American, and British Students. In F. Sigmund Topor (ed.), *Handbook of Research on Individualism and Identity in the Globalized Digital Age*, 478–499. Hershey: Information Science Reference. <http://dx.doi.org/10.4018/978-1-5225-0522-8.ch021> (accessed 15 April 2020).
- Colson, Jean-Pierre. 2017. The IdiomSearch experiment: Extracting Phraseology from a Probabilistic Network of Constructions. In Ruslan Mitkov (ed.), *Computational and Corpus-Based Phraseology*, 16–28. Cham, Switzerland: Springer. http://dx.doi.org/10.1007/978-3-319-69805-2_2 (accessed 15 April 2020).
- Ebeling, Jarle & Signe Ebeling. 2013. *Patterns in Contrast*. Amsterdam & Philadelphia: John Benjamins. <http://dx.doi.org/10.1075/scl.58> (accessed 15 April 2020).
- Ebeling, Signe & Hilde Hasselgård. 2015. Learner corpora and phraseology. In Sylviane Granger, Gaëtanelle Gilquin & Fanny Meunier (eds.), *The Cambridge Handbook of Learner Corpus Research*, 207–230. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781139649414.010> (accessed 15 April 2020).
- Eisenberg, Peter. 2004. *Der Satz. Grundriß der deutschen Grammatik*. Stuttgart & Weimar: J. B. Metzler.
- Espunya, Anna. 2013. Investigating lexical difficulties of learners in the error-annotated UPF learner translation corpus. In: Sylviane Granger & Gaëtanelle Gilquin (eds.), *Twenty Years of Learner Corpus Research. Looking Back, Moving Ahead*, 129–137. Louvain-la-Neuve: Presses universitaires de Louvain.
- Guerberof, Ana. 2012. *Productivity and quality in the post-editing of outputs from translation memories and machine translation*. Tarragona: Universitat Rovira i Virgili. PhD thesis. <https://www.tdx.cat/bitstream/handle/10803/90247/Guerberof%20Thesis%20Final.pdf?sequence=1&isAllowed=y> (accessed 15 April 2020).
- González Rey, M^a Isabel (ed.). 2014. *Didáctica y traducción de las unidades fraseológicas*. Santiago de Compostela: Universidade de Santiago de Compostela, Servizo de Publicacións e Intercambio Científico.
- Granger, Sylviane & Marie-Aude Lefer. 2018. MUST: A collaborative corpus collection initiative for translation teaching and research. In Sylviane Granger, Marie-Aude Lefer & Laura Aguiar de Souza Penha Marion (eds.), *Book of Abstracts. Using Corpora in Contrastive and Translation Studies Conference (5th edition)*, 72–73. Louvain-la-Neuve: Centre for English Corpus Linguistics/Université catholique de Louvain.
- Gries, Stefan Th. 2008. Phraseology and linguistic theory: A brief survey. In Granger, Sylviane & Fanny Meunier (eds.), *Phraseology: An interdisciplinary perspective*, 3–25. Amsterdam

- & Philadelphia: John Benjamins. <http://dx.doi.org/10.1075/z.139.06gri> (accessed 15 April 2020).
- Konecny, Christine, Andrea Abel, Erica Autelli & Lorenzo Zanasi. 2016. Identification and Classification of Phrasemes in an L2 Learner Corpus of Italian. In Gloria Corpas Pastor (ed.), *Computerised and Corpus-based Approaches to Phraseology*, 533–542. Geneva: Editions Tradulex.
- Leiva Rojo, Jorge. 2013. La traducción de unidades fraseológicas (alemán-español/español-alemán) como parámetro para la evaluación y revisión de traducciones. In Carmen Mellado Blanco (coord.), Patricia Buján, Nely M. Iglesias, M. Carmen Losada & Ana Mansilla (eds.), *La fraseología del alemán y el español: lexicografía y traducción*, 31–42. Munich: Peniöpe.
- Paquot, Magaly & Sylviane Granger. 2012. Formulaic Language in Learner Corpora. *Annual Review of Applied Linguistics* 32. 130–149. <http://dx.doi.org/10.1017/S0267190512000098> (accessed 15 April 2020).
- Sanz-Villar, Zuriñe. 2018. Interference and the Translation of Phraseological Units in a Parallel and Multilingual Corpus. *Meta* 63 (1). 72–93. <http://dx.doi.org/10.7202/1050515ar> (accessed 15 April 2020).
- Sardelli, Maria Antonella. 2014. La fraseología en las clases de traducción: aproximación metodológica y aplicaciones prácticas. In M.^a Isabel González Rey (ed.), *Didáctica y traducción de las unidades fraseológicas*, 199–212. Santiago de Compostela: Universidade de Santiago de Compostela, Servizo de Publicacións e Intercambio Científico.
- Serrano Lucas, Lucía Clara. 2010. Metodología para la enseñanza de la fraseología en traducción: la ficha fraseológica como tarea final. *Paremia* 19. 197–206.
- Sycz-Opoń, Joanna & Ksenia Gałuska. 2017. Machine Translation in the Hand of Trainee Translators – An empirical Study. *Studies in Logic, Grammar and Rhetoric* 49 (62). 195–212. <http://dx.doi.org/10.1515/slgr-2017-0012> (accessed 15 April 2020).
- Valero Cuadra, Pino. 2015. Fraseologismos en el aula de traducción general. In Pedro Mogorrrón Huerta & Fernando Navarro Domínguez (eds.), *Fraseología, didáctica y traducción*, 309–320. Bern: Peter Lang.
- Wray, Alison. 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511519772> (accessed 15 April 2020).

Lena Stutz, Rita Finkbeiner

Veni, vidi, veggio

A contrastive corpus linguistic analysis of the phraseological construction *Veni, vidi, X* and its German equivalent *X kam, sah und Y*

Abstract: This paper contributes to the growing research on lexically partly fixed and partly open phraseological constructions that allow for the productive formation of new instances. A crucial question is how one can determine the degree to which a given phraseological construction serves as a productive pattern for the formation of new instances. Stumpf (2016) suggested drawing a distinction, on empirical grounds, between modificatory patterns that allow for occasional substitution of a lexical position, and true phraseological schemata whose open slots are varyingly filled on a regular basis. Following this proposal, we examine in this paper the filler potential of the German phraseological construction [*X kam, sah und Y*] ‘*X came, saw, and Y*’, carrying out slot analyses with the open source corpus linguistic tool *Lexpan* (Steyer and Brunner 2014; Steyer 2018). Contrasting the construction with its Latin equivalent [*veni, vidi, X*], we show that while both constructions are productively used in German and allow for a variety of fillers, [*X kam, sah und Y*] has the status of a true phraseological schema, while [*veni, vidi, X*] is to be classified as a modificatory pattern.

Keywords: phraseological construction, phraseological schema, modificatory pattern, productivity.

1 Introduction

In recent years, phraseological constructions, i.e. lexically partly fixed and partly open constructions that allow for the productive formation of new instances, have gained increasing attention in Construction Grammar and phraseology. As lexically semi-fixed complex constructions, they are situated at the

Lena Stutz, Heinrich Heine University Düsseldorf, lstutz@phil.hhu.de

Rita Finkbeiner, Johannes Gutenberg-University Mainz, finkbeiner@uni-mainz.de

<https://doi.org/10.1515/9783110520569-012>

much-debated boundary between grammar and lexicon.¹ In fact, it is this class of constructions – also known as *Phraseoschablonen* (Fleischer 1982), *formal or schematic idioms* (Fillmore, Kay, and O'Connor 1988), or *constructional idioms* (Booij 2002; Taylor 2012) – that are at the heart of the increasing convergence of the two research traditions. While Construction Grammar and phraseology have developed rather independently of each other over many years, it has become more and more evident that the two strands of research have a lot to offer to each other (cf. Gries 2008: 14–15; cf. also Finkbeiner 2008a; Dobrovolskij 2011, 2018; Ziem 2018). While Construction Grammar can benefit from a long tradition of descriptive and classificatory work in phraseology, in the opposite direction, phraseology can profit from the elaborate theoretical apparatus and usage-based methodology of Construction Grammar (cf. Gries 2008: 15).

The increasing convergence of Construction Grammar and phraseology has led to a growing number of studies on a variety of phraseological constructions, both from language-specific and contrastive perspectives.² The main task of a descriptive approach to phraseological constructions that is inspired by Construction Grammarian concepts is to provide an exact definition of their specific structural and semantic constraints. In particular, what needs to be specified for each phraseological construction are the structural and semantic properties that constrain the set of potential lexical fillers that may be inserted into the open slots of the construction. It is clear that an adequate linguistic description of the properties of potential fillers calls for solid empirical analysis.

Corpus linguistic approaches to phraseological constructions have made headway recently in the development of tailor-made methods for the automatic analysis of open slot positions (cf. Bubenhofer 2015). Steyer and colleagues (Steyer and Brunner 2014; Steyer 2018) developed the freely available analytical tool *Lexpan*³ that imports the results of corpus searches on a certain phraseological construction and allows for clustering, computing and analysis of the filler items in the different instances. Tools like *Lexpan* have made it possible to analyse in more detail the syntactic, semantic and pragmatic restrictions of phraseological constructions (cf., e.g., Finkbeiner in press) and to make more fine-grained distinctions between different degrees of schematicity in phraseological constructions (e.g., Stumpf 2016). Stumpf (2016), in a study using the *Lexpan* tool, argues for a distinction between modificatory patterns (*Modifikationsmus-*

1 C.f. e.g., Jackendoff 2008; Jacobs 2008; Engelberg, Holler, and Prost 2011; Finkbeiner 2017.

2 Cf. the recent collections in Steyer (ed.) 2018; Ziem (ed.) 2018; Mellado Blanco, Mollica, and Schafroth (eds.) in press.

3 *Lexical Pattern Analyzer*, <http://uwv.ids-mannheim.de/lexpan/>

ter) that allow for occasional substitution of a lexical position, and true phraseological schemata (*Modellbildungen*) with open slots that are regularly filled with a variety of lexical items. In an attempt to distinguish between the two on empirical grounds, Stumpf (2016: 337) suggests, as a rough guide, that if at least 50% of the attestations of the construction in the corpus are instances of the phraseme in its lexicalised form, one is dealing with a case of modificatory pattern. If less than 50% of the attestations are instances of the phraseme in its lexicalised form, one is dealing with a phraseological schema. This also implies the assumption of a potential diachronic path from a fully lexicalised phraseme via (occasional) phraseme modification to a full-fledged phraseological schema.

Although the 50% benchmark suggested by Stumpf (2016) seems, unquestionably, to be an arbitrarily drawn distinction, and surely requires some further (also theoretical) discussion, it nevertheless offers an initial, viable guideline for the corpus-based determination of different degrees of productivity of phraseological constructions. It is clear that the question is disputable whether 50% of attestations is an adequate value; it is also clear that the problem of how to address the various cases with fluent transitions between pattern types is not yet satisfactorily solved. But only the testing of this benchmark within the context of numerous analyses of phraseological constructions will show its range and confirm or disprove the need for it to be adapted. In this sense, we take the 50% benchmark as a ‘pragmatic makeshift solution’ for the moment.

The present study joins the existing research on phraseological constructions in general and sets out to test Stumpf’s (2016) proposal in particular, focusing on a well-known pattern in German that is widely used in a variety of text and discourse types, but that nonetheless has not been subject to comprehensive linguistic analysis so far,⁴ namely the construction [X *kam, sah und Y*], which goes back to the famous Caesar quote *veni, vidi, vici* and its German translation *Ich kam, sah und siegte* ‘I came, saw, and conquered’. The overall aim of the study is to determine the degree of schematicity of [X *kam, sah und Y*] with the help of extensive automatic slot analyses based on *DeReKo*⁵ corpus data. While both the German and the Latin construction are frequently used in German, our contrastive analysis of [*veni, vidi, X*] and [X *kam, sah und Y*] indicates

4 Mieder (2013) compiles a sample of 65 (modified) attestations of the German and Latin equivalents, occasionally found in poems, short prose, cartoons, graffiti and advertising, though without providing a systematic, in-depth study of the modificatory processes and the degree of productivity of the patterns.

5 *DeReKo* = *Deutsches Referenzkorpus*, Institut für deutsche Sprache, Mannheim. The *DeReKo* archive of *DeReKo* comprises roughly 9,5 billion words.

that there are clear differences in the productive utilisation of the two patterns. Following Stumpf's (2016) distinction, it will therefore be suggested that [*veni, vidi, X*] is a modificatory pattern, while [*X kam, sah und Y*] is a true phraseological schema.

The structure of the paper is as follows. We will first take a look at the basic properties and constraints of the Latin construction as attested in *DeReKo* (Section 2). It will turn out that the Latin dictum *veni, vidi, vici* allows for modification of the third verb position to some extent, the potential substitutes being required to fulfil mainly phonological restrictions. At the same time, the corpus analysis reveals that the majority of instances of the construction are in fact instances of the original (lexicalised) quote. In a second step, we examine the German construction [*X, kam, sah und Y*] (Section 3). It turns out that the German construction exhibits two open slots, i.e. the subject position and the third verb position. In contrast to the Latin construction, the potential fillers on the verb slot are restricted as to their syntactic category and semantic class, but less so as to their phonological properties. Moreover, the corpus analysis reveals that the translated quote in the first person singular *Ich kam, sah und siegte* is not the instance that is dominantly attested in the corpus. Rather, the majority of instances are productive instances of [*X kam, sah und Y*] in the third person singular. Section 4 summarises the findings.

2 *Veni, vidi, X*

Veni, vidi, vici, the famous Caesar quote after the rapid victory of his troops over King Pharnakes II. in the year 47 BC, is still one of the most well-known quotes used in its Latin wording in German (cf. Büchmann [1864] 2014: 386). Because of its high degree of familiarity, it has today the status of a so-called winged word, or dictum, and as such it is part of the inventory of phraseological multiword units in German. However, as a quick look into the world of advertisement reveals, *veni, vidi, vici* does not only occur in its fixed, phraseological form but is also frequently exploited as a substitutive pattern, cf. (1)–(3).

- (1) *Veni. Vidi. Codi.*
(Apple T-Shirt at Worldwide Developers Conference 2006)
- (2) *Veni – Vidi – Wiici.*
(Advertisement for Nintendo game console *Wii*, 2006)

(3) *Veni – Vidi – Vista.*(Introduction of the Microsoft operation system *Vista*, 2007)

This raises the question as to whether such instances are occasional modifications of the original dictum employed by speakers in order to reach certain communicative effects, or whether the original dictum has already achieved the status of a full-fledged phraseological schema. We will first examine this question in a corpus-based analysis and then compare the Latin dictum to its German equivalent. To begin with, we take a look at the characteristic formal properties of the original dictum.

In a first approximation, *veni, vidi, vici* can be characterised syntactically as a triad of the first-person singular present perfect active indicative verb forms *veni* ‘I came’, *vidi* ‘I saw’ and *vici* ‘I conquered’, which are asyndetically coordinated. As Latin is a pro-drop language, however, it is more apt to say that this is not a coordination of three finite verbs but a coordination of three main clauses with a null-instantiated pronominal subject. The irreversible ordering of the three clauses is not arbitrary, but seems to be motivated by a more general principle of enumeration, according to which less important precedes more important. From a cognitive perspective, coming can be seen as a prerequisite of seeing, and seeing as a prerequisite of conquering. As is often the case in coordinate structures, the mere additive aspect that is encoded in the semantics of coordination is further enriched pragmatically. The process at work here can be regarded as an inference triggered by Grice’s (1975: 46) fourth submaxim of manner (‘Be orderly’), resulting in a reading according to which the three actions reported also follow a temporal ordering, in the sense that the coming event preceded the seeing event, which in turn preceded the conquering event. Finally, on the phonological level, the dictum is characterised by specific prosodic properties which contribute to its mnemotechnic effects (cf. Agricola [1962] 1977: 30). More specifically, the three elements each form a trochee, relating them on the metrical level, and they exploit both alliteration (in the initial [v]) and end rhyme (in the final [i]), relating them on the rhyme level.

Turning to the corpus analysis, we first take a look at the proportions of instances of the canonical dictum and modified instances in the corpus. In principle, one might expect that all three verbal positions (V_1 , V_2 , or V_3) should equally allow for substitution. However, the analysis reveals that it is predominantly the third verb position (V_3) that undergoes substitution in actual usage. This is shown in Table 1, which summarises the results of automatic slot analyses for the three potential patterns [V_1 , *vidi*, *vici*], [*veni*, V_2 , *vici*] and [*veni*, *vidi*, V_3].

Tab. 1: Slot analysis results for [V₁, *vidi*, *vici*], [*veni*, V₂, *vici*] and [*veni*, *vidi*, V₃]

	V ₁ , <i>vidi</i> , <i>vici</i>	<i>veni</i> , V ₂ , <i>vici</i>	<i>veni</i> , <i>vidi</i> , V ₃
DeReKo search query (number of hits)	\$vidi /+w3 \$vici (486)	\$Veni /+w5 \$vici (489)	\$Veni /+w3 \$vidi (671)
Lexpan search pattern (number of hits imported by Lexpan)	(# Vidi vidi Vici vici) 469 (≅100%)	(Veni veni # Vici vici) 464 (≅100%)	(Veni veni Vidi vidi #) 642 (≅100%)
number of hits with the filler item...	... <i>Veni</i> / <i>veni</i> : 447 (95.31%)	... <i>Vidi</i> / <i>vidi</i> : 440 (94.83%)	... <i>Vici</i> / <i>vici</i> : 443 (69.01%)
number of hits with a different filler item	22 (4.69%)	24 (5.17%)	199 (30.99%)

Whereas substitution on V₁ and V₂ each occurs in only about 5% of the cases, substitution on V₃ occurs in about 31% of the cases. That V₃ is the most productive position is also indicated by the fact that among the 642 instances with variation on V₃, *Lexpan* identifies 106 different types. Furthermore, one finds a great number of hapaxes on V₃, i.e. instances that occur only once. Of the 106 different instances, 73 (69%) are hapaxes. The frequent occurrence of hapaxes can be taken as a major indicator of the productivity of a pattern (cf. Plag 2006: 542–544; Booij 2002: 52; Hein 2015: 474). Thus, in terms of productivity, while all three verb slots in principle are available for productive use, only V₃ is used productively in the sense of being not only available, but also profitable to a considerable degree (cf. Bauer 2001: 209–211; Plag 2006: 539/553; Finkbeiner 2008b: 392/401–404). Still, while variation on V₃ does occur regularly, it does not exceed the 50% benchmark suggested by Stumpf (2016): in total, the number of instances of the lexicalised dictum *veni, vidi, vici* (1,330 attestations, about 84%) clearly exceeds the number of instances that are modified in one way or another (245 attestations, about 16%).

From a cognitive perspective, it is plausible to assume that in order for the modification to be comprehensible, recall and recognition of the frozen pattern must be secured. If the initial elements remain unmodified, the pattern may be easily recalled and the modification in the last element may be processed without difficulty. However, if the initial elements are substituted, recall of the pattern will be hindered and comprehension will be more difficult (cf. also Ptashnyk 2009: 84). This may explain the higher profitability of V₃ compared to V₁ and V₂. What is more, the climactic ordering of the three slots makes it highly expected that the climax, i.e. V₃, will be preferably substituted in order to

achieve special communicative effects. This is because the substitute on V_3 , in a sense, will inherit the climax semantics from the basic pattern, and thus receive a special communicative weight.

Let us now take a closer look at the different filler items on V_3 . Table 2 provides the 40 first filler items for V_3 as provided by *Lexpan* (sorted by frequency).

Tab. 2: First 40 of 106 different V_3 filler items in the Latin construction, sorted by frequency

Filler	Number	Percentage	Filler	Number	Percentage
<i>vici</i>	322	50.16	<i>viagri</i>	3	0.47
<i>Vici</i>	121	18.85	<i>Abii</i>	2	0.31
<i>Veekend</i>	14	2.18	<i>Spezi</i>	2	0.31
<i>Canti</i>	13	2.02	<i>Vegil</i>	2	0.31
<i>Vicious</i>	11	1.71	<i>Vice</i>	2	0.31
<i>Verdi</i>	10	1.56	<i>Vicki</i>	2	0.31
<i>Visa</i>	5	0.78	<i>Vista</i>	2	0.31
<i>vino</i>	5	0.78	<i>Wiki</i>	2	0.31
<i>Vegi</i>	4	0.62	<i>abi</i>	2	0.31
<i>Vichy</i>	4	0.62	<i>ici</i>	2	0.31
<i>Viki</i>	4	0.62	<i>und vici</i>	2	0.31
<i>veto</i>	4	0.62	<i>vegi</i>	2	0.31
<i>Abi</i>	3	0.47	<i>video</i>	2	0.31
<i>Fidschi</i>	3	0.47	<i>Berlusconi</i>	1	0.16
<i>Video</i>	3	0.47	<i>Celebravi</i>	1	0.16
<i>Vinum</i>	3	0.47	<i>Da Vinci</i>	1	0.16
<i>fick di</i>	3	0.47	<i>Deus vixit</i>	1	0.16
<i>inflammavi</i>	3	0.47	<i>Drinki</i>	1	0.16
<i>non vici</i>	3	0.47	<i>Fiji</i>	1	0.16
<i>sorry</i>	3	0.47	<i>Gaudium</i>	1	0.16

If we look at the fillers that are most frequently used beyond *vici/Vici*, it becomes clear from the table that phonological restrictions seem to play a major role in filler selection. As Table 3 reveals, the majority of fillers (61.3%) exhibit an initial labiodental fricative ([v], [f]) or bilabial approximant ([w]) in the onset of the first syllable. To a large extent, these are followed by a tense [i]-nucleus vowel (e.g., *Vichy*, *Video*, *Visa*, *vino*, *Viva*, *Wiese*) or its lax variant [ɪ] (e.g., *Vicious*, *Vista*, *fick di*, *Fiji*, *Whiskey*). Thus, with respect to their first syllable, these

fillers are phonologically identical ([vi]) or highly similar ([fi], [wi], [vi], [fi], [wi]) to the original filler *vici*. Furthermore, a considerable number of fillers also display sound similarity with the last syllable of the original filler *vici*. In 63.2% of the cases, the last syllable's nucleus is also a tense [i]-nucleus as in *vici*, often preceded by a [ʃ], [tʃ], [ts] or [dʒ]-onset (e.g., *Vichy*, *ici*, *Ricci*, *Spezi*, *Fidschi*, *arrivederci*, *Veggie*).

Tab. 3: Phonological qualities of the 106 different V₃ filler items in [*veni*, *vidi*, V₃]

Phonological qualities of the V ₃ filler items	Number	Percentage (106 fillers in total $\cong 100\%$)	Examples
Identical or highly similar to the initial sound of <i>vici</i> , based on ...	65	61.3	...
Initial sound			
1. a voiced labiodental fricative [v]	57	53.8	<i>Vichy</i> , <i>vicious</i> , <i>Vista</i> , <i>veggie</i> , <i>Verdi</i> , <i>Veto</i> , <i>vino</i> , <i>Vintage</i> , <i>Weserdeich</i> , ...
2. a voiceless labiodental fricative [f]	4	3.7	<i>fick di</i> , <i>Fidschi</i> , <i>Fiji</i> , <i>vrici</i>
3. a bilabial approximant [w]	4	3.7	<i>Web</i> , <i>Weekend</i> / <i>Veekend</i> , <i>Whiskey</i>
Initial sound other than [v], [f] or [w]	41	38.7	<i>Abi</i> , <i>arrivederci</i> , <i>audi</i> , <i>Berlusconi</i> , <i>gelati</i> , <i>klinsi</i> , <i>nix war's</i> , <i>paff</i> , <i>Spezi</i> , <i>Schnuffi</i> , ...
Identical or highly similar to the first syllable of <i>vici</i> , based on ...	49	46.2	...
First syllable			
1. a tense [i]-nucleus vowel ([vi], [fi], [wi])	36	34	<i>Vichy</i> , <i>Video</i> , <i>Visa</i> , <i>Vista</i> , <i>Vitus</i> , <i>vino</i> , <i>violine</i> , <i>Viva</i> , <i>Weekend</i> , ...
2. its lax variant [ɪ]-vowel ([vi], [fi], [wi])	13	12.2	<i>vicious</i> , <i>Vincenz</i> , <i>Vintage</i> , <i>Vista</i> , <i>Winkewinke</i> , <i>fick di</i> , <i>Fiji</i> , <i>Whiskey</i> , ...
First syllable other than [vi], [fi], [wi] or [vi], [fi], [wi]	57	53.8	<i>Abi</i> , <i>audi</i> , <i>Berlusconi</i> , <i>boni</i> , <i>canti</i> , <i>gelati</i> , <i>inflammavi</i> , <i>paff</i> , <i>Schnuffi</i> , <i>sorry</i> , <i>Toni</i> , <i>Vegil</i> , <i>vanitas</i> , <i>Wasem</i> , <i>Web</i> , <i>Zladdi</i> , ...
Last syllable			
Identical tense [i]-nucleus as in <i>vici</i>	67	63.2	<i>arrivederci</i> , <i>boni</i> , <i>Da Vinci</i> , <i>Fidschi</i> , <i>gaudi</i> , <i>ici</i> , <i>klinsi</i> , <i>Ricci</i> , <i>Schnuffi</i> , <i>sorry</i> , <i>Spezi</i> , <i>Veggie</i> , <i>Vichy</i> , <i>Wiki</i> , ...

Phonological qualities of the V ₃ filler items	Number	Percentage (106 fillers in total $\hat{=}$ 100%)	Examples
Nucleus other than [i]	39	36.8	<i>paff, vanitas, vertigo, Veto, vicious, Video, vino, Vintage, Visa, Vista, Visum, Web, Weserdeich, Wikipedia, ...</i>

By contrast, categorial and semantic restrictions do not seem to play a major role. As examples (4)–(9) illustrate, the slot is by no means restricted to (first-person) finite verbs. On the contrary, a broad variety of syntactic categories is used on V₃, verbs being rather rare. For example, one can find common nouns (4), proper nouns (5), adjectives (6), interjections (7), prepositional phrases (8), or even complete clauses (9), employing both native German material (4), (5), (8) and material from other languages such as English (6), Italian (7), or Swiss German (9).

- (4) *Weil Sie aber alles von A wie „Aha, Afrika kann tatsächlich eine WM organisieren“ bis V wie „Veni, vidi, Videobeweis“ noch so oft hören werden, ... hier bloß ein kleines, rasches „Best of“ der Nebensächlichkeiten, die von der WM bleiben werden ...*

‘As you will get to hear everything from A like ‘PRT Africa actually can organise a World Cup’ until V like ‘veni, vidi, video proof’ many more times, ... here we present only a small best-of of the trivial things that will remain after the world cup ...’

(DPR, 04.07.2010)

- (5) *Seit Tim veni, vidi, Wiese nicht mehr im Tor steht, suchen sie in Bremen eine echte Nummer 1.*

‘Ever since Tim veni vidi Wiese does not keep goal any more they are looking for a real number 1 in Bremen.’

(BZT, 13.02.2017)

- (6) *Veni Vidi Vicious: Der Titel des Albums basiert auf dem Ausspruch „Veni, Vidi, Vici“ ... von Julius Cäsar nach seinem erfolgreichen Gallien-Feldzug ...*

‘Veni Vidi Vicious: The album title is based on the dictum ‘Veni, vidi, vici’ ... by Julius Caesar after his successful campaign against Gallia ...’

(Wikipedia 2011, http://de.wikipedia.org/wiki/Veni_Vidi_Vicious)

- (7) *Für die Turniersiegerin 2010 hieß es also diesmal „veni, vidi, arrivererci“. Die Deutsche darf indes weiter hoffen.*
 ‘This time, ‘veni, vidi, arrivererci’ applied to the champion of 2010. The German, however, may still be hopeful.’
 (LTB, 16.10.2014)
- (8) *Er fuhr nach Moskau. Als er wiederkam, zitierte er auf dem Flughafen den berühmten Ausspruch Cäsars ... Jedoch mit einer kleinen Änderung ...: Veni, vidi, in Arsch.*
 ‘He went to Moscow. When he came back, he quoted Caesar’s famous dictum at the airport ... But with a little modification ...: Veni, vidi, in [my] ass.’
 (Stern, 19.02.1998)
- (9) *Das Video „Veni, vidi, fick di“ ist mit beinahe 19.000 Clicks einer der meist beachtetten romanischen Clips überhaupt.*
 ‘The video ‘Veni, vidi, fuck you’ is with about 19,000 clicks one of the most famous romance clips ever.’
 (SOZ, 11.12.2013)

As the examples indicate, the V_3 position not only seems to be indifferent as to the category and the linguistic code of the filler items, but also as to the semantics of the filler items. That is, the fillers do not belong to a restricted lexical-semantic set. It seems that the great tolerance towards fillers that neither match the syntactic category, the linguistic code, nor the semantics of the original item *vici* has to do with the status of *veni, vidi, vici* as a loan phrase that may not be transparent syntactically and semantically to speakers of German. Therefore, the prime linguistic characteristics that speakers can rely on when modifying this phrase is a feature that is perceptible at the surface, namely the phonological shape of V_3 .

Taken together, the corpus data indicate that [*veni, vidi, X*] indeed is a phraseological construction that allows for the productive formation of new instances to a certain degree. However, as about 69% of the instances are instances of the lexicalised quote, it does not seem warranted to speak of a true phraseological schema. That is, V_3 is not conceptually entrenched as an open slot, but rather as a position that is lexically specified, but shows a certain affinity to modification along the lines of the phonological form of its original filler item. At present, one would thus assume that [*veni, vidi, X*] is a modificatory pattern rather than a true phraseological schema.

In the next section, we turn to the German construction [X *kam, sah und Y*]. Comparing the usage of this construction with the usage of the Latin construction, it will become clear that the potential to be modified is higher in [X *kam, sah und Y*] compared to [*veni, vidi, X*]. We will argue that the German construction, despite its ‘secondary’ status as a translation, proves especially apt for productive usage because of its specific structural and semantic properties.

3 X *kam, sah und Y*

The Latin dictum *veni, vidi, vici* has a German equivalent that is codified in slightly different versions in different dictionaries, cf. (10)–(12).

- (10) *Ich kam, ich sah, ich siegte.*
 ‘I came, I saw, I conquered.’
 (Duden [1993] 2008; Büchmann 2014)
- (11) *Ich kam, sah und siegte.*
 ‘I came, saw and conquered.’
 (Mackensen 1973)
- (12) *Ich kam, sah, siegte.*
 ‘I came, saw, conquered.’
 (Reichert [1948] 1956)

A corpus search in the W-archive in *DeReKo* reveals that (10) and (11) are, relatively speaking, the most frequently realised variants (61 and 64 attestations, respectively), while (12) is rather rare (11 attestations). On this empirical basis, we will assume here that (11) is the basic form of expression of the dictum in German, with (10) being a commonly used variant.

As to their syntactic structure, (10) comes closest to the Latin original, in asynchronously conjoining three structurally identical main clauses with first-person singular subjects. As German is not a pro-drop language, the subject pronoun must, in principle, be realised. However, as is the case in variant (11), coordinate structures typically undergo a process of ellipsis, where an element that occurs identically in the different conjuncts – here, the subject pronoun – may be deleted in the non-initial conjunct(s). Furthermore, in enumerations of three or more elements, a strong convention is to conjoin (only) the last element via the conjunction *und* ‘and’

with the next-to-last element (Duden [1973] 2016: §1408). Together, these processes can explain the transformation of variant (10) to variant (11).

Obviously, in the translations in (10)–(12), the lexical content of the Latin verbs is largely maintained. This primacy of semantic equivalence leads to a loosening of the internal phonological bonds that are present in the Latin version and to a metric structure that is different from the Latin original. While the German variant (10) displays an iambic structure, variant (11), with pronominal ellipsis, exhibits a metric break between the first and the second verb, both of which are monosyllabic. The prosodic similarities between the three conjuncts are weaker in the German translation than in the Latin original. However, there is still an assonance in the nucleus between V_1 and V_2 (*kam, sah*), as well as in the onset between V_2 and V_3 (*sah, siegte*). Overall, in the German translation, V_3 stands out not only as the cognitive-semantic climax element, but also because of its prosodic properties, in particular its light nucleus vowel that contrasts with the dark nucleus vowels in V_1 and V_2 , and its disyllabic structure that contrasts with the monosyllabic structures of V_1 and V_2 .

Examples like (13)–(15), which are taken from the Internet, indicate that in analogy with the Latin construction, the German construction also allows for the productive formation of new instances via substitution of the final verb slot.

- (13) *Ich kam, sah und reanimierte*
 ‘I came, saw and reanimated’
 (book title)
- (14) *Ich kam, sah und hatte direkt keinen Bock*
 ‘I came, saw and immediately wasn’t in the mood’
 (t-shirt print)
- (15) *Ich kam, sah und musste spülen*
 ‘I came, saw and had to do the dishes’
 (dishcloth print)

However, a crucial difference between the Latin and the German construction is that the latter has an additional subject position at its disposal. The subject position can be assumed to be frequently exploited for variable filling as well, because subject variation will open the construction for usage in a broad range of contexts beyond speaker-related ones. We may thus assume that speakers of

German have at their disposal a phraseological construction [X *kam, sah und Y*]⁶ that can be exploited for productive use.

This assumption can be verified in the corpus. A search in the *DeReKo* W-archive results in 3,256 instances of the pattern [X *kam, sah und Y*]⁷. This number by far exceeds the number of instances of the fully lexicalised dictum (*Ich kam, sah und siegte*), which only has 136 attestations in total, i.e. all three codified variants together. Taking first a closer look at the subject position (S), the automatic slot analysis yields roughly 1,250 different filler items for S. Figure 1 shows the distribution over syntactic categories in percentage terms. As is apparent, the two largest classes by far are pronouns and proper nouns, followed by common nouns.

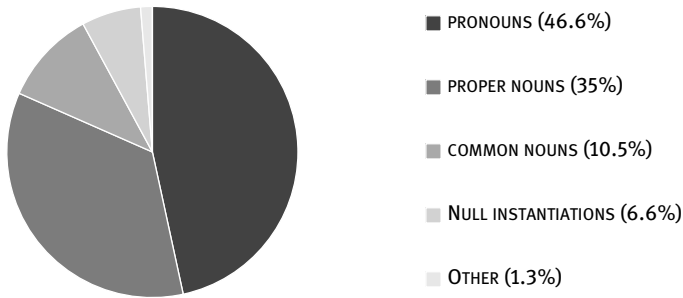
⁸


Fig. 1: Distribution of all S-slot fillers (total number of hits: 3,256) (*DeReKo* W-archive)

⁶ The schema [X *kam, sah und Y*] is somewhat simplified, disregarding the fact that the verb forms change their agreement features depending on the choice of subject (person, numerus), tense (e.g., past tense, present tense, future tense), and mood (indicative, subjunctive, imperative), so that the schema [X *kommen*_{finite}, *sehen*_{finite} und Y] seems to be more apt to represent the discussed phraseological construction. Furthermore, one can also find occasional instances with more than three conjuncts. Nevertheless, we decided to choose the simplified representation for the sake of better readability.

⁷ The search string “&kommen /+w1 &sehen” yielded a total of 19,198 hits, which were manually checked. Of 19,198 hits, 3,256 turned out to be attestations of the construction in question (for more details, see also footnote 6).

⁸ For reasons of space, we will not go into the rather small class of null instantiations here; these are represented by examples such as, e.g., *Kam, sah und traf: Brasiliens exzentrischer Stürmer Romario* [‘Came, saw, and scored: Brazil’s eccentric forward Romario’] (RHZ, 07.06.2014).

While it is expected, from the original dictum, that pronouns will make a large proportion of S fillers, it is rather surprising that the great majority of the pronouns are third-person rather than first-person personal pronouns, cf. Figure 2.

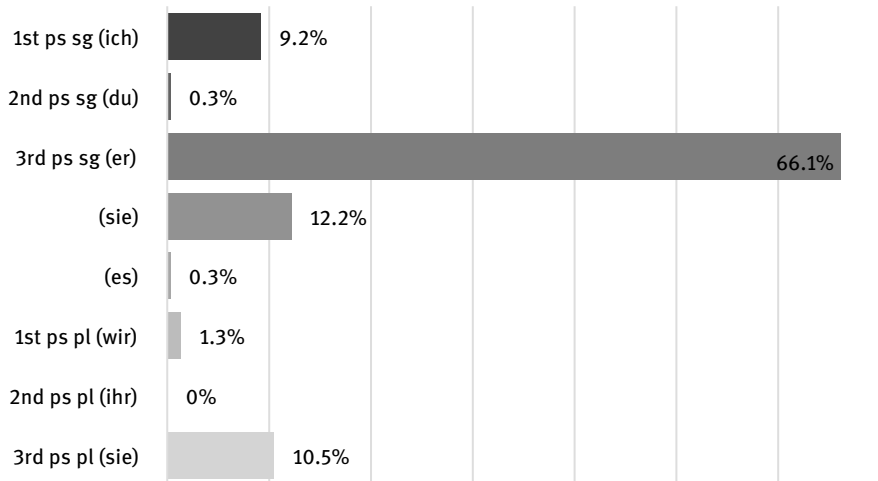


Fig. 2: Distribution of all S-slot pronoun fillers (total number of hits: 1,519) (*DeReKo W*-archive)

As indicated in Figure 2, third person singular masculine personal pronouns are most frequent (66.1%), whereas first-person singular and plural pronouns together only make up about 10.5% of the usages. This suggests that the codified form of the phraseme as given in the dictionary entries (cf. (10)–(12) above) does not seem to match the actual usage of the pattern in present-day German. In German, the verb forms of first and third person singular past tense are the same; therefore, the substitution of *ich* ‘I’ by *er* ‘he’ is accomplished very easily, without requiring any adaption of the verbs with regard to their inflection. This grammatical fact, together with the fact that a third person pronoun allows for a much broader referential potential than the first-person pronoun, may explain the preference for third person pronouns.

The second largest class are proper nouns, with 35% of S fillers. Naturally, as proper nouns refer to specific individuals, this is also the class with most hapaxes in the corpus (of 1,141 proper nouns, 780 are hapaxes). Thus, proper nouns contribute considerably to the productive usage of the S-slot. The majority of proper nouns used in the S-slot refer to celebrities, mostly from politics and sport, cf. (16)–(17). This is due to the high proportion of newspaper texts in *DeReKo*.

- (16) *Erdogan kam, sah und giftete*
 ‘Erdogan came, saw, and said nasty things’
 (HMP, 03.11.2011)
- (17) *Klinsi kam, sah und schwieg*
 ‘Klinsi came, saw, and remained silent’
 (HMP, 01.07.2008)

However, beyond anthroponyms, one can also find toponyms (18), names of organizations (19), or brand names (20), whose referents are conceptualised as agents, i.e. (groups of) human referents.

- (18) *Russland kam, sah und siegte wieder*
 ‘Russia came, saw, and triumphed again’
 (NZZ, 09.11.2007)
- (19) *Ver.di kam, sah und streikte*
 ‘Ver.di [a German trade union, L.S., R.F.] came, saw, and went on strike’
 (SZ, 24.03.2001)
- (20) *Disney kam, sah und schmierte ab*
 ‘Disney came, saw, and crashed’
 (BMP, 20.10.2017)

Turning to common nouns in the S-slot, it is expected that the verbs *kam* and *sah* will semantically restrict the fillers to human referents. This is also what we find in the corpus. As Table 4 shows, the nouns that are inserted either explicitly refer to persons or groups of persons (e.g., *Trainer* ‘coach’, *Publikum* ‘public’, *Favorit* ‘favorite’, *Gäste* ‘guests’, *Präsident* ‘president’) or they implicitly conceptualise a person or group of persons (e.g., *Polizei* ‘police’, *Streife* ‘patrol’, *Kommission* ‘committee’, *Kongreß* ‘congress’, *Kreisauswahl* ‘county selection’).

Tab. 4: First 30 of 225 different head nouns in S position, sorted by frequency

Filler	Translation	Nr.	Perc.	Filler	Translation	Nr.	Perc.
<i>Joker</i>	‘joker’	14	0.46	<i>Jungstar</i>	‘young star’	2	0.07
<i>Besucher</i>	‘visitor’	11	0.36	<i>Kandidat</i>	‘candidate’	2	0.07
<i>Trainer</i>	‘coach’	8	0.26	<i>Kanzler</i>	‘chancellor’	2	0.07

Filler	Translation	Nr.	Perc.	Filler	Translation	Nr.	Perc.
<i>Star</i>	'star'	7	0.23	<i>Kirchenfürst</i>	'church dignitary'	2	0.07
<i>Menschen</i>	'humans'	6	0.20	<i>Kommission</i>	'committee'	2	0.07
<i>Polizei</i>	'police'	6	0.20	<i>Kongreß</i>	'congress'	2	0.07
<i>Publikum</i>	'public'	6	0.20	<i>Kontrolleur</i>	'inspector'	2	0.07
<i>Prinz</i>	'prince'	5	0.16	<i>Kreisauswahl</i>	'county selection'	2	0.07
<i>König</i>	'king'	4	0.13	<i>Minister</i>	'secretary'	2	0.07
<i>Mann</i>	'man'	4	0.13	<i>Neuling</i>	'newcomer'	2	0.07
<i>Streife</i>	'patrol'	4	0.13	<i>Premier</i>	'prime minister'	2	0.07
<i>Weltstar</i>	'champion'	4	0.13	<i>Präsident</i>	'president'	2	0.07
<i>Favorit</i>	'favorite'	3	0.10	<i>Teamchef</i>	'team boss'	2	0.07
<i>Gäste</i>	'guests'	3	0.10	<i>Touristen</i>	'tourists'	2	0.07
<i>Kanzlerin</i>	'chancellor'	3	0.10	<i>Volk</i>	'people'	2	0.07

As to the syntactic complexity of the subject NPs, there don't seem to be any restrictions, as (21)–(22) illustrate.

(21) *Dieser quirlige, unbekümmerte, laufkräftige und schußgewaltige Junge kam, sah und schaffte das Aber.*

'This lively, easygoing, powerfully running and kicking boy came, saw and accomplished the 'But'.'

(DPR, 03.05.1993)

(22) *Auch das Juniorenteam, welches auf den „Hebepositionen“ komplett neu aufgestellt ist, kam, sah und siegte.*

'The junior team which was newly assembled on the 'lift positions' also came, saw and conquered.'

(SGT, 05.05.2011)

However, as pronouns and proper nouns make a large proportion of fillers, empirically, there is a tendency towards N⁰-categories in subject position, i.e. towards monolexical fillers.

With respect to the placement of the subject NP, the construction behaves wholly regularly. As in canonical German verb-second clauses, the subject may take the prefield position in front of the finite verb, as, e.g., in (21)–(22), or it may be situated in the middle field, cf. (23).

(23) *Trotzdem kamen, sahen und eroberten die „New Dubby Conquerors“ die Herzen und Tanzbeine der Partywilligen auf Antrieb.*

lit. ‘Nevertheless came, saw and conquered the ‘New Dubby Conquerors’ the hearts and dancing legs of the party people in an instance.’

(MM, 16.07.2001)

We turn now to the Y-slot of the pattern, i.e. the V_3 position. A first observation is that only about 40% of the instances display the canonical filler item *siegte(n)* ‘conquered’ on V_3 , cf. Table 5⁹. This is clearly below the 50% benchmark suggested by Stumpf (2016) and thus an argument in favor of a true phraseological schema. Moreover, the slot analysis reveals a large number and variety of other fillers, among them 320 hapax legomena, indicating that V_3 is indeed conceptually entrenched as an empty slot position to be variably filled in accordance with certain syntactic and semantic restrictions.

Tab. 5: First 40 of 512 different V_3 filler items in the German construction, sorted by frequency

Filler	Translation	Nr.	Perc.	Filler	Translation	Nr.	Perc.
<i>siegte</i>	‘conquered’	871	35.74	<i>lachte</i>	‘loughed’	12	0.49
<i>traf</i>	‘scored’	205	8.41	<i>siegt</i>	‘conquers’	12	0.49
<i>siegten</i>	‘(they) conquered’	76	3.12	<i>trifft</i>	‘scores’	12	0.49
<i>brüllte</i>	‘shouted’	44	1.81	<i>schoss</i>	‘kicked’	11	0.45
<i>ging</i>	‘went’	31	1.27	<i>machte</i>	‘made’	10	0.41
<i>staunte</i>	‘marveled’	27	1.11	<i>eroberte</i>	‘conquered’	9	0.37
<i>verlor</i>	‘lost’	26	1.07	<i>rettete</i>	‘saved’	9	0.37
<i>wurde</i>	‘became’	26	1.07	<i>flog</i>	‘flew’	8	0.33
<i>siegen</i>	‘(they) conquer’	24	0.98	<i>half</i>	‘helped’	8	0.33
<i>spielte</i>	‘played’	24	0.98	<i>feiern</i>	‘(they) celebrate’	7	0.29
<i>kaufte</i>	‘bought’	23	0.94	<i>gewann</i>	‘won’	7	0.29
<i>fuhr</i>	‘went’, ‘drove’	19	0.78	<i>hat</i>	‘has’	7	0.29
<i>liebte</i>	‘loved’	19	0.78	<i>schlug</i>	‘defeated’	7	0.29
<i>sagte</i>	‘said’	18	0.74	<i>strahlte</i>	‘gleamed’	7	0.29
<i>kauften</i>	‘(they) bought’	17	0.70	<i>feierte</i>	‘celebrated’	6	0.25

⁹ In addition to *siegte* and *siegten*, the following word forms of *siegen* are verifiable in the corpus, but clearly less frequent: *siegt* (12x), *siegen* (24x), *habe gesiegt* (1x), *hatte gesiegt* (1x), *wird siegen* (1x), *würde siegen* (2x) and *hätte gesiegt* (1x).

Filler	Translation	Nr.	Perc.	Filler	Translation	Nr.	Perc.
<i>schoß</i>	'kicked'	15	0.62	<i>hörte</i>	'heard'	6	0.25
<i>staunten</i>	'(they) marveled'	15	0.62	<i>ließ</i>	'let'	6	0.25
<i>blieb</i>	'remained'	14	0.57	<i>lobte</i>	'praised'	6	0.25
<i>sang</i>	'sang'	14	0.57	<i>raste</i>	'rushed'	6	0.25
<i>sägte</i>	'sawed'	13	0.53	<i>schrieb</i>	'wrote'	6	0.25

Taking a closer look at the V_3 fillers, the slot analysis reveals three crucial differences between the Latin and the German construction. First, in the German construction, the V_3 filler items are syntactically clearly restricted to finite verbs that show the required agreement features. Second, the set of verbs to fill the V_3 slot stems from a restricted number of semantic verb classes. And third, phonological restrictions seem to play a minor role in the choice of fillers.

Let us look at these three aspects. Table 5 indicates, first, that the great majority of fillers in this position are finite verbs (in past tense), with both singular or plural forms possible depending on the chosen subject. Occasionally, the verbs also occur in present or perfect tense. Of the entire 3,256 instances, there are only two non-verbal fillers, cf. (24)–(25).

(24) *Scout kommt, sieht, Vertrag – fertig!*
 'Scout comes, sees, contract – ready!'
 (FOC, 07.05.2007)

(25) *Er kam, sah – und Bam Bam!*
 'He came, saw – and bam bam!'
 (HMP, 10.01.2012)

While in the Latin construction, due to its status as a loan phraseme, the opacity of the pattern seems to license all kinds of categories in V_3 position, in the (fully transparent) German construction, the filler category is bound to the syntactic category regularly required in this position, namely, a finite verb showing agreement features with the subject.

As to the semantics of the fillers, the slot analysis in Table 5 shows that beyond *siegen* 'to conquer', a variety of verbs, e.g., *treffen* 'to score', *brüllen* 'to shout', *verlieren* 'to lose', *kaufen* 'to buy' is used in third position. While this may seem, at first glance, an arbitrary set of verbs, on closer investigation, one can distinguish four broad semantic classes of verbs. First, there is a set of verbs that can be regarded as roughly synonymous with *siegen* 'to conquer', either because

they explicitly denote a victorious action, e.g., *gewinnen* ‘to win’, *erobern* ‘to conquer’, or because they imply some kind of victorious action in the relevant contextual setting, often in a sports context, e.g., *treffen* ‘to score’, *erzielen* ‘to achieve’, *punkten* ‘to score’. Second, there is a set of verbs explicitly denoting a successful accomplishment, e.g., *beeindrucken* ‘to impress’, *bezaubern* ‘to charm’, as well as verbs that imply some kind of successful accomplishment in a relevant contextual setting, e.g., *singen* ‘to sing’, *kaufen* ‘to buy’ in the contexts of (26) and (27). In these contexts, the verbs denote what can be seen as successful accomplishment from the perspective of the audience of the concert (26), or the organisers of the Christmas market (27).

(26) *Die ersten Fans hatten sich schon am Vormittag auf den Weg gemacht, um den vom Privatfernsehen zum Superstar ausgerufenen Sänger Pietro Lombardi auf dem Trammplatz zu sehen ... Der Superstar kam, sah und sang – und alle waren glücklich.*

‘The first fans already hit the road in the morning in order to meet the singer Pietro Lombardi on the Tramm Square, who had been declared a superstar by private television ... The superstar came, saw, and sang – and everybody was happy.’

(HAZ, 27.06.2011)

(27) *Etwa 1.500 Besucher beim Weihnachtsmarkt in Preungesheim kamen, sahen und kauften.*

‘Around 1,500 visitors at the Christmas market of Preungesheim came, saw and bought.’

(FR, 11.12.1997)

Third, there is a class of verbs that denote a more passive behavior and thus imply stagnation rather than climax, e.g., *bleiben* ‘to remain’, *warten* ‘to wait’, *schweigen* ‘to remain silent’, e.g., in contexts like (28).

(28) *Der Täter wirkte unentschlossen. Er kam, sah und – wartete.*

‘The offender seemed indecisive. He came, saw and – waited.’

(DPR, 14.03.1998)

Finally, a fourth class comprises verbs that are in strong semantic contrast with the first class and that function like an anti-climax, i.e. verbs that denote some kind of defeat or failure, e.g., *verlieren* ‘to lose’, *versagen* ‘to fail’, *unterliegen* ‘to suffer defeat’. This category also includes verbs that explicitly conceptualise the

failure as a result of a ‘mishap’, e.g., *stolpern* ‘to stumble’, *stürzen* ‘to fall’, *humpeln* ‘to limp’. The four classes with illustrative examples are given in Table 6.

Tab. 6: Different semantic sets of filler verbs on V₃

Set	Examples	English translation
1	<p>Semantics and climax structure: roughly synonymous with <i>siegen</i> ‘to conquer’; create a climax</p> <ul style="list-style-type: none"> – 1a: denote a victorious action – 1b: imply some kind of victorious action in a relevant contextual setting 	
1a	<i>obsiegen, besiegen, gewinnen, (auf)trumpfen, brillieren, triumphieren, erobern, ungeschlagen sein, den Titel holen, ...</i>	to win, to defeat, to conquer, to crow, to shine with sth., to triumph, to take sth. by storm, to be unbeaten, to lift the trophy, ...
1b	<i>treffen, punkten, scoren, (er)zielen, netzen, knipsen, versenken, abstauben, das Spiel drehen, davonrasen/-flitzen, schlagen, ...</i>	to score, to score, to score, to strike (to achieve), to kick (the ball) into the net, to snap, to pot, to scrounge, to turn the match, to run away, to beat, ...
2	<p>Semantics and climax structure: similar to <i>siegen</i> ‘to conquer’ but with an attenuated denotation; still create a climax</p> <ul style="list-style-type: none"> – 2a: denote a successful accomplishment – 2b: imply some kind of successful accomplishment in a relevant contextual setting 	
2a	<i>begeistern, faszinieren, (be)zaubern, überzeugen, beeindrucken, beflügeln, Fähigkeiten beweisen, in den Bann ziehen, ...</i>	to thrill, to fascinate, to charm, to convince, to impress, to inspire, to demonstrate one’s ability, to enthral, ...
2b	<i>singen, kaufen, shoppen, entrümpeln, Birne auswechseln, servieren, basteln, kichern, reformieren, verhandeln, ...</i>	to sing, to buy, to shop, to declutter, to change the bulb, to serve, to do handicrafts, to giggle, to reform, to negotiate, ...
3	<p>Semantics and climax structure: contrary to <i>siegen</i> ‘to conquer’ by neutralizing its denotation; denote passive behavior and imply stagnation</p>	
	<i>gähnen, sinnieren, überlegen, bleiben, einstauben, warten, zaudern, streiken, ignorieren, schweigen, vergessen, ...</i>	to yawn, to ponder, to consider, to remain, to get dusty, to wait, to hesitate, to strike, to ignore, to remain silent, to forget, ...
4	<p>Semantics and climax structure: completely contrary or antonymous to <i>siegen</i> ‘to conquer’; create an anti-climax</p> <ul style="list-style-type: none"> – 4a: denote a defeat or failure – 4b: explicitly conceptualise the failure as a result of a ‘mishap’ 	
4a	<i>verlieren, versagen, scheitern, enttäuschen, unterliegen, untergehen, Nieder-</i>	to lose, to fail, to collapse, to disappoint, to suffer defeat, to go under, to suffer defeat,

Set	Examples	English translation
	<i>lage kassieren, aufgeben, ...</i>	to quit, ...
4b	<i>stolpern, stürzen, über sich selbst fallen, abrutschen, humpeln, trampeln, patzen, den Pfosten treffen, ein Foul begehen, ...</i>	to stumble, to fall, to trip over oneself, to slide, to limp, to tramp, to screw up, to strike the post, to foul, ...

Thus, it becomes clear that, contrary to the Latin construction, the German construction exhibits specific semantic restrictions that govern the choice of filler verbs. By contrast, phonological properties seem to play a minor role for the selection of verbs in the German construction. Having said that, the corpus analysis shows that there is in fact a small lexical set of verbs that seem to be chosen primarily because of their prosodic qualities, cf. Table 7.

Tab. 7: Filler verbs on V₃ that exhibit phonological similarity with *siegte* ‘conquered’

Set	Examples	English translation
5	Phonological shape: close phonological similarity with <i>siegte</i> ‘conquered’ – 5a: share a phonological feature with <i>siegte</i> , e.g., the onset, assonance or rhyme – 5b: share their onset with <i>siegte</i> and also show an assonance with the preceding verb <i>sah</i> ‘saw’	
5a	<i>siebte, piepte, (ver)liebte, quiekte, siechte, nieste, servierte, versiegte, signierte, versiegelte, sichtete, kriegte, nickte, sägte, saugte, sorgte, säuberte, seufzte, sendete</i>	sieved, beeped, loved (fell in love), squeaked, was sick, sneezed, served, run dry, signed, sealed, got, nodded, sawed, sucked, cared, cleaned, sighed, sent
5b	<i>sahnt ab, sanierte, sang (mit), sagte (aus), sagte zu (ab)</i>	creams off, reconstructed, sang (along), said, confirmed (canceled)

The verbs in Table 7 are characterised by close phonological similarity with *siegte* ‘conquered’, for example, in that they form minimal pairs with *siegte* (*siebte* ‘sieved’, *siechte* ‘was sick’, *sägte* ‘sawed’, *saugte* ‘sucked’, ...), share their onset (*sorgte* ‘cared’, *seufzte* ‘sighed’, *servierte* ‘served’, *signierte* ‘signed’, ...), their rhyme (*kriegte* ‘got’, *quiekte* ‘squeaked’, ...) or show assonances in the stressed syllable (*piepte* ‘beeped’, *(ver)liebte* ‘loved’, *nieste* ‘sneezed’, ...). A subset of phonologically determined verbs are verbs like *sahnt (ab)* ‘creams off’ and *sanierte* ‘reconstructed’, which not only share their onset with *siegte*, but also show an assonance with the preceding verb *sah*, e.g.

- (29) *Er kam, sah und sahnt ab*
 ‘He came, saw and creams off’
 (SZ, 11.07.2012)
- (30) *Herr Volkswagen: Er kam, sah und sanierte*
 ‘Mr Volkswagen: He came, saw and reconstructed’
 (ZEIT online, 20.09.2000)

However, as the phonologically motivated verbs only amount up to 28 types (142 tokens) of a total of 512 types (3,256 tokens), phonology still seems to be a minor selection criterion. What is more, one can observe that the aforesaid verbs are never chosen solely because of their phonological form, but also because of their semantic fit, as evidenced in examples like (31) and (32).

- (31) *Er kam, sah und sägte: Sven Bürki bezwang die globale Konkurrenz und wurde am Donnerstagabend zum Weltmeister in der Kategorie Möbelschreiner gekürt.*
 ‘He came, saw and sawed: Sven Bürki defeated the global competitors and was elected world champion in the category cabinet maker on Thursday evening.’
 (SGT, 20.10.2017)
- (32) *Er kam, sah und quiekte: Albert, das Hauptrollenschwein aus dem neuen Disney-Film „College Road Trip“.*
 ‘He came, saw and squeaked: Albert, the star role pig from the new Disney movie ‘College Road Trip’.’
 (SZ, 05.03.2008)

As we have seen above, the Latin construction [*veni, vidi, X*] is an instance of asyndetic coordination, being highly fixed as to the juxtaposition of three elements without connectives. By contrast, as is already apparent from the different codified variants (10)–(12) as mentioned above, the German construction [*X kam, sah und Y*] allows for a certain variability as to the occurrence of connective elements, as well as to graphematic marking. We will therefore, in a final step, take a closer look at the syntactic and graphematic connectivity of the construction.

A look at our *DeReKo* results reveals that of 3,256 attestations of the construction, 3,063 are instances with exactly three conjuncts. 95% of these cases (2,909 tokens) exhibit the connective *und* ‘and’ between V_2 and V_3 , thus repre-

senting the monosyndetic pattern [X kam, sah und Y].¹⁰ Only 5% of the cases are realizations of asyndetic coordination (i.e. [X kam, sah, Y]). Interestingly, among the 95% monosyndetic instances, there is a broad variety of graphematic means that occur together with the connective. Table 8 provides an overview.

Tab. 8: Graphematic variants in monosyndetic coordination with three conjuncts (total number of hits: 2,909)

Connective syntax	Number of hits	Example
V ₁ , V ₂ und V ₃	2,561	<i>Er kam, sah und rettete 100 Arbeitsplätze.</i> ['He came, saw and saved 100 job positions.'] (SGT, 09.09.2016)
V ₁ , V ₂ & V ₃	3	<i>Ein fast perfekter Einstand: Petric kam, sah & siegte.</i> ['An almost perfect draw: Petric came, saw & conquered.'] (HMP, 24.08.2008)
V ₁ , V ₂ – und V ₃	299	<i>Die Polizei kam, sah – und siegte nicht.</i> ['The police came, saw – and did not conquer.'] (VAN, 10.12.1998)
V ₁ , V ₂ , und V ₃	28	<i>Die Lady kam, sah, und beraubte uns des liebsten Feindbildes: der Briten.</i> ['The Lady came, saw, and stole from us our favorite concept of an enemy: the British.'] (SZ, 19.05.2011)
V ₁ , V ₂ und ... V ₃	9	<i>Rattle kam, sah und ... Nein, das tat er nicht.</i> ['Rattle came, saw and ... No, he did not do that.'] (BMP, 25.08.2007)
V ₁ , V ₂ und – V ₃	6	<i>Arnie kam, sah und – schwelgt in Erinnerungen.</i> ['Arnie came, saw and – wallows in memories.'] (DPR, 10.10.1997)
V ₁ , V ₂ ... und V ₃	2	<i>Helmut Kohl kam, sah ... und meckerte.</i> ['Helmut Kohl came, saw ... and bellyached.'] (NUN, 12.12.1998)
V ₁ , V ₂ , und ... V ₃	1	<i>Wir kamen, sahen, und ... haben gelesen.</i> ['We came, saw, and ... have read.'] (NON, 19.11.2007)
Total	2,909	

As Table 8 shows, in total eight different variants in graphematic realization of the monosyndetic instances can be found in the corpus. While the realizations

¹⁰ Also among the 193 instances that include more than three conjuncts, the majority (175 instances) represent a monosyndetic pattern such as *Er kam, sah, griff in die Harfe und siegte: Der Niedersachse Holger Schäfer (40) ist Minnesänger des Jahres 2012.* ['He came, saw, played the harp, and conquered: the man from Lower Saxony, Holger Schäfer (40), is minnesinger of the year 2012.'] (RHZ, 14.06.2012).

with ampersand instead of *und* seem to be mere typographic variants, the realizations with pre- or postposed punctuation mark + *und* may be regarded as graphematic means that are used in order to achieve specific stylistic effects, for example, to mark a break in order to increase tension in awaiting the last element in the chain. Interestingly, the more than 300 (299+6) instantiations with an en dash are particularly frequent with verbs of the fourth, “anti-climax” set, or with negated verbs of conquering, e.g.

(33) *Er kam, sah – und patzte!*

‘He came, saw – and screwed up!’

(HMP, 04.10.2010)

(34) *Sie kamen, sahen – und siegten nicht: Barack Obama, Wen Jiabao und andere mächtige Staatschefs wollten der Welt einen Mini-Klimakompromiss diktieren.*

‘They came, saw – and did not conquer: Barack Obama, Wen Jiabao and other potent heads of state wanted to dictate the world a mini climate compromise.’

(SPIEGEL online, 19.12.2009)

In such instances, the en dash marks a graphematic, as well as a prosodic break that reinforces the semantic break induced by the choice of the final verb. A similar function can be ascribed to the less frequently occurring variants with ellipsis (...), the combination of comma + *und* (, *und*) or the combination of comma, *und*, and ellipsis (, *und* ...).

If we take our findings together, the corpus data clearly indicate that [X *kam, sah und* Y], with 3,256 instances, is a productive phraseological construction. About 60% of the instances of [X *kam, sah und* Y] display a filler item different from the canonical filler *siegte(n)* ‘conquered’. By contrast, the codified dictum in the first person singular, in its different variants (*Ich kam, (ich) sah (und) (ich) siegte*), is only attested 136 times in the corpus. Therefore, we may assume that [X *kam, sah und* X] is a true phraseological schema.

4 Summary

The aim of the present study was to contribute to the growing research on lexically semi-fixed phraseological constructions that are at the heart of Construction Grammar and phraseology, two research traditions that have seen an in-

creasing tendency to converge in recent years. Using newly developed corpus linguistic tools especially designed for pattern analysis, we examined the construction [X *kam, sah und* Y], which has not previously been subject to comprehensive linguistic analysis. This construction proves to be an especially interesting case because of its dual existence in German as both a loan phrase from Latin and its codified German translation.

Our analysis revealed, in fact, variability in both cases, but interesting differences in the exact modifying potential between the Latin and the German construction. The empirical data suggest that [veni, vidi, X] is a modificatory pattern in the sense of Stumpf (2016), because in 69% of the attestations (443 instances) the X position is filled by the lexicalised archetype *vici*, while only 31% (199 instances) can be qualified as modifications of this lexicalised form. The Latin pattern thus is profitable only to a certain degree. This can be explained mainly by the strong phonological restrictions that lexical items have to satisfy to be eligible as slot fillers in the Latin pattern.

[X *kam, sah und* Y], on the other hand, is a true phraseological schema containing two conceptually open slots that are regularly filled with a variety of lexical items, namely the additional subject position and the third verb position. In well over 50% of cases, both slots are filled with lexemes other than those in the according positions of the original quote. In particular, the additional autonomous subject position enables the application of the construction beyond speaker-related usages, while the third verb position allows for the formation of “anti-climaxes” due to its categorial and semantic rather than phonological restrictions. Together, these aspects change the German pattern version into a flexible piece of “sprachliche Fertigware” [‘linguistic convenience product’] (Mieder 2006: 199), adaptable to almost any occasional need. [X *kam, sah und* Y] thus allows for verbalizing virtually any result of any chain of action that is carried out by any person in any context.

From a contrastive perspective, it was thus demonstrated that it is necessary to precisely determine the specific factors (e.g., the structural and semantic properties) which constrain the actual exploitation of certain phraseological constructions in present-day German. These factors have a decisive impact on the different patterns’ varying productive potential and therefore are crucial for their precise localisation within the continuum between occasional modification and systematic schematicity. Working out these specific constraints on productivity will be an important task for future research in the field of lexically semi-fixed phraseological constructions. Stumpf’s (2016) distinction between ‘modificatory patterns’ and ‘true phraseological schemata’, based on the 50% benchmark discussed at the beginning, has turned out to be a valid empirical

instrument for the localisation of such constructions on this continuum, at least for the present case.

References

- Agricola, Erhard. 1977 [1962]. *Wörter und Wendungen. Wörterbuch zum deutschen Sprachgebrauch*, 8th edn. Leipzig: VEB Bibliographisches Institut.
- Bauer, Laurie. 2001. *Morphological productivity*. Cambridge: Cambridge University Press.
- Booij, Geert. 2002. *Construction morphology*. Oxford: Oxford University Press.
- Bubenhofer, Noah. 2015. Muster aus korpuslinguistischer Sicht. In Christa Dürscheid & Jan Georg Schneider (eds.), *Handbuch Satz – Äußerung – Schema*, 485–502. Berlin & New York: de Gruyter.
- Büchmann, Georg. 2014 [1864]. *Geflügelte Worte. Der klassische Zitatenschatz*, 43th edn. by Winfried Hofmann. Berlin: Ullstein.
- Dobrovolskij, Dmitrij. 2011. Phraseologie und Konstruktionsgrammatik. In Alexander Lasch & Alexander Ziem (eds.), *Konstruktionsgrammatik III. Aktuelle Fragen und Lösungsansätze*, 111–130. Tübingen: Stauffenburg.
- Dobrovolskij, Dmitrij. 2018. Sind Idiome Konstruktionen? In Kathrin Steyer (ed.), *Sprachliche Verfestigung. Wortverbindungen, Muster, Phrasem-Konstruktionen*, 11–24. Tübingen: Narr Francke Attempto.
- Duden. 2008 [1993]. *Zitate und Aussprüche. Herkunft und aktueller Gebrauch – rund 7.500 Zitate von der klassischen Antike bis zur Gegenwart*, 3rd edn. Mannheim, Wien & Zürich: Dudenverlag.
- Duden. 2016 [1973]. *Die Grammatik*, 9th edn. Berlin: Dudenverlag.
- Engelberg, Stefan, Anke Holler & Kristel Proost (eds.). 2011. *Sprachliches Wissen zwischen Lexikon und Grammatik*. Berlin: de Gruyter.
- Fillmore, Charles J., Paul Kay & Mary Catherine O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language* 64 (3). 501–538.
- Finkbeiner, Rita. 2008a. *Idiomatische Sätze im Deutschen. Studien zu ihrer Syntax, Semantik und Pragmatik und Untersuchung ihrer Produktivität*. Stockholm: Acta Universitatis Stockholmiensis (Stockholmer Germanistische Forschungen 72).
- Finkbeiner, Rita. 2008b. Zur Produktivität idiomatischer Konstruktionsmuster. Interpretierbarkeit und Produzierbarkeit idiomatischer Sätze im Test. *Linguistische Berichte* 216 (4). 391–430.
- Finkbeiner, Rita. 2017. *Argumente hin, Argumente her*. Regularity and idiomaticity in German *N hin, N her*. *Journal of Germanic Linguistics* 29 (3). 205–258.
- Finkbeiner, Rita. in press. *Kein ZDF-Film ohne Küsse im Heu. Kein X ohne Y* zwischen Lexikon, Grammatik und Pragmatik. In Carmen Mellado Blanco, Fabio Mollica & Elmar Schafroth (eds.), *Konstruktionen zwischen Lexikon und Grammatik: Phrasem-Konstruktionen monolingual, bilingual, multilingual*. Berlin & Boston: de Gruyter (Linguistik – Impulse und Tendenzen).
- Fleischer, Wolfgang. 1982. *Phraseologie der deutschen Gegenwartssprache*. Leipzig: VEB Bibliographisches Institut.

- Grice, H. Paul. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Syntax and semantics 3: Speech acts*, 41–58. New York: Academic Press.
- Gries, Stefan T. 2008. Phraseology and linguistic theory: A brief survey. In Sylviane Granger & Fanny Meunier (eds.), *Phraseology: An interdisciplinary perspective*, 3–25. Amsterdam & Philadelphia: John Benjamins.
- Hein, Katrin. 2015. *Phrasenkomposita im Deutschen. Empirische Untersuchung und konstruktionsgrammatische Modellierung*. Tübingen: Narr Francke Attempto.
- Jackendoff, Ray. 2008. Construction after construction and its theoretical challenges. *Language* 84 (1). 8–28.
- Jacobs, Joachim. 2008. Wozu Konstruktionen? *Linguistische Berichte* 213. 3–44.
- Leibniz-Institut für Deutsche Sprache. 2019. *Deutsches Referenzkorpus / W-archive*. Mannheim: Leibniz-Institut für Deutsche Sprache. <http://www.ids-mannheim.de/DeReKo> (accessed 28 August 2019).
- lexpan. *Lexical Pattern Analyzer. Ein Analysewerkzeug zur Untersuchung syntagmatischer Strukturen auf der Basis von Korpusdaten*. Developed by the project „Usuelle Wortverbindungen“. Mannheim: Leibniz-Institut für Deutsche Sprache. <http://uwv.ids-mannheim.de/lexpan/> (accessed 28 August 2019).
- Mackensen, Lutz. 1973. *Zitate, Redensarten, Sprichwörter*. Brugg: Fackelverlag.
- Mellado Blanco, Carmen, Fabio Mollica & Elmar Schafroth (eds.). in press. *Konstruktionen zwischen Lexikon und Grammatik: Phrasem-Konstruktionen monolingual, bilingual, multilingual*. Berlin & Boston: de Gruyter (Linguistik – Impulse und Tendenzen).
- Mieder, Wolfgang. 2006. *Andere Zeiten, andere Lehren. Sprichwörter zwischen Tradition und Innovation*. Baltmannsweiler: Schneider-Verlag Hohengehren.
- Mieder, Wolfgang. 2013. *Veni, vidi, vici*. In Wolfgang Mieder (ed.), *Neues von Sisyphus. Sprichwörtliche Mythen der Antike in der modernen Literatur, Medien und Karikaturen*, 350–374. Wien: Praesens Verlag.
- Plag, Ingo. 2006. Productivity. In Bas Aarts & April McMahon (eds.), *Handbook of English linguistics*, 537–556. Oxford: Blackwell.
- Ptashnyk, Stefaniya. 2009. *Phraseologische Modifikationen und ihre Funktionen im Text. Eine Studie am Beispiel der deutschsprachigen Presse*. Baltmannsweiler: Schneider-Verlag Hohengehren.
- Reichert, Heinrich G. 1956 [1948]. *Urban und human: Unvergängliche lateinische Spruchweisheit*, 4th edn. Hamburg: von Schröder.
- Steyer, Kathrin & Annelen Brunner. 2014. Contexts, patterns, interrelations – new ways of presenting multi-word expressions. *Proceedings of the 10th Workshop on Multiword Expressions (MWE 2014), Gothenburg, Sweden, 26-27 April 2014*, 82–88.
- Steyer, Kathrin. 2018. Lexikalisch geprägte Muster. Modell, Methoden und Formen der Online-Präsentation. In Kathrin Steyer (ed.), *Sprachliche Verfestigung. Wortverbindungen, Muster, Phrasem-Konstruktionen*, 227–264. Tübingen: Narr Francke Attempto.
- Steyer, Kathrin (ed.). 2018. *Sprachliche Verfestigung. Wortverbindungen, Muster, Phrasem-Konstruktionen*. Tübingen: Narr Francke Attempto.
- Stumpf, Sören. 2016. Modifikation oder Modellbildung? Das ist hier die Frage – Abgrenzungsschwierigkeiten zwischen modifizierten und modellartigen Phrasemen am Beispiel formelhafter (Ir-)Regularitäten. *Linguistische Berichte* 247. 317–342.
- Taylor, John. 2012. *The mental corpus. How language is represented in the mind*. Oxford: Oxford University Press.

Ziem, Alexander. 2018. Construction grammar meets phraseology: eine Standortbestimmung. *Linguistik Online* 90 (3/18). 3–19.

Ziem, Alexander (ed.). 2018. *Muster im Sprachgebrauch. Construction grammar meets phraseology*. [Special issue]. *Linguistik Online* 90 (3/18).

Index

- admiration 191, 196–198
- affirmation 57–58, 70, 74
- agreement 16, 53–67, 73, 304
- alignment 4, 272
- analysability 17, 133, 158–159
- analysis 37–47, 56–74, 87–98, 121–129, 140–155, 178–184, 200–211, 224–231, 272–283, 291–310
 - automatic analysis 288
 - contrastive analysis 289
 - Corpus Pattern Analysis (CPA) 2
 - corpus-based analysis 30, 88, 291
 - filler analysis 165, 168, 170
 - frame-based analysis 19, 217, 222
 - interlingual analysis 15
 - multiword term analysis 18, 216, 220
 - pattern analysis 311
 - qualitative analysis 141, 195, 270, 283
 - semantic analysis 220, 222
 - slot analysis 87, 91, 292–304
 - syntactic analysis 168
 - translation analysis 275, 284
- attenuation 34, 135
- backchannel 16, 60, 62
- Basque 15, 19, 20, 265–286
- bewilderment 191–197
- binomial 139
- calque 15, 19, 252, 255
 - calque in Basque 282
 - calques of French 235, 236, 252
 - constructional calque 119
 - English calques 236
- cliché 10, 107
- (to) coerce (*also* coercion) 119
- cohesion 133, 141, 154
- collocation 55–74, 266
 - collocation profile 165–180
 - collocation lemma 55
 - collocational framework 2
 - collocational partner 15
 - idiomatic collocation 274
 - semi-idiomatic collocation 274
 - verb preposition collocation 15, 19
- combination 56–57, 66–72, 165–185, 191–192, 226–231, 266–268
 - fixed combination 195
 - free word combination 18, 191–192, 211
 - frequent combination 66
 - preposition-noun combination 18, 165–169
 - stable combination 136
 - verb-adjective combination 274
 - word combination 18, 84, 94, 135, 149, 153, 191–211, 215–226, 266–275
- comparison 29, 30, 35, 41–48
- compositionality 17, 19, 117, 191, 211, 226, 237, 266, 267
- concordance tool 170
- constraint, *see also* restriction 20, 32, 83, 86, 108–109, 119, 129, 228–232, 288, 290, 311
 - morpho-syntactic constraint 140
 - ontological constraint 228, 231
 - part-of-speech constraint 129
 - phonological constraint 20
 - semantic constraint 32, 108, 119, 288
- construct 15–17, 40, 79, 99, 100, 136, 143, 149, 153–159
- construction 37–47, 88–98, 107–110, 135–140, 191–192, 287–290
 - abstract construction 17, 86, 116–117, 149, 155, 158, 166
 - comparative construction 35–36
 - completely filled construction 133, 148
 - Construction Grammar 4, 9–13, 30–33, 47–49, 79–84, 101–102, 109, 116–117, 135, 139–140, 168, 191, 287–288
 - contaminating construction 46
 - family of constructions (*also* constructional family) 82, 87, 99–102
 - free syntactic construction 107, 116
 - fully lexically specified construction 107, 116
 - German construction 20, 290–301
 - grammatical construction 33, 245
 - Incredulity Response Construction 6, 11

<https://doi.org/10.1515/9783110520569-013>

- intensifying construction 8, 133, 138, 143, 147–149, 152, 154
- Latin construction 253–254, 289–308
- macro-construction 99
- meso-construction 100
- micro-construction 16, 29–31, 43, 46, 99–101, 153
- partially filled constructions 17–18, 149–153
- phraseological construction, *see also* constructional phraseme 20, 133, 148, 287–312
- productive construction 153
- resultative construction 85
- Russian construction 18, 191–214
- schematic construction 8, 30–35, 48, 84, 153
- semischematic construction 9–20, 29, 33
- substantive construction 12, 17, 100
- true construction 6, 8
- constructionalisation 9, 17, 141, 159
- continuum 6–13, 33–35, 84, 116–117, 154, 158–159, 226, 311–312
- continuum of constructionalisation 17, 159
- idiomaticity continuum 35
- lexicon-grammar continuum, *see also* lexis-grammar continuum 5, 33–34, 48, 84
- lexis-grammar continuum 7, 12–17
- syntax-lexicon continuum, *see also* lexis-grammar continuum 17, 107, 116–117, 129
- conventionalization 1, 5, 259–261
- conversational routine 15–16, 53–75
- co-occurrence 3, 135, 178, 240, 266
- corpus, corpora 29–30, 37, 54–57, 87–88, 110, 140–141, 172, 191–193, 224–225, 269–275
- corpus analysis 8, 79, 88, 99, 200, 290–291
- corpus-based 13–20, 30, 79, 88, 101, 133, 141, 171–172, 289, 291
- corpus-driven 169–170
- learner translation corpus 19, 265–284
- parallel corpus 18, 191, 193, 269
- spoken corpus 55–57
- written corpus 55–57
- creativity 8–13, 29–30, 44–47, 192
- Czech 16, 53–75
- dictum 290–312
- disagreement 16, 61–62, 73
- discourse marker 53–75, 165–172
- doubt 16, 61, 63
- English 1–13, 200–211, 220–231, 236–261
- Middle English 14–19, 235–261
- equality 111, 120
- equivalence (*also* equivalent) 120, 200–211, 215, 235, 275–283, 297–298
- English equivalent 171, 174
- frequent equivalent 206–211
- full equivalent 87
- functional equivalent (*also* functionally equivalent) 87
- German equivalent 20, 208, 287–312
- Latin equivalent 248, 287–289
- Swedish equivalent 200–210
- systematic equivalent 200
- translation equivalent 200–211
- zero equivalent 191–211
- evaluation 62–66, 71, 114–115, 118, 125, 152, 181, 231
- expression 31–32, 61–65, 87, 166, 171–172, 184, 192
- compositional expression 32, 54
- fixed expression 13, 184, 237
- idiomatic expression 31, 155, 180, 183
- multiword expression, *see also* multiword-lexeme 2
- nominal expression 62, 65, 74
- phraseological expression 217
- positive expression 64
- temporal expression 165, 172
- time expression, *see also* temporal expression 55, 63, 176, 184
- extension 38–40, 44–47, 168–169, 181–183
- analogical extension 8–16, 44–49
- creative extension 29, 46
- external extension 168
- internal extension 181–183
- lexical extension 179

- optional extension 169
- extraction 140–141, 147, 224–231
 - data extraction 133, 140, 147
 - multiword term extraction 215–232
- filler 15, 41, 86–93, 168–170, 178–184, 195–196, 287–310
 - adjective filler 169
 - canonical filler 303, 310
 - frequent filler 182, 196
 - lexical filler 34, 41, 79, 87–88, 288
 - non-verbal filler 304
 - noun filler (*also* nominal filler) 169, 178, 300
 - potential filler 48, 288, 290
 - pronoun filler 300
 - slot filler, *see also* filler 4–20, 165–185, 311
- fixedness 33, 141–149, 166, 191, 237–238
- flexibility 13, 153
 - semantic flexibility 5
 - syntactic flexibility 267, 284
- formula 235–261
 - acknowledgement formula 240–249
 - astonishment formula 240–249
 - benediction formula 240–258
 - closing formula 240–257
 - endorsement formula 240–251
 - haste formula 240–250
 - health formula 240–257
 - information formula 240–249
 - letter formula 235–259
 - place and date formula 240–253
 - recommendation formula 240–258
 - request formula 240–250
 - salutation formula 240–257
 - structural formula 9–11
 - subscription formula 240–257
 - transfer of letter formulae 238, 259
 - welfare formula 240–257
- frame 7, 12–13, 215–232
 - frame-based 18–19, 215–232
 - Frame Semantics 14, 19, 221
 - primary frame element 229
 - secondary frame element 229–230
 - syntactic frame 7, 12
- French 247–261, 284
 - Medieval French 235
- frequency 63, 91–94, 123, 126, 149–153, 210–211, 237, 266–267, 293, 301–303
 - absolute frequency 150–152
 - frequency of co-occurrence 3, 267
 - frequency of translation equivalents 210
 - frequency of use 129–130, 136, 149, 153, 159
 - high frequency 3, 57, 66, 94, 100, 168, 196, 211, 274
 - low frequency 37, 40, 66
 - normalised frequency 150–152
 - token frequency 7, 9, 45, 48, 100
 - type frequency 7, 30, 44–45
- (to) generate (*also* generated, generating) 8–9, 11, 34, 181, 192
- German 15–20, 79–102, 147, 165–185, 191–211, 215–261, 265–284, 287–312
- (Modern) Greek 107–130
- hapax 46, 292, 300, 303
- idiom 1–20, 19–49, 53–55, 74–75, 83, 94, 107–109, 112–117, 173–174, 180, 191–192, 211, 237, 266, 283
 - constructional idiom 1–20, 30–49
 - comparative constructional idiom 29–49
 - creative play of idioms, *see also* modification and wordplay 13
 - idiom principle 1
 - idiom variant 30
 - open-slot idiom 30
 - schematic idiom 6, 34, 79, 83, 288
- idiomaticity 2, 108, 111–117, 154, 181, 191, 266
- indifference 18, 191–211
- creative instance (*also* modified instance, new instance, non-canonical instance, occasional instance) 11, 16, 30, 44–45, 149, 153, 291, 297, 296, 298–299
- instantiation 20, 100–101, 299, 310
- intensification 35–36, 128, 133–137, 155–158
 - intensification strategies 133, 155
 - morphological intensification 134

- phonological intensification 134
- syntactic intensification 134
- intensifier 69–71, 133–158
 - constructional intensifier 15–17, 133–159
 - lexical intensifier 134–136
- intonation 59, 61, 67, 134
- irony 37, 71
- Italian 15–19, 119, 133–159, 215–232, 275
- judgment 118, 197–199
- language 1–2, 8, 44, 79, 83, 133–134, 137, 149, 191–193, 215
 - contact language 20, 236
 - foreign language 15–18, 121, 128, 168, 174, 192, 211, 270, 277
 - formulaic language 1, 14, 19, 235–261
 - language specificity 18, 191–211
 - language transfer 19, 235–261
 - prefabricated language 32
 - source language 193, 265–270
 - spoken language 16, 53–75, 87
 - target language 209–210, 269–279
 - technical language 215–218
- lexeme 121–129, 134–136, 141–159, 166
 - mono-lexeme 18, 184
 - multiword-lexeme, *see also* multi-word expression 18, 55, 87, 99, 166, 170–184, 274
 - simple lexeme 102
- lexicalisation 17, 133–159
- (Bilingual) Lexicography 18, 191–211
- (to) license (*also* licensed, licensing) 14, 16, 30, 34, 44, 260
- meaning 3–6, 107–108, 111–130, 148–155
 - abstract meaning 86
 - actual meaning 115
 - canonical meaning 10
 - cognitive meaning 144
 - compositional meaning 34, 275
 - constructional meaning 10, 34, 141
 - core meaning 82, 87, 90, 169
 - denotational meaning 144
 - Extended-unit-of-meaning model 2
 - extension of meaning 39–40
 - form-meaning pair (*also* form-meaning pairing) 29, 31, 38, 79, 149
 - generalised meaning 111–114
 - holistic meaning 166–172
 - idiomatic meaning 47, 53, 147, 182, 275
 - idiosyncratic meaning 158
 - intensified meaning (*also* intensifying meaning) 19, 134–146, 155
 - ironic meaning 61
 - lexical meaning 83, 113, 117, 192, 275
 - meaning of equality 111
 - meaning of intensification, *see also* intensified meaning 128
 - metaphorical meaning 146
 - non-compositional meaning 34
 - original meaning 10, 148
 - pairing of form and meaning, *see also* form-meaning pair 15, 36
 - pattern meaning 169, 178
 - polysemous meaning 19, 149
 - pragmatic meaning 29–48, 155
 - qualifying meaning 150
 - telic meaning 146
 - terminological meaning 223
 - typical meaning 107–130
 - unit of meaning 2
 - vague meaning 61, 64
- modification, *see also* creative play of idioms and wordplay 6, 9–13, 45–49, 155, 259, 289–292
 - creative modification 9–13
 - occasional modification 20, 46, 291, 311
- multitude 17, 107–130, 141
- negation 17, 113–138
- negative 41–42, 60–65, 94, 114–115, 125, 128, 152, 179, 197–199, 210, 281
 - negative adjective 47
 - negative affirmation 74
 - negative agreement 16, 60–62
 - negative attitude 75, 210
 - negative connotation 18, 60, 125, 128, 152
 - negative evaluation 59, 71, 114–128, 152
 - negative information 60
 - negative judgment 197
 - negative response 64
 - negative situation 41–48
 - negative surprise 67, 198

- n-gram 1–14, 33–34, 53–75
 - bigram 53–75
 - tetragram 16, 53, 58, 65
 - trigram 16, 53–75, 195
- non-compositionality 17–19, 117, 237, 266
- pattern 1–12, 59–66, 82–86, 136–140, 168–170, 267–268, 311
 - co-occurrence pattern 135
 - extension pattern 181, 183
 - formulaic pattern 5
 - grammatical pattern 2, 14, 36, 267
 - lexical pattern 3–4, 15, 18, 56, 168–184
 - Lexical Pattern Analyzer (LexPan) 20, 79–87, 170, 287–293
 - modificatory pattern 3–4, 15, 18, 56, 168–169, 173, 184
 - multiword pattern 147
 - pattern of coining 8
 - pattern productivity 153
 - productive pattern 11, 13, 17, 45, 99, 101, 115, 159, 267, 287
 - schematic pattern 34
 - speech pattern 16
 - verbal pattern 19, 265–284
- phraseme, *see also* formulaic unit, phraseological unit and multiword unit and phraseologism 1, 83
 - constructional phraseme, *see also* phraseological construction 15–17, 79–102, 107–130, 192
 - loan phraseme 296, 304, 311
 - family of constructional phrasemes 80
 - partially lexicalized phraseme 80
 - phraseme annotation 53
 - specialised phraseme 218
- phraseologism, *see also* phraseme, formulaic unit, phraseological unit and multiword unit 1–6, 107, 109, 116–117, 267–269
 - canonical phraseologism 10–13, 302–303, 310
 - lexical phraseologism 114
 - lexically specified phraseologism 116
- Phraseology 1–9, 31–33, 47–48, 82–84, 109, 191, 216–217, 266–269, 287–288
 - Specialised Phraseology 215
 - Didactics of Phraseology 268–270, 283
- predictability 18, 133–159, 211
- prefab 1, 13–14, 237
- preposition-noun combination 18, 165–185
- productivity 6–9, 12–13, 15–16, 44–45, 48–49, 147, 155, 158–159, 289, 292
 - degrees of productivity 6–7, 136, 149, 289
 - E-Productivity 16
 - productivity of a pattern, *see also* pattern productivity 292
- profile 15, 165–180
 - collocation profile 165–178, 180
 - syntagmatic profile 2, 15
- semantic prosody 2, 152
- proverb 9–13, 107, 109, 185, 237
 - anti-proverb 9
 - modification of proverbs 9
 - proverb pattern 9
- pun 46
- quotation 9–15, 185, 240
- reduplication 134, 137–138, 144–145
- restriction, *see also* constraint 7–8, 94, 122, 149, 252, 302–304, 311
 - lexical restriction 192
 - morphological restriction 11
 - morpho-syntactic restriction 149
 - ontological restriction 231
 - part-of-speech restriction 122
 - phonological restriction 290–311
 - pragmatic restriction 31, 48, 288
 - semantic restriction 32, 34, 99, 295–303
 - syntactic restriction 149
- Russian 15, 18, 107–121, 191–211
- satisfaction 16, 61, 63, 166–167
- schema (*also* scheme, schemata) 6, 109, 299
 - abstract schema 154
 - phraseological schema 33, 46, 49, 287, 289
 - true phraseological schema 10, 46, 287–311
- schematicity 6–8, 35, 86, 143, 153
 - degrees of schematicity 17, 153, 288

- pattern schematicity 133
- systematic schematicity 20, 311
- schematism 147
- sequence 2, 56, 139–146, 158–159, 266–268
- formulaic sequence 237, 266
- slot 4, 15, 30, 41–42, 47–49, 107–108, 122–129, 141–142, 168–170, 192, 238, 288–310
- fixed slot 149, 151
- open slot 30, 41, 48, 108–130, 192, 242, 287–311
- slot filler, *see also* filler 4–20, 165–181, 299, 311
- slot (filler) analysis 87, 91, 165–170, 181, 292–304
- variable slot 149
- snowclone 10, 11
- Spanish 1, 8, 15, 20, 29–49, 139, 146, 170, 218, 265–284
- surprise 16, 18, 61–75, 191–211
- term 215–232
- complex term 218
- multiword term 15, 18, 215–232
- multiword term extraction 216, 224–231
- term variation 216, 220, 226, 231
- translation 200, 259–260, 268–271, 275–284
- literal translation 87, 111–114, 255, 259
- loan translation 171, 235, 238
- machine translation 56, 108, 268, 271
- translation equivalent 200, 210–211
- translation tools 271
- verbal pattern's translation 277
- word for word translation 280
- unit 166–168, 175, 184–185, 237–238, 266–269
- multiword unit, *see also* phraseme and phraseologism 148, 170, 266, 290
- phraseological unit, *see also* phraseme and phraseologism 1–19, 31–35, 48, 86, 99, 107, 109, 143, 158, 165, 172–173, 184, 191–192, 215, 226, 265, 268–269, 273, 274
- formulaic unit, *see also* phraseme and phraseologism 196
- partially filled unit 149–154
- variant 12–13, 30, 216–220, 222–223, 226–227, 308–310
- idiom variant 30
- individual variant 12
- intensifying variant 69
- lax variant 293–294
- multiword term variant 218–231
- variant cluster 220, 222, 231
- variant extraction 220, 227, 231
- variant shape 225–231
- variant typology 231
- variation 216–223, 230–232, 251–252, 292
- ad hoc variation 219
- diachronic variation 238, 242
- diatopic variation 37
- intra- and cross-writer variation 238
- lexical variation 133, 154
- morphological variation 218–219
- non-diasystemic variation 217, 231
- prosodic variation 214
- subject variation 299
- synonymous variation 216, 222
- syntactic variation 147, 218–219
- systematic variation 13
- term variation 215–231
- variation typology 281, 230
- verb 73–74, 274–280, 301–310
- complex verb 274
- filler verb 306–307
- phrasal-prepositional verb 274
- prepositional verb 274–275
- semantic verb class 304
- syntagmatic verb 275
- verb-adjective-combination 274
- verb slot 290, 292, 298
- wordplay, *see also* modification and creative play of idioms 13