



EURALEX XIX
Congress of the
European Association
for Lexicography

Lexicography for inclusion

7-9 September 2021
Virtual

www.euralex2020.gr

Proceedings Book
Volume 2

Edited by Zoe Gavriilidou, Lydia Mitits, Spyros Kiosses

EURALEX Proceedings

ISSN 2521-7100

ISBN 978-618-85138-2-2

Published by: SynMorPhoSe Lab, Democritus University of Thrace

Komotini, Greece, 69100

e-edition

Publication is free of charge

Edited by: Zoe Gavriilidou, Lydia Mitits, Spyros Kiosses

English Language Proofreading: Lydia Mitits and Spyridon Kiosses

Technical Editor: Kyriakos Zagliveris



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

2021 Edition

Verbal multiword expressions: a preliminary study on the fixedness degree, application to Modern Greek and French

Constant M., Fotopoulou A.

Université de Lorraine Nancy France, Institute for Language and Speech Processing Athens Greece
mathieu.constant@univ-lorraine.fr, afotop@athenarc.gr

Abstract

Multiword expressions display multidimensional properties and a varying degree of compositionality. In this paper, we show a preliminary study to systematically characterize multiword expression types using a set of lexical, morphosyntactic and semantic features, in order to identify their fixedness degree. In particular, we built two sample lexical databases of 100 verbal (mainly emotion) multiword expressions for French and for Modern Greek, systematically encoding these features. We then explore the correlation between semantic features and lexical/morphosyntactic features, in order to better understand the link between lexical and morphosyntactic fixedness and semantic compositionality. This pilot study opens an interesting path of lexicographic research that would consist in systematically exploring a larger spectrum of linguistic features and of types of multiword expressions.

Keywords: multiword expressions; degree of fixedness; modelling and encoding MWEs

1 Introduction

Multiword expressions (MWEs) are combinations of several lexical items that display some idiosyncrasy at one or several linguistic levels. They cover a large number of linguistic phenomena (Sag et al. 2001), including nominal and verbal idioms, support verb constructions, phrasal verbs, complex grammatical words, named entities. They have been the focus of a wide body of research both in linguistics and natural language processing for some decades. As an example, we can mention the European research network PARSEME (COST Action 2013 - 2017) that brought together researchers of multiple scientific fields, enabling substantial progress in modelling and processing MWEs.

This paper aims at presenting a preliminary work on the identification of the fixedness degree of verbal multiword expressions. We specifically focus on Modern Greek and French as verbal MWE databases exist with a large coverage of encoded syntactic and semantic properties, that are organized in a lexicon-grammar (M. Gross 1986). Our study is built on the work of M. Gross (1982) for French and the work of Fotopoulou (1993) and Mini (2009) for Modern Greek. Our objective is to build a model representing the different fixedness cases (also for lexicographic purposes), by means of a spectrum of linguistic features. The verbal MWEs in these databases are treated as elementary sentences for which all possible fixed and non-fixed (or variable) arguments (if any) are consistently and uniformly encoded. The MWE structures are represented as part-of-speech sequences. Selectional restrictions over the non-fixed or variable elements of MWEs as well as syntactic phenomena (i.e., clitic and passive alternation, etc) - if any - are also encoded formally. Finally, other grammatical phenomena such as agreement features (in person and number) are accounted for.

The article is structured as follows. First, it provides some background on multiword expressions (section 2). Section 3 describes the linguistic data, the methodology as well as the criteria used for our experiments. Section 4 presents the encoding of the sample lexical databases. And finally, we provide a detailed analysis of the obtained results (section 5).

2 Background

Linguistic studies in the literature often characterize MWEs by their fixedness and non-compositionality. There are these multiword expressions whose idiomatic meaning cannot be deduced from the meaning of their parts (Fraser 1970; Bobrow & Bell. 1973; Swinney & Cutler 1979; Chomsky 1980; Gross 1982; Van der Linden 1992). In this case, one cannot derive the idiomatic meaning of the idiom *bite the dust* (i.e. cease to exist) based on the meanings of the words *bite*, *the*, *dust*. More precisely, there are three criteria that can help identify them. These criteria define the “fixedness” of these expressions: (a) the semantic criterion: the meaning of an expression is holistic or non-compositional. This means that its meaning cannot be derived from the regular combination of its constituents; (2) the lexical criterion: at least one of the expression’s constituents does not have (or nearly so) paradigmatic variation; (3) the morpho-syntactic criterion: at least one of the constituents cannot be treated as it would be in free sentences, because constraints for example on the determiner or on the transformations, e.g. in passive (Lamiroy 2003). These criteria, however, do not always apply uniformly, and the observed variability leads to the notion ‘degree of fixedness’ (Gross 1996).

Nunberg (1979) introduced the notion of idioms as combining expressions and, in these terms, semantic compositionality refers to idiom elements that “carry identifiable parts of the idiomatic meaning” (Nunberg et al. 1994: 496). Based on this notion, while distinguishing semantically decomposable idioms (*spill the beans*) from non-decomposable ones (*kick the*

bucket), he supports that idioms vary in function of their semantic decomposability. What differentiates them is the interaction between the literal and figurative meanings of their parts. In the example *pop the question*, there is a clear association between *pop* and *question*, and their corresponding parts of the figurative meaning ‘propose marriage’ (normally decomposable idioms), whereas in the MWE *meet your maker*, the word *maker* makes a metaphorical reference to a deity (abnormally decomposable idioms). It is, however, to be noted that an expression may be considered fixed or less fixed depending on the theoretical framework adopted and the criteria used.

From a more computational linguistics perspective, the degree of compositionality of MWEs is often characterized by numerical scores: for instance, using scales from 0 to n (McCarthy et al. 2003; Reddy et al. 2011; Roller et al. 2013; Ramisch et al. 2016) with several individual judgements by MWE type, or based on binary judgments (Farahmand et al. 2015). The encoding can also be categorial like in (Gurrutxaga & Alegria 2013) where the expressions are classified in three categories (idiomatic, collocational and free combination).

Concerning the European research network PARSEME (COST Action 2013 - 2017), various surveys have been produced on either their representation, their grammatical modeling, and their processing (Sailer & Markantonatou 2018; Parmentier & Waszczuk 2019). One of the greatest outcomes is a multilingual corpus annotated in verbal MWEs for 20 languages relying on unique guidelines based on decision diagrams integrating precise linguistic tests (Savary et al. 2017; Ramisch et al. 2018).

3 Linguistic data - methodology - criteria

For our study, we constructed a small set of 65 verbal multiword expressions for both languages. This list of modern Greek and French MWEs that mainly pertain to the semantic field of *emotions* was manually compiled from data listed in existing lexicon-grammar tables for Greek and French. The extracted expressions have different syntactic structures in order to test the fixedness degree with respect to a series of tests. These tests are those which, in principle, enable to define a fixed expression like more and more studies show such as in M. Gross (1982), G. Gross (1996), Lamiroy (2003), Vincze (2011), Sailer & Wintner (2014), Stone (2015):

1. **Lexical criteria:** Fixedness can be identified by testing whether there exists a paradigmatic break on each lexical elements, e.g. while having *Max casse (le jouet+le verre) à Marie* with the compositional meaning *Max breaks Marie's (toy+glass)* (lit. *Max breaks the (toy+glass) to Marie*), we get *Max a cassé (les pieds +*le jouet+*le verre) à Marie* with the meaning ‘*Max gets on Marie's nerves*’ (lit. *Max broke the (feet +*toy+*glass) to Marie*); These criteria enable the evaluation of the exclusive co-occurrence of the expression components.

2. **Morphosyntactic criteria:** non-regular restrictions apply on the determiner distribution or on the morphological variants (e.g. number), as well as on some transformations like passivation or pronominalization. For instance, the verbal expression *Μου κόπηκαν τα ήπατα* (lit. *I have the livers cut*) with the meaning ‘*I was very frightened*’ do not allow any modification over the fixed constituents: **Μου κόπηκε το ήπαρ* (lit. *I have the liver cut*).

3. **Semantic criteria:** traditionally, the main criterion is the following: the meaning of the expression is non-compositional, i.e. it is not predictable from the meaning of its components. But generally speaking, the meaning of a verbal expression can emerge from different combinatorics. Consequently, to detect whether an expression is semantically compositional or not, we have examined independently (and in correlation) the verb and the nominal arguments of the expression similarly to Mini et al. (2011), with tests like:

- i. the element keeps its literal meaning,
- ii. the element has a metaphorical meaning or is an extension of the literal meaning.
- iii. the element or the meaning of the whole sequence has nothing to do with the literal meaning

Mini et al. (2011) examined in each expression the type of relationship between the verb and the nominal arguments. They studied whether the configuration verb + nominal complements in a given expression is unique and specific to derive the global meaning, or the given elements, on the contrary, always keep the same meaning by combining with other elements. In the example *Luc nage dans le bonheur* [FR] (lit. *Luc swims in happiness*), meaning ‘*Luc is happy*’ the word *nage* (swims) is not related to its literal meaning but it is a metaphorical usage (case ii) and the word *bonheur* (happiness) has its literal meaning (case i). On the other hand, *lui casser les pieds* [FR] (lit. *him (Ppv) to-break the feet*), meaning ‘*get on his nerves*’ the words *casser* (break) and *pieds* (feet) are not related to their literal meaning (case iii.). On the contrary, in the example *μου ράγισε την καρδιά* [GR] (lit. *he broke my heart*) the verb retains the core meaning of ‘*cracking without being cut into pieces*’ that it has in the sentence *ράγισε το ποτήρι* (“the glass cracked”) by a semantic extension to a non-tangible/abstract level (case ii). On the other hand, the noun *καρδιά* (“heart”) is a semantically autonomous constituent of the sentence since it refers to the inner emotional world *η καρδιά μου* ‘*my heart cracked*’ (case ii). In example *Marie monte sur ses grands chevaux* [FR] (lit. *Marie (climbs/rises/gets) on her big horses*) the meaning of

the verb *monter* and its complement *ses grands chevaux* has nothing to do with the meaning *Marie gets on her big/high horse* (case iii).¹

We also used standard tests to identify support verb constructions (Gross 1988; Fotopoulou 1992). They are specific multiword expressions as the nominal element keeps its literal meaning, while the verb has a neutral value. We examined three criteria: reduction in a noun phrase with deletion of the support verb, nominalization of the expression *prendre une décision* (lit. *take a decision*, with the meaning *make a decision*) = *prise de décision* (meaning *decision making*), unique relation of the nominal element to the subject. We considered that encoding such tests would contribute to modeling the fixedness degree.

4 Encoding

The expressions are encoded in a table: a row corresponds to a lexical entry (a multiword expression), a column corresponds to a lexical, syntactic or semantic feature like in M. Gross (1986). More precisely, the features encoded are:

- the source lexicon-grammar table in French and modern Greek from which each individual MWE has been extracted,
- the lexical values of the components of the MWEs at every syntactic position,
- valency properties of the verbal expressions, including its arguments and their distributional properties; for instance, the French MWE *casser les pieds* (lit. *break the feet*, ‘*get on one’s nerve*’) has two arguments: a non-restricted subject and a human complement introduced by the preposition *à* (to),
- properties on the lexicalisation of the MWE components: non-substitution by a semantic neighbour (i.e. paradigmatic break), non-deletion,
- Syntactic study of the whole sentence: pronominalisation, passive/ergative transformation, support-verb construction,
- Semantic study of the individual lexical elements of the expressions: literal, metaphoric, or meaning extension,
- Global meaning of the sentence: non-compositional- metaphoric/extension

A sample of these tables are provided in Table 1 for French and in Table 2 for Modern Greek. The description of feature labels and their possible values are defined in table 3. Note that, for clarity, the features given in this table constitute a subset of all encoded properties:

- the list of lexical components at various syntactic positions: V (verb), P (preposition), D (determiner), A (adjective), N (noun),
- some lexical, morphosyntactic and semantic features regarding the lexical MWE components (cf. table 3),
- the glosses and translations in English of the MWEs

We do not show, for instance, the selected prepositions, the list of lexical elements (if any) that can substitute the MWE elements when there is no paradigmatic break using the lexical criteria, and other syntactic properties (ex. pronominalization).

MWE					LEX		MORPHOSYNT				SEM			GLOSS	translation
V	P	D	A	N	L1	L2	M1	M2	M3	M4	S1	S2	S3		
essuyer		une		insulte	+	+	+	+	+	+	1	1	1	endure an insult	receive an insult
trembler	de			peur	+	+	-	+	+	/	0	1	0	shake of fear	be terrified
remuer		l’		âme	+	+	+	+	+	+	0	1	0	stir the soul	touch
briser		le		coeur	-	+	+	+	-	+	0	0	0	breat the heart	break one’s heart

¹ According Mini (2009) and Mini et al. (2011), the combinatorics of the above criteria led to two general groups: non-compositional/typical (*τα φόρτωσε στο κόκορα* (literally *He loaded them on the rooster*) meaning ‘*I did not act at all, as I was feeling lazy*’ and compositional/non typical expressions that can be divided into quasi-typical expressions such as *ράγισε η καρδιά μου*, (literally *my heart cracks*) meaning ‘*I am in deep grief*’ or ‘*it broke my heart*’ and the conventional expressions such as *μου άνοιξαν νεοί ορίζοντες* (literally *they opened new horizons to me*).

monter	à	la		tête	-	+	+	-	-	/	-1	0	0	climb to the head	got to one's head
monter	sur	Poss	grands	chevaux	-	-	+	-	-	/	-1	-1	-1	climb on Poss big horses	get on Poss high horses
casser		les		pieds	-	-	-	-	-	-	-1	-1	-1	break the feet	get on one's nerves

Table 1: Sample of the French MWE table.

MWE					LEX		MORPHOSYNT				SEM			GLOSS	translation
V	P	D	A	N	L1	L2	M1	M2	M3	M4	S1	S2	S3		
ανοίγω				δρόμο (για)	-	-	+	-	-	-	1	0	0	The road is open for...	There are no obstacles left for
χάνω		τα		λόγια Poss-0	+	-	-	-	-	+	0	1	0	N loses his words	be speechless (with emotion)
πάγωσε		το		αίμα Poss-0	-	-	+	-	-	/	0	0	0	My blood freezes	I was terrified
άναψαν		τα		λαμπάκι α Poss-0	-	-	-	-	-	/	-1	-1	-1	They lit the lamps	I was very angry
δεν μου καίγεται				καρφί	-	-	-	-	-	/	-1	-1	-1	Not care a scrap	'I couldn't care less'

Table 2: Sample of the Greek MWE table

CAT	Label	Description	Value
LEX	L1	possible substitution of the noun	true (+) or false (-)
	L2	possible substitution of the verb	true (+) or false (-)
MORPHO-SYN	M1	the determiner is flexible	true (+), false (-) or a class label (ex: Poss = possessive determiner)
	M2	possible ellipsis on the verb	true (+) or false (-)
	M3	possible ellipsis on the noun	true (+) or false (-)
	M4	possible passivation	true (+), false (-) or N/A (/)
SEM	S1	semantic meaning of the verb	literal (1), metaphorical (0) or no relationship (-1)
	S2	semantic meaning of the noun	literal (1), metaphorical (0) or no relationship (-1)
	S3	semantic meaning of the MWE	literal (1), metaphorical (0) or no relationship (-1)

Table 3: Sample of feature label descriptions

5 Results and discussions

Results show a large variety of behaviors. In general, we can observe a correlation between the different types of criteria in the extreme cases of the continuum between entirely non-compositional and almost free expressions. We also observe a grey zone with some unexpected behaviors. A traditional assumption considers that the more an expression does not accept lexical substitutions, the more it does not accept syntactic transformations like passivation, the more it has a non-compositional meaning. However, in some cases, this is not true. For instance, in French, *casser du sucre sur le dos* (lit. *break some sugar on the back*), meaning 'talk about someone behind her/his back' accepts passivation, while it does not accept any substitution for its lexical components.

In order to better visualize the correlation between lexical/morphosyntactic features and semantic ones, we decided to provide a numerical score for each of the two types of features. To do so, we first replaced each non-numeric value in the tables by numeric ones using the following rule: a + symbol is assigned a +1 value, a - symbol is assigned a -1 value, other symbols are given a 0 value, except when it is not appropriate (/ symbol in the tables). Note that for the semantic features, a given lexical element of a given expression is associated with the value 1 when it has its literal meaning, 0 when it has its metaphoric meaning, -1 when it has no relation with the literal meaning. The global score for a type of feature for each individual MWE is the average of the scores of all encoded features of this type.

For instance, the expression *briser le coeur* in Table 1 encodes two - values (-1) and four + values (+1) regarding lexical/morphosyntactic features. Therefore, by averaging the corresponding numerical values, it reaches a score of 0.33 (2/6). Regarding semantic features, the average of the corresponding values (i.e. 0 for the three semantic tests) is 0. Thus, the expression *briser le coeur* is associated with the pair of values (0.33,0). It is the same for Greek expressions like *πάγωσε το αίμα μου* (litt. *My blood freezes*) ‘*I was terrified*’. In this case, lexical and morphosyntactic features are encoded with four - values and one + value, leading to the score -0.6 (-3/5). The semantic encoding includes three 0 values, leading to a 0 score. The expression then corresponds to the pair (-0.6,0).

Such an approach means that we consider that all features have the same weight. Thereafter, we consider that the global score for the lexical/morphosyntactic features corresponds to the degree of lexical/morphosyntactic fixedness, and the global score for the semantic features corresponds to the degree of semantic non-compositionality. For the two examples above, *briser le coeur* with a positive value (0.33) tends to be flexible from a lexical and morphosyntactic point of view. The greek expression *πάγωσε το αίμα μου* with a negative value (-0.6) tends to show lexical and morphological fixedness. Results on French are given in the bubble chart in Figure 1. The horizontal axis (resp. vertical axis) corresponds to the global score of the lexical/morphosyntactic features (resp. semantic features). The bubble size depends on the number of multiword expression entries having the corresponding pair of scores: the more numerous the bigger.

The score pairs corresponding to entirely frozen and non-compositional expressions are positioned in the lower-left corner such as *casser les pieds* with only - values (-1) for lexical/morphosyntactic tests and -1 values for semantic tests. It is the same for Greek expressions like *δεν μου καίγεται καρφί* (lit. *care a scrap*) ‘*I couldn't care less*’, which only encodes - values both for semantic and lexical/morphosyntactic features, showing it is semantically non-compositional. The score pairs corresponding to almost free and compositional expressions are positioned in the upper-right corner such as *essuyer une insulte* ‘*receive an insult*’. The results seem to confirm the existence of a grey zone in between the extreme cases, as one can observe a continuum between the extreme cases, i.e. the part of the graph in between the extreme cases is fully filled with bubbles.

Furthermore, the chart overall shows a rough linear correlation between the degree of lexical/morphosyntactic fixedness and the degree of semantic non-compositionality. Bubbles tend to be in the lower-left and upper-right parts of the chart, tending to show that the more the expression is lexically and morpho-syntactically flexible the more it is semantically compositional.

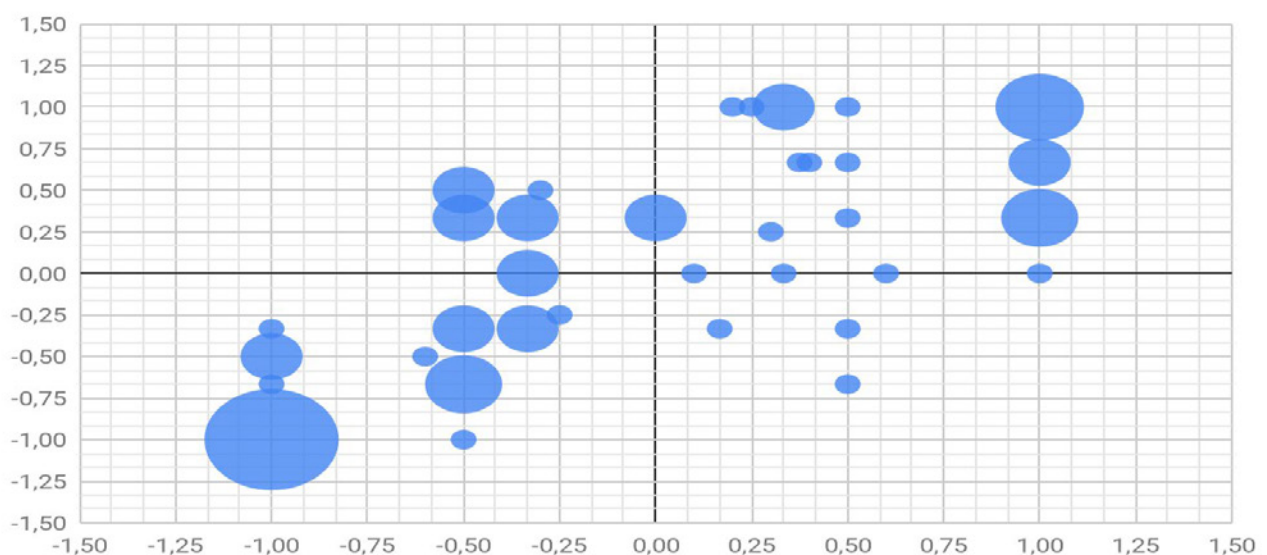


Figure 1: Bubble chart showing the correlation between lexical/morphosyntactic features and semantic ones for French. The horizontal axis (resp. the vertical -axis) corresponds to the global score of the lexical/morphosyntactic features (resp. semantic features). The bubble size depends on the number of multiword expression entries having a given pair of scores.

6 Conclusion

This paper presented a methodology to encode the fixedness degree of verbal multiword expressions that is characterized by a set of lexical, morphosyntactic and semantic features. This pilot study with a specific focus on two languages and a limited set of multiword expressions has shown that formal lexical and morphosyntactic properties tend to approximate semantic compositionality degree though this correlation is somewhat unclear in a 'grey zone'.

Future work should consist in extending this encoding to a larger set of multiword expressions for both languages. This methodology can be applied to other languages, in the same way as it has been applied in the lexicon-grammar methodology for multiple languages. We also plan to compare such an approach with statistical and distributional methods.

7 References

- Bobrow, A., & Bell, S.S.M. (1973). On catching on to idiomatic expressions. *Memory and Cognition*, 1, pp. 343-346.
- Chomsky, N. (1980). Rules and Representations. Columbia Classics in Philosophy. Columbia University Press.
- Fotopoulou, A. (1992). Dictionnaires électroniques des phrases figées. Traitement d'un cas particulier: phrases figées/phrases à *Vsup*. In *COMPLEX. Papers in Computational Lexicography*. Hungarian Academy of Sciences, pp. 147-161.
- Fotopoulou, A. (1993). *Une classification des phrases à compléments figés en grec moderne - étude morphosyntaxique des phrases figées*. Ph.D. thesis, Université Paris VIII - St. Denis.
- Fraser, B. (1970). Idioms within a Transformational Grammar. *Foundations of language*, 6(1), pp. 22-42.
- Farahmand, M., Smith, A. & Nivre, J. (2015). A multiword expression data set: Annotating non-compositionality and conventionalization for english noun compounds. In *Proceedings of the 11th Workshop on Multiword Expressions*, pp. 29-33
- Gross, M. (1982). Une classification des phrases "figées" du français. *Revue Québécoise de Linguistique*, 11(2), pp. 151-185.
- Gross, M. (1986). Lexicon-Grammar. The Representation of Compound Words. In *Proceedings of the 11th conference on Computational linguistics (COLING '86)*.
- Gross, M. (1988). Les limites de la phrase figée. *Langages*, 90, pp. 7-22.
- Gross, G. (1996). *Les expressions figées en français. Noms composés et autres locutions*. Ophrys.
- Gurrutxaga, A. & Alegria, I. (2013). Combining different features of idiomaticity for the automatic classification of noun+verb expressions in Basque. In *Proceedings of the 9th Workshop on Multiword Expressions*, pp. 116-125, Atlanta, Georgia, USA, June. Association for Computational Linguistics.
- Lamiroy, B. (2003). Les notions linguistiques de figement et de contrainte. *Linguisticae Investigationes*, 26(1), pp. 53-66.
- McCarthy, D., Keller, B. & Carroll, J. (2003). Detecting a continuum of compositionality in phrasal verbs. In *Proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, 18, pp. 73-80, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Mini, M., Diakogiorgi, K. & Fotopoulou, A. (2011). What can children tell us about idiomatic phrases' fixedness: the psycholinguistic relevance of a linguistic model". *DISCOURS 9 (Revue de linguistique, psycholinguistique et informatique)*.
- Mini, M. (2009). *Linguistic and Psycholinguistic Study of Fixed Verbal Expressions with Fixed Subject in Modern Greek: A Morphosyntactic Analysis, Lexicosemantic Gradation and Processing by Elementary School Children*. Unpublished doctoral dissertation. Ph.D. thesis, University of Patras.
- Nunberg, G., Sag, I. A. & Wasow T. (1994). Idioms. In Stephen Everson, editor, *Language*, pp. 491-538. Cambridge University Press.
- Nunberg, G. (1979). The non-uniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy*, 3, pp. 43-184.
- Parmentier, Y., & Waszczuk, J. (2019). Representation and Parsing of Multiword Expressions: Current trends. *Language Science Press, Phraseology and Multiword Expressions*.
- Ramisch, C., Cordeiro, S.R., Zilio, L., Idiart, M., Villavicencio, A., Wilkens, R. (2016). "How Naked is the Naked Truth? A Multilingual Lexicon of Nominal Compound Compositionality", *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Berlin, Germany.
- Ramisch, C., Cordeiro, S.R., Savary, A., Vincze, V., Barbu Mititelu, V., Bhatia, A., Buljan, M., Candito, M., Gantar, P., Giouli, V., Güngör, T., Hawwari, A., Nurrieta, U., Kovalevskait'e, J., Krek, S., Lichte, T., Liebeskind, Ch., Monti, J., Parra Es-carl in, C., QasemiZadeh, B., Ramisch, R., Schneider, N., Stoyanova, I., Vaidya, A. & Walsh, A. (2018). Edition 1.1 of the PARSEME Shared Task on Automatic Identification of Verbal Multi-word Expressions. In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG 2018)*, Santa Fe, New Mexico, USA, August. Association for Computational Linguistics.
- Reddy, S., McCarthy, D., & Manandhar. S. (2011). An empirical study on compositionality in compound nouns. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pp 210-218, Chiang Mai, Thailand, November. Asian Federation of Natural Language Processing.
- Roller, S., Schulte S. im Walde. & Scheible, S. (2013). The (un)expected effects of applying standard cleansing models to human ratings on compositionality. In *Proceedings of the 9th Workshop on Multiword Expressions*, pp 32-41, Atl Santa, Georgia, USA, June. Association for Computational Linguistics.

- Sag, I., Baldwin, T., Bond, F., Copestake, A. & Flickinger, D. (2001). Multiword Expressions: A Pain in the Neck for NLP. In *Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2002)*, pp. 1–15.
- Sailer, M. & Wintner S. (2014). Multiword expressions: linguistic properties and lexical representation. *Slides from the PARSEME Prague Training School*.
- Sailer, M. & Markantonatou S. (2018). Multiword expressions: insights from a multi-lingual perspective. *Phraseology and Multiword Expressions series*. Language Science Press.
- Savary, A., Ramisch, C., Cordeiro, S., Sangati, F., Vincze, V., QasemiZadeh, B., Candito, M., Cap, F., Giouli, V., Stoyanova, I., & Doucet, A. (2017). The PARSEME Shared Task on Automatic Identification of Verbal Multiword Expressions. In *Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017)*, pp. 31–47, Valencia, Spain, April. Association for Computational Linguistics.
- Stone, M. S. (2015). Systematic flexibility in Verb-Object Idioms. In *Proceedings of the 8th Brussels Conference on Generative Linguistics*.
- Swinney, D. & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior*, 18, pp. 523–534.
- Van der Linden, E.J. (1992). Incremental processing and the hierarchical lexicon. *Computational Linguistics*, 18, pp. 219–238.
- Vincze, V. (2011). *Semi-Compositional Noun + Verb Constructions: Theoretical Questions and Computational Linguistic Analyses*. Ph.D. thesis, University of Szeged, August.

Acknowledgments

This work has been partially funded by the French Agence Nationale pour la Recherche, through the PARSEME-FR project (ANR-14-CERA-0001).