The role of constituents in multiword expressions

An interdisciplinary, cross-lingual perspective

Edited by

Sabine Schulte im Walde

Eva Smolka

Phraseology and Multiword Expressions 4

language science press

Phraseology and Multiword Expressions

In this series:

1. Manfred Sailer & Stella Markantonatou (eds.). Multiword expressions: Insights from a multilingual perspective.

ISSN: 2625-3127

The role of constituents in multiword expressions

An interdisciplinary, cross-lingual perspective

Edited by

Sabine Schulte im Walde

Eva Smolka

language
science
press

Schulte im Walde, Sabine & Eva Smolka (ed.). 2019. *The role of constituents in multiword expressions: An interdisciplinary, cross-lingual perspective* (Phraseology and Multiword Expressions 4). Berlin: Language Science Press.

This title can be downloaded at:
http://langsci-press.org/catalog/book/239

ISBN: 978-3-96110-184-9 (Digital)
978-3-96110-185-6 (Hardcover)

ISSN: 2625-3127
no DOI
Source code available from www.github.com/langsci/239
Collaborative reading: paperhive.org/documents/remote?type=langsci&id=239

Cover and concept of design: Ulrike Harbort
Fonts: Linux Libertine, Libertinus Math, Arimo, DejaVu Sans Mono
Typesetting software: XeLaTeX

Language Science Press
Unter den Linden 6
10099 Berlin, Germany
langsci-press.org

Storage and cataloguing done by FU Berlin

Contents

The role of constituents in multiword expressions: An interdisciplinary, cross-lingual perspective

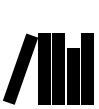
Sabine Schulte im Walde
Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart

Eva Smolka
Fachbereich Sprachwissenschaft, Universität Konstanz

1 Introduction

The processing and representation of multiword expressions (MWEs), ranging from noun compounds (such as *nickname* in English, and *Ohrwurm* in German) to complex verbs (such as *give up* in English, and *aufgeben* in German) and idiomatic expressions (such as *break the ice* in English, and *das Eis brechen* in German) has remained an unsettled issue over the past 20+ years.

Our research question concerns semantically transparent MWE combinations as well as MWE combinations that result in a meaning shift. For example, in the absence of situational experience, even complex verbs that appear to be fully semantically transparent such as *aufstehen* ('stand up') do not necessarily have whole-word meanings that are easily predictable from their constituents. Even more difficult are complex verbs such as *verstehen* ('understand') and *zustehen* ('legally due'), which contain only a remote resemblance to the meaning of *stehen* ('stand'). Similarly, the constituents of noun compounds do not necessarily contribute to their whole-word meanings in a straightforward way. The meaning contribution may range from relatively semantically transparent as in *Familienname* ('family name') to semantically opaque, as in *Spitzname* ('nickname', lit.



Sabine Schulte im Walde & Eva Smolka. 2019. The role of constituents in multiword expressions: An interdisciplinary, cross-lingual perspective. In Sabine Schulte im Walde & Eva Smolka (eds.), *The role of constituents in multiword expressions: An interdisciplinary, cross-lingual perspective*, iii–xi. Berlin: Language Science Press. DOI:??

'pointy name') or *Zwickmühle* ('dilemma', lit. 'pinch mill'), which contain a modifier (i.e. the left constituent) or a head (i.e. the right constituent) that renders the compound semantically more opaque. The most extreme meaning shifts across types of MWEs occur in idiomatic constructions, such as *kick the bucket* and *reach for the stars*, where the literal meanings of the constituents do not seem to contribute to the overall figurative meanings 'die' and 'strive for something unachievable' at all. MWEs of the idiomatic type are typically assumed to be semantically opaque, even though some idioms like *spill the beans* are stronger in reflecting the figurative meaning ('reveal a secret') in a metaphoric way than others.

This edited volume exploits complementary evidence across different types of MWEs to shed light on the interaction of constituent properties and meanings of MWEs. Specialists across languages and across research disciplines contribute to this issue and provide a cross-linguistic perspective integrating linguistic, psycholinguistic, corpus-based and computational studies.

2 Contributions

In the following, the six contributions in this volume discuss MWEs that are composed of different types of constituents, including the combination of particle+stem in complex verbs (e.g., *aufstehen*, 'stand up'), the combination of stem+stem and stem+preposition+stem in noun compounds (e.g., *nickname, government assessment, juego de niños*), the combination of modifier+stem in modifier-noun phrases (e.g., *the brown dog*) and idiomatic combinations of words (e.g., *reach for the stars*).

Sections 2.1 to 2.3 discuss the interdisciplinary perspectives separately for complex verbs, noun compounds and idiomatic expressions, and for each of these three categories of MWEs we summarise the contributions to this collection.

2.1 Complex verbs

Seminal psycholinguistic studies have applied the manipulation of semantic transparency to study whether verbal MWEs of the type prefix+stem, particle+stem and stem+suffix are lexically represented and processed via the constituents or as a whole-word unit (e.g., Taft & Forster 1975; Marslen-Wilson et al. 1994; Longtin et al. 2003).

Recurrent findings in English and French showed that semantically transparent words facilitate their base (e.g., *distrust–trust, confessor–confess*). This facilitation result, however, was not obtained for semantically opaque primes (e.g.,

retreat–treat, successor–success). Lexicon-based models concluded from these findings that a semantically transparent word like *confessor* possesses a lexical entry that corresponds to its base and is represented as the stem (*–confess–*) and suffix (*–or*), whereas *successor* is represented in its full form (e.g., Rastle et al. 2000; Feldman et al. 2004; Diependaele et al. 2005; Meunier & Longtin 2007; Marslen-Wilson et al. 2008; Diependaele et al. 2009; Taft & Nguyen-Hoan 2010).

Semantic transparency effects emerge also when transparency is manipulated in a more graded way (Gonnerman et al. 2007): Strong facilitation effects showed for strongly phonologically and semantically related word pairs (e.g., *preheat–heat*), intermediate effects for moderately similar pairs (e.g., *midstream–stream*), and no priming for low semantically related word pairs (*rehearse–hearse*). Within learning-based approaches, such as the convergence-of-codes account, form and meaning relatedness between word pairs determines lexical processing (Plaut & Gonnerman 2000; Gonnerman et al. 2007).

Findings in German, however, indicate that lexical processing in German occurs via the stem and irrespective of semantic transparency (i.e., meaning composition of the complex verb). Low semantically related word pairs (*entwerfen–werfen*, 'design–throw') induced facilitation of the stem to the same extent as semantically related word pairs did: *bewerfen–werfen*, 'throw at'–'throw' (e.g., Smolka et al. 2009; 2014; 2015; 2018). Most importantly, these findings stress the importance of cross-language comparisons: what is true for the processing in one language is not necessarily true for the processing in another language (Günther et al. 2018).

Computational approaches regarding the meanings of complex verbs have mainly focused on predicting the degree of transparency of complex verbs. These approaches typically rely on the distributional hypothesis (Harris 1954; Firth 1957) and empirical co-occurrence information from large corpora, and are realised as vector space models (Turney & Pantel 2010). Regarding English, computational approaches explored variants of distributional models and distributional similarity, comparing word-based and syntax-based descriptions, large-scale vs. dimensionality-reduced representations, and verb-specific vs. general information (Baldwin et al. 2003; McCarthy et al. 2003; Bannard 2005; Cook & Stevenson 2006: i.a.). Regarding German, an initial series of papers (Aldinger 2004; Schulte im Walde 2004; 2005; 2006) studied particle verbs from a large-scale corpus-based perspective, with an emphasis on salient distributional features at the syntax-semantics interface. Schulte im Walde (2006) and Bott & Schulte im Walde (2018) integrated the subcategorisation transfer of German particle verbs with respect to their base verbs into models of compositionality. Kühner & Schulte im Walde (2010); Bott & Schulte im Walde (2017); Köper & Schulte im Walde (2017a) used

clustering to distinguish between multiple senses, and common cluster membership to determine compositionality. Köper & Schulte im Walde (2016) and Aedmaa et al. (2018) applied classifiers to identify non-literal language usage of German and Estonian particle verbs in context.

So far, most approaches that have dealt with complex verbs –across disciplines and across languages– have considered semantic transparency as the meaning relation between the whole word meaning of the MWE and the meaning of its base constituent, disregarding the contribution of the often ambiguous prefix or particle, e.g., they were concerned with the question: to what degree is the meaning of *stand* reflected in *understand*? Apart from a series of formal word-syntactic analyses in the framework of *Discourse Representation Theory* (Kamp & Reyle 1993) for German particle verbs with the particles *auf* (Lechler & Roßdeutscher 2009), *ab* (Kliche 2011), *an* (Springorum 2011) and *nach* (Haselbach 2011), this gap of knowledge has recently been addressed from experimental perspectives: Frassinelli et al. (2017) demonstrated in a lexical decision experiment that the particle *an* in German particle verbs is primarily associated with a horizontal directionality, while *auf* is primarily associated with a vertical directionality. Schulte im Walde et al. (2018) and Köper & Schulte im Walde (2018) present data collections to assess meaning components in German complex verbs. The former dataset contains source- and target-domain characterisations of the base verbs and the complex verbs, respectively, and a selection of arrows to add spatial directional information to user-generated contexts; the latter dataset contains ratings for strengths of particle-related pairs of German base verbs and particle verbs.

As part of the present collection, **Springorum and Schulte im Walde** also focus on the meaning contribution of the particle to the overall meaning of German particle verbs. They combine nine particles (e.g., *auf* 'up') with 30 base verbs (e.g., *geben* 'give') and examine how the particles are perceived in adding directionality (e.g., up, down, left, right) to the meaning of the particle verb (e.g., *aufgeben* 'give up'). That is, the participants in their study saw a base verb or a particle verb and decided which type of directionality in form of two-dimensional arrows best reflects the verb meaning. Their qualitative and quantitative analyses indicate that the particles exhibit individual spatial profiles, but also that the particles vary in their flexibility to provide predominant directions, in interaction with the abstractness of the semantic base verb domains.

2.2 Noun compounds

Also compounds lie on a continuum between relatively transparent and rather opaque with respect to the meaning of their constituents. Psycholinguistic re-

search so far has been intrigued by the question whether the compound is lexically represented and processed via the constituents or as a whole-word unit. For example, findings on the processing of noun-noun-compounds indicate a competition between the compound's constituents that correspond to independent words and their whole-word counterparts. Hence, upon seeing the compound *doughnut*, the constituent [nut] may compete with the whole word *nut* (e.g., Libben 2006; Frisson et al. 2008; Monahan et al. 2008; Fiorentino & Fund-Reznicek 2009; Gagné & Spalding 2009; 2014; Libben 2014). Another question concerns whether the semantic transparency of the constituents affect the processing of the MWE they compose, and if so, how? Indeed, semantically opaque compounds are generally processed more slowly than semantically transparent ones, and are less likely to show constituent activation – probably because the semantic opacity of the whole compound makes its constituents less relevant to lexical comprehension (e.g., Taft & Forster 1976; Sandra 1994; Zwitserlood 1994; Isel et al. 2003; Libben et al. 2003). Furthermore, recent studies indicate that the influence of semantic transparency is language-specific. The semantic transparency of the head has been found to affect the processing of noun-noun compounds in English and Italian (e.g., Marelli et al. 2009; Marelli & Luzzatti 2012) but not in German (e.g., Smolka & Libben 2017).

Computational approaches to predicting the transparency of noun compounds can be subdivided into two subfields: (i) approaches that aim to predict the *meaning* of a compound by composite functions, relying on the vectors of the constituents (e.g., Mitchell & Lapata 2010; Coecke et al. 2011; Baroni et al. 2014; Hermann 2014); and (ii) approaches that aim to predict the *degree of compositionality* of a compound, typically by comparing the compound vectors with the constituent vectors (e.g., Reddy et al. 2011; Salehi & Cook 2013; Schulte im Walde et al. 2013; Salehi et al. 2014a,b; 2015; Schulte im Walde et al. 2016; Köper & Schulte im Walde 2017b). As for complex verbs, the computational models typically rely to a large extent on the distributional hypothesis and empirical co-occurrence information from large corpora. Individual research studies noticed differences in the contributions of modifier and head constituents towards the composite functions predicting compositionality (Reddy et al. 2011; Schulte im Walde et al. 2013), but only a very limited number of approaches zoomed into potentially relevant properties of MWEs and their constituents, such as ambiguity, frequency, productivity (Bell & Schäfer 2016; Schulte im Walde et al. 2016).

In this collection, **Pezzelle and Marelli** apply a distributional semantic model to show that the semantic properties of the compound and its constituents may explain syntactically-based classes of compounds as suggested in linguistic theories (Bisetto & Scalise 2005). They differentiate between types of compounds

such as subordinate, attributive, and coordinate compounds, on the basis of the underlying syntactic relation between the compound constituents. In particular, Pezzele and Marelli provide measures that quantify (a) the degree of semantic similarity between the constituents, and (b) the contribution of each constituent to the overall compound meaning, and show that these semantic measures are effective in capturing the different syntactic linguistic classes. In other words, the continuous quantitative semantic aspects of the meanings of compounds parallel the discrete qualitative grammatical distinctions between compounds.

Iordăchioaia, van der Plas, and Jagfeld study the compositionality of English deverbal compounds. These deverbal nouns are ambiguous between compositionally interpreted "Argument Structure Nominals", which inherit verbal structure and realise arguments (e.g., *assessment of the budget by the government*), and more lexicalized "Result Nominals", which preserve no verbal properties or arguments (e.g., *budget assessment*), cf. Grimshaw (1990). While the former are fully compositional, the latter remain ambiguous because the non-head (*budget*) can be interpreted as either subject or object. The authors apply machine-learning techniques to evaluate corpus data and human annotations to support their hypothesis and find that different properties of the head contribute to the interpretation of the deverbal compound.

Two papers deal with MWEs that are untypical compound constructions for which linguistic theories in general refer to the notions of lexicon and syntax and debate whether these MWEs are to be considered as compounds or not. **Hennecke** examines the formation of MWEs of the type "N Prep N" in Romance languages, such as Spanish, French and Portuguese (e.g. *juego de niños*, 'kid's game') and takes a constructionist approach to analyse the constructions as abstract templates. In a qualitative analysis, she examines the variation that the preposition in a construction may undergo (e.g. *juego de niños* vs. *juego para niños*, both meaning 'kid's game'). To this end, she analyses the semantic relations between the nominal constituents and the semantic transparency of the constructions. Her findings indicate that variability of the prepositional element occurs only in semantically transparent constructions. Furthermore, prepositional variability largely varies across the three Romance languages.

Also **Gagné, Spalding, Burry, and Adams** examine MWEs that are not typically classified as compounds and compare modifier-noun phrases (e.g., *the brown dog*) with full phrases (e.g., *the dog that was brown*). They examine how modifying information that refers to recently encountered information is used in the production of MWEs and manipulate the property of the head noun between normal (e.g., *brown*) and distinctive (e.g., *blue*). Participants showed a strong overall bias

toward using a modifier-noun phrase structure (regardless of whether they previously saw a modifier-noun phrase or a full phrase), and were more likely to include distinctive properties (*the blue dog*) than normal properties (*the brown dog*) when referring to the concept. These findings indicate that modifier-noun phrases have a privileged status among MWEs and provide a good compromise between conveying sufficient information and using simple syntactic structures.

2.3 Idioms

Idiomatic expressions are the MWEs which may be considered as showing the strongest semantic shift that the constituents undergo, because the figurative meaning is usually not even remotely connected with the meaning of its constituents, as in *hit the road*. Rather, idiomatic expressions are considered semantically fixed, since the figurative meaning does not allow the replacement of any of the word constituents (e.g., **she hit the street; *she beat the road*), and the modification of an idiomatic constituent is assumed to change the figurative meaning into a literal meaning.

The processing and representation of idioms has thus remained an unsettled issue in psycholinguistic research: How is the figurative meaning processed and stored in lexical memory? In particular, is the figurative meaning of an idiom represented separately from the meaning of its constituents, and how is the figurative meaning assembled (e.g., Cacciari & Tabossi 1988; Gibbs Jr. 1992; Cacciari & Glucksberg 1994; Titone & Connine 1999; Hamblin & Gibbs Jr. 2003). Seminal studies thus assumed a "non-compositional" representation in which the whole figurative meaning of an idiom is stored as a distinct entry in the mental lexicon similar to the representation of a complex word like *Finanzmarktaufsichtsbehörde* 'financial market supervisory authority' (e.g., Bobrow & Bell 1973; Swinney & Cutler 1979; Gibbs Jr. 1980). More recent hybrid models try to integrate the assumption that idioms are both compositional and unitary: On the one hand each, an idiom is composed of single constituents that are activated to some degree, and on the other hand each idiom possesses its own lexical entry that stores the whole meaning of the idiom (e.g., Cacciari & Tabossi 1988; Gibbs Jr. et al. 1992; Cutting & Bock 1997; Titone & Connine 1999; Sprenger et al. 2006; Caillies & Butcher 2007; Holsinger & Kaiser 2013; Titone & Libben 2014).

As far as computational work on idiomatic expressions is concerned, several research studies measured the syntactic flexibility of idiomatic expressions, to a large extent focusing on verb–object combinations (e.g., Bannard 2007; Fazly et al. 2009). These measures varied the constituents of the target MWEs, explored modifiability and passivisation, etc. in order to distinguish between lit-

eral vs. idiomatic interpretations. A large number of automatic classification approaches addressed idioms as non-literal language identification across various types of MWEs, mostly relying on contextual indicators to distinguish between literal and idiomatic interpretations (e.g., Sporleder & Li 2009; Turney et al. 2011; Köper & Schulte im Walde 2016), such as distributional similarity, text cohesion graphs, and contextual abstractness. The variation-based approaches further provide some insight into the flexibility of MWE constituents and their meaning contributions.

The last paper by **Smolka and Eulitz** deals with idioms and how the meaning of the constituents adds to the figurative meaning. They present three experiments, in which participants rate the meaning similarity between an idiomatic phrase (e.g., *She always reached for the stars*) and a paraphrase of its figurative meaning (e.g., *She always strove for something unreachable*). They exchange the noun, verb, or prepositional idiomatic constituent by a close semantic associate (e.g., *She always reached/grasped for/at the stars/planets*) and find that a modified constituent still preserves the figurative meaning. This study adds to the understanding that there is no completely fixed "unitary entry" and that the idiomatic constituents do contribute to the figurative meaning of the idiom, even though the figurative meaning is semantically opaque.

Acknowledgements

This collection was supported by the DFG Collaborative Research Centre SFB 732 and the DFG Heisenberg Fellowship SCHU-2580/1 (Sabine Schulte im Walde), and by the Volkswagen Foundation Grant FP 561/11 (Eva Smolka). Special thanks go to our student researcher Anurag Nigam who type-set this volume. Last but not least we thank our experts from the interdisciplinary fields who ensured a qualitatively high-standing reviewing process:

- Melanie Bell (Anglia Ruskin University, UK)
- Jens Bölte (University of Münster, Germany)
- Paul Cook (University of New Brunswick, Canada)
- Christina Gagné (University of Alberta, Canada)
- Giannina Iordăchioaia (University of Stuttgart, Germany)
- Alessandro Lenci (University of Pisa, Italy)

- Marco Marelli (University of Milano-Bicocca, Italy)
- Carlos Ramisch (Aix-Marseilles University, France)
- Martin Schäfer (University of Jena, Germany)
- Nils Schiller (Leiden University, The Netherlands)
- Thomas Spalding (University of Alberta, Canada)
- Lonneke van der Plas (University of Malta, Malta)
- Aline Villavicencio (Universidade Federal do Rio Grande do Sul, Brazil / University of Essex, UK)

References

Aedmaa, Eleri, Maximilian Köper & Sabine Schulte im Walde. 2018. Combining abstractness and language-specific theoretical indicators for detecting non-literal usage of Estonian particle verbs. In *Proceedings of the naacl 2018 student research workshop*, 9–16. New Orleans, LA, USA.

Aldinger, Nadine. 2004. Towards a dynamic lexicon: Predicting the syntactic argument structure of complex verbs. In *Proceedings of the 4th international conference on language resources and evaluation*, 427–430. Lisbon, Portugal.

Baldwin, Timothy, Colin Bannard, Takaaki Tanaka & Dominic Widdows. 2003. An empirical model of multiword expression decomposability. In *Proceedings of the acl workshop on multiword expressions: analysis, acquisition and treatment*, 89–96. Sapporo, Japan.

Bannard, Colin. 2007. A measure of syntactic flexibility for automatically identifying multiword expressions in corpora. In *Proceedings of the acl workshop on "A broader perspective on multiword expressions"*, 1–8. Prague, Czech Republic.

Bannard, Collin. 2005. Learning about the meaning of verb–particle constructions from corpora. *Computer Speech and Language* 19. 467–478.

Baroni, Marco, Raffaella Bernardi & Roberto Zamparelli. 2014. Frege in space: A program for compositional distributional semantics. *Linguistic Issues in Language Technologies* 9(6). 5–110.

Bell, Melanie J. & Martin Schäfer. 2016. Modelling semantic transparency. *Morphology* 26. 157–199.

Bisetto, Antonietta & Sergio Scalise. 2005. The classification of compounds. *Lingue e Linguaggio* 4(2). 319–332.

Bobrow, Samual A. & Susan M. Bell. 1973. On catching on to idiomatic expression. *Memory and Cognition* 1. 343–346.

Bott, Stefan & Sabine Schulte im Walde. 2017. Factoring ambiguity out of the prediction of compositionality for German multi-word expressions. In *Proceedings of the 13th workshop on multiword expressions*, 66–72. Valencia, Spain.

Bott, Stefan & Sabine Schulte im Walde. 2018. German particle verbs: Compositionality at the syntax-semantics interface. *Journal of Language Modelling* 6(1). 41–86.

Cacciari, Cristina & Sam Glucksberg. 1994. Understanding figurative language. In Morton Ann Gernsbacher (ed.), *Handbook of Psycholinguistics*, 447–477.

Cacciari, Cristina & Patrizia Tabossi. 1988. The comprehension of idioms. *Memory and Language* 27(6). 668–683.

Caillies, Stephanie & Kirsten Butcher. 2007. Processing of idiomatic expressions: Evidence for a new hybrid view. *Memory and Language* 22(1). 79–108.

Coecke, Bob, Mehrnoosh Sadrzadeh & Stephen Clark. 2011. Mathematical foundations for a compositional distributional model of meaning. *Linguistic Analysis* 36(1-4). 345–384.

Cook, Paul & Suzanne Stevenson. 2006. Classifying particle semantics in English verb-particle constructions. In *Proceedings of the acl/coling workshop on multiword expressions: identifying and exploiting underlying properties*, 45–53. Sydney, Australia.

Cutting, J. Cooper & Kathryn Bock. 1997. That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory and Cognition* 25(1). 57–71.

Diependaele, Kevin, Dominiek Sandra & Jonathan Grainger. 2005. Masked cross-modal morphological priming: Unravelling morpho-orthographic and morpho-semantic influences in early word recognition. *Language and Cognitive Processes* 20(1–2). 75–114.

Diependaele, Kevin, Dominiek Sandra & Jonathan Grainger. 2009. Semantic transparency and masked morphological priming: The case of prefixed words. *Language and Cognitive Processes* 37(6). 895–908.

Fazly, Afsaneh, Paul Cook & Suzanne Stevenson. 2009. Unsupervised type and token identification of idiomatic expressions. *Computational Linguistics* 35(1). 61–103.

Feldman, Laurie Beth, Emily G. Soltano, Matthew John Pastizzo & Sarah E. Francis. 2004. What do graded effects of semantic transparency reveal about morphological processing? *Brain and Language* 90. 17–30.

Fiorentino, Robert & Ella Fund-Reznicek. 2009. Masked morphological priming of compound constituents. *The Mental Lexicon* 4(2). 159–193.

Firth, John R. 1957. *Papers in linguistics 1934-51*. London, UK: Longmans.

Frassinelli, Diego, Alla Abrosimova, Sylvia Springorum & Sabine Schulte im Walde. 2017. *Meaning (mis-)match in the directionality of German particle verbs*. Poster at the 30th Annual CUNY Conference on Human Sentence Processing. MIT, Cambridge.

Frisson, Steven, Elizabeth Niswander-Klement & Alexander Pollatsek. 2008. The Role of Semantic Transparency in the Processing of English Compound Words. *British Journal of Psychology* 99(1). 87–107.

Gagné, Christina L. & Thomas L. Spalding. 2009. Constituent Integration during the Processing of Compound Words: Does it involve the Use of Relational Structures? *Journal of Memory and Language* 60. 20–35.

Gagné, Christina L. & Thomas L. Spalding. 2014. Conceptual Composition: The Role of Relational Competition in the Comprehension of Modifier-Noun Phrases and Noun-Noun Compounds. *The Psychology of Learning and Motivation* 59. 97–130.

Gibbs Jr., Raymond W. 1980. Spilling the beans on understanding and memory for idioms in conversation. *Memory and Cognition* 8. 149–156.

Gibbs Jr., Raymond W. 1992. What do idioms really mean? *Memory and Language* 31. 485–506.

Gibbs Jr., Raymond W., Nandini P. Nayak & Cooper Cutting. 1992. How to kick the bucket and not decompose: Analyzability and idiom processing. *Memory and Language* 28. 576–593.

Gonnerman, Laura M., Mark S. Seidenberg & Elaine S. Andersen. 2007. Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology. *Journal of Experimental Psychology: General* 136(2). 323–345.

Grimshaw, Jane. 1990. *Argument Structure*. Cambridge, MA, USA: MIT Press.

Günther, Fritz, Eva Smolka & Marco Marelli. 2018. "Understanding" differs between English and German: Capturing systematic language differences of complex words. *Cortex* 11(10). DOI https://doi.org/10.1016/j.cortex.2018.09.007.

Hamblin, Jennifer L. & Raymond W. Gibbs Jr. 2003. Processing the meanings of what speakers say and implicate. *Discourse Processes* 35. 59–80.

Harris, Zellig. 1954. Distributional structure. *Word* 10(23). 146–162.

Haselbach, Boris. 2011. Deconstructing the meaning of the German temporal verb particle *"nach"* at the syntax-semantics interface. In *Proceedings of generative grammar in geneva*, 71–92. Geneva, Switzerland.

Hermann, Karl Moritz. 2014. *Distributed Representations for Compositional Semantics*. University of Oxford dissertation.

Holsinger, Edward & Elsi Kaiser. 2013. Processing (non) compositional expressions: Mistakes and recovery. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 39(3). 866.

Isel, Frédéric, Thomas C. Gunter & Angela D. Friederici. 2003. Prosody-assisted head-driven access to spoken German compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29(2).

Kamp, Hans & Uwe Reyle. 1993. *From Discourse to Logic*. Kluwer Academic Publishers.

Kliche, Fritz. 2011. Semantic variants of German particle verbs with *"ab"*. *Leuvense Bijdragen* 97. 3–27.

Köper, Maximilian & Sabine Schulte im Walde. 2016. Distinguishing literal and non-literal usage of German particle verbs. In *Proceedings of the conference of the north american chapter of the association for computational linguistics: human language technologies*, 353–362. San Diego, California, USA.

Köper, Maximilian & Sabine Schulte im Walde. 2017a. Applying multi-sense embeddings for German verbs to determine semantic relatedness and to detect non-literal language. In *Proceedings of the 15th conference of the european chapter of the association for computational linguistics*, 535–542. Valencia, Spain.

Köper, Maximilian & Sabine Schulte im Walde. 2017b. Complex verbs are different: Exploring the visual modality in multi-modal models to predict compositionality. In *Proceedings of the 13th workshop on multiword expressions*, 200–206. Valencia, Spain.

Köper, Maximilian & Sabine Schulte im Walde. 2018. Analogies in complex verb meaning shifts: The effect of affect in semantic similarity models. In *Proceedings of the 16th annual conference of the north american chapter of the association for computational linguistics: human language technologies*, 150–156. New Orleans, LA, USA.

Kühner, Natalie & Sabine Schulte im Walde. 2010. Determining the degree of compositionality of German particle verbs by clustering approaches. In *Proceedings of the 10th conference on natural language processing*, 47–56. Saarbrücken, Germany.

Lechler, Andrea & Antje Roßdeutscher. 2009. German particle verbs with *"auf"*. Reconstructing their composition in a DRT-based framework. *Linguistische Berichte* 220. 439–478.

Libben, Gary. 2006. Why study compounds? An overview of the issues. In Gary Libben & Gonia Jarema (eds.), *The representation and processing of compound words*, 1–21. Oxford: Oxford University Press.

Libben, Gary. 2014. The Nature of Compounds: A Psychocentric Perspective. *Cognitive Neuropsychology* 31(1–2). 8–25.

Libben, Gary, Martha Gibson, Yeo Bom Yoon & Dominiek Sandra. 2003. Compound fracture: the role of semantic transparency and morphological headedness. *Brain and Language* 84. 50–64.

Longtin, Catherine-Marie, Juan Segui & Pierre A. Halle. 2003. Morphological priming without morphological relationship. *Language and Cognitive Processes* 18. 313–334.

Marelli, Marco, Davide Crepaldi & Claudio Luzzatti. 2009. Head position and the mental representation of nominal compounds – A constituent priming study in italian. *The Mental Lexicon* 4(3). 430–454.

Marelli, Marco & Claudio Luzzatti. 2012. Frequency effects in the processing of Italian nominal compounds: Modulation of headedness and semantic transparency. *Journal of Memory and Language* 66(4). 644–664.

Marslen-Wilson, William D., Mirjana Bozic & Billi Randall. 2008. Early decomposition in visual word recognition: Dissociating morphology, form, and meaning. *Language and Cognitive Processes* 23(3). 394–421.

Marslen-Wilson, William D., Lorraine K. Tyler, Rachelle Waksler & Lianne Older. 1994. Morphology and meaning in the English mental lexicon. *Psychological Review* 101. 3–33.

McCarthy, Diana, Bill Keller & John Carroll. 2003. Detecting a continuum of compositionality in phrasal verbs. In *Proceedings of the acl workshop on multiword expressions: analysis, acquisition and treatment*, 73–80. Sapporo, Japan.

Meunier, Fanny & Catherine-Marie Longtin. 2007. Morphological decomposition and semantic integration in word processing. *Journal of Memory and Language* 56(4). 457–471.

Mitchell, Jeff & Mirella Lapata. 2010. Composition in Distributional Models of Semantics. *Cognitive Science* 34. 1388–1429.

Monahan, Philip J., Robert Fiorentino & David Poeppel. 2008. Masked Repetition Priming using Magnetoencephalography. *Brain and Language* 106. 65–71.

Plaut, David C. & Laura M. Gonnerman. 2000. Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes* 15(4–5). 445–485.

Rastle, Kathleen, Matthew H. Davis, William Marslen-Wilson & Lorraine Komisarjevsky Tyler. 2000. Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes* 15(4–5). 507–537.

Reddy, Siva, Diana McCarthy & Suresh Manandhar. 2011. An empirical study on compositionality in compound nouns. In *Proceedings of the 5th international joint conference on natural language processing*, 210–218. Chiang Mai, Thailand.

Salehi, Bahar & Paul Cook. 2013. Predicting the compositionality of multiword expressions using translations in multiple languages. In *Proceedings of the 2nd joint conference on lexical and computational semantics*, 266–275. Atlanta, GA, USA.

Salehi, Bahar, Paul Cook & Timothy Baldwin. 2014a. Detecting non-compositional MWE components using wiktionary. In *Proceedings of the conference on empirical methods in natural language processing*, 1792–1797. Doha, Qatar.

Salehi, Bahar, Paul Cook & Timothy Baldwin. 2014b. Using distributional similarity of multi-way translations to predict multiword expression compositionality. In *Proceedings of the 14th conference of the european chapter of the association for computational linguistics*, 472–481. Gothenburg, Sweden.

Salehi, Bahar, Paul Cook & Timothy Baldwin. 2015. A word embedding approach to predicting the compositionality of multiword expressions. In *Proceedings of the conference of the north american chapter of the association for computational linguistics/human language technologies*, 977–983. Denver, Colorado, USA.

Sandra, Dominiek. 1994. The morphology of the mental lexicon: Internal word structure viewed from a psycholinguistic perspective. *Language and Cognitive Processes* 9(3). 227–269.

Schulte im Walde, Sabine. 2004. Identification, quantitative description, and preliminary distributional analysis of German particle verbs. In *Proceedings of the coling workshop on enhancing and using electronic dictionaries*, 85–88. Geneva, Switzerland.

Schulte im Walde, Sabine. 2005. Exploring features to identify semantic nearest neighbours: A case study on German particle verbs. In *Proceedings of the international conference on recent advances in natural language processing*, 608–614. Borovets, Bulgaria.

Schulte im Walde, Sabine. 2006. Can human verb associations help identify salient features for semantic verb classification? In *Proceedings of the 10th conference on computational natural language learning*, 69–76. New York City, NY.

Schulte im Walde, Sabine, Anna Hätty & Stefan Bott. 2016. The role of modifier and head properties in predicting the compositionality of English and German noun-noun compounds: A vector-space perspective. In *Proceedings of the 5th joint conference on lexical and computational semantics*, 148–158. Berlin, Germany.

Schulte im Walde, Sabine, Maximilian Köper & Sylvia Springorum. 2018. Assessing meaning components in German complex verbs: A collection of source–target domains and directionality. In *Proceedings of the 7th joint conference on lexical and computational semantics*, 22–32. New Orleans, LA, USA.

Schulte im Walde, Sabine, Stefan Müller & Stephen Roller. 2013. Exploring vector space models to predict the compositionality of German noun-noun compounds. In *Proceedings of the 2nd joint conference on lexical and computational semantics*, 255–265. Atlanta, GA, USA.

Smolka, Eva, Matthias Gondan & Frank Rösler. 2015. Take a stand on understanding: Electrophysiological evidence for stem access in German complex verbs. *Frontiers in Human Neuroscience* 9(62).

Smolka, Eva, Sarolta Komlosi & Frank Rösler. 2009. When semantics means less than morphology: The processing of German prefixed verbs. *Language and Cognitive Processes* 24(3). 337–375.

Smolka, Eva & Gary Libben. 2017. "Can you wash off the hogwash?" – Semantic transparency of first and second constituents in the processing of German compounds. *Language, Cognition and Neuroscience* 32(4). 514–531.

Smolka, Eva, Gary Libben & Wolfgang U. Dressler. 2018. When morphological structure overrides meaning: Evidence from German prefix and particle verbs. *Language, Cognition, and Neuroscience*. DOI: https://doi.org/10.1080/23273798.2018.1552006.

Smolka, Eva, Katrin H. Preller & Carsten Eulitz. 2014. 'Verstehen' ('understand') primes 'stehen' ('stand'): Morphological structure overrides semantic compositionality in the lexical representation of German complex verbs. *Journal of Memory and Language* 72. 16–36.

Sporleder, Caroline & Linlin Li. 2009. Unsupervised recognition of literal and non-literal use of idiomatic expressions. In *Proceedings of the 12th conference of the european chapter of the acl*, 754–762. Athens, Greece.

Sprenger, Simone A., Willem J.M. Levelt & Gerard Kempen. 2006. Lexical access during the production of idiomatic phrases. *Journal of Memory and Language* 54(2). 161–184.

Springorum, Sylvia. 2011. DRT-based analysis of the German verb particle *"an"*. *Leuvense Bijdragen* 97. 80–105.

Swinney, David A. & Anne Cutler. 1979. The access and processing of idiomatic expression. *Journal of Verbal Learning and Verbal Behavior* 18. 523–534.

Taft, Marcus & Kenneth I. Forster. 1975. Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior* 14. 638–648.

Taft, Marcus & Kenneth I. Forster. 1976. Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior* 15. 607–620.

Taft, Marcus & Minh Nguyen-Hoan. 2010. A sticky stick? The locus of morphological representation in the lexicon. *Language and Cognitive Processes* 25(2). 277–296.

Titone, Debra & Cynthia Connine. 1999. On the compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics* 31. 1655–1674.

Titone, Debra & Gary Libben. 2014. Time-dependent effects of decomposability, familiarity and literal plausibility on idiom priming: A cross-modal priming investigation. *The Mental Lexicon* 9(3). 473–496.

Turney, Peter D., Yair Neuman, Dan Assaf & Yohai Cohen. 2011. Literal and metaphorical sense identification through concrete and abstract context. In *Proceedings of the conference on empirical methods in natural language processing*, 680–690. Edinburgh, UK.

Turney, Peter D. & Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research* 37. 141–188.

Zwitserlood, Pienie. 1994. The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes* 9. 341–368.

Chapter 1

Aiming with → arrows ← at particles: Towards a conceptual analysis of directional meaning components in German particle verbs

Sylvia Springorum
Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart

Sabine Schulte im Walde
Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart

This article presents a case study on the contributions of prepositional particles to the meanings of German particle verbs (such as *anstrahlen* 'to beam/smile at' and *aufgeben* 'to give up'). Based on a set of 16 "concept images", two-dimensional directional arrow pictographs, 60 experiment participants selected one or more concept images for a systematically composed set of 270 German particle verbs and their 30 base verbs. We formulate a series of hypotheses for the meanings of nine constituent particle types (*ab, an, auf, aus, ein, mit, nach, vor, zu*) and investigate them in the light of the concept image selections. Qualitative and quantitative analyses indicate that our hypotheses are largely confirmed, across three source domains varying in their abstractness (`Machines & Tools`, `Force`, `Sound`), as well as across well-known vs. unknown particle verbs. The particles exhibit individual concept image profiles, and they vary in their flexibility to provide predominant directions; for example, while *auf* is rather consistently perceived as contributing an upward right direction to a particle verb meaning, *an* shows similarly strong preferences for a set of concept images; in both cases, these tendencies are observed across source domains.

Sylvia Springorum & Sabine Schulte im Walde. 2019. Aiming with → arrows ← at particles: Towards a conceptual analysis of directional meaning components in German particle verbs. In Sabine Schulte im Walde & Eva Smolka (eds.), *The role of constituents in multiword expressions: An interdisciplinary, cross-lingual perspective*, 1–33. Berlin: Language Science Press. DOI:??

1 Introduction

German particle verbs (PVs) are complex verb structures such as *anstrahlen* 'to beam/smile at' that combine a prefix particle (*an*) with a base verb (*strahlen* 'to beam').[1] PVs represent a type of multiword expressions, which are generally known as a *"pain in the neck for NLP"* (Sag et al. 2002). Even more, German PVs pose a specific challenge, because the particles are highly ambiguous; e.g., the particle *an* has a partitive meaning in *anbeißen* 'to take a bite', a cumulative meaning in *anhäufen* 'to pile up', and a topological meaning in *anbinden* 'to tie to' (Springorum 2011). In addition, the particles often trigger meaning shifts of the base verbs (BVs), cf. Springorum, Utt, et al. (2013); Frassinelli et al. (2017); Köper & Schulte im Walde (2018); Schulte im Walde et al. (2018); e.g., the PV *abschminken* with the BV *schminken* 'to put on make-up' has a literal meaning in a concrete context ('to remove make-up'), as in example (1), and a metaphorical meaning in an abstract context ('to forget about something'), as in example (2).

(1) Den Lippenstift kannst Du Dir abschminken.
the lipstick can you yourself [ab] put on make-up
'You can remove the lipstick.'

(2) Den Job kannst Du Dir abschminken.
the job can you yourself [ab] put on make-up
'You can forget about the job.'

Not only the particle types but also the particle verbs as a whole often have more than a single reading. For example, the PV *anstrahlen* not only means 'to beam at' but also 'to smile at', when derived from the metaphorical meaning of *strahlen* ('to beam'), i.e. 'to smile'. The PV *abnehmen* not only means 'to take off/away', but can also be used to express 'to reduce' as an incremental interpretation of 'to take off/away'; in addition, it has obtained the specific sense 'to reduce weight'. The semantic decomposition in the latter two examples seems to be less transparent than in the previous ones, thus indicating different degrees of PV compositionality. Accordingly, we also find opaque compositions such as *aufhören* ('to stop'), where the semantics of the BV *hören* ('to hear') does not seem to provide any contribution to the PV meaning at all. Such examples are the reason why PV composition is often deemed idosyncratic, cf. Kratzer (2003).

[1]German distinguishes between separable particle verbs and non-separable prefix verbs. Our focus is on particle verbs.

In this chapter, we explore the meaning contribution of particle types to the meanings of German particle verbs across three semantic domains of base verbs, which vary in their degree of abstractness: `Machines & Tools, Force, Sound`. Within our study, we focus on prepositional particle types and the role of directionality. In this vein, Section 2 will motivate our assumptions about particle meanings in German PVs in more detail, before Section 3 presents the design, hypotheses and results of an experiment that collected human judgements on directionality in particle meanings. Section 4 discusses the experiment data and reflects on our preceding assumptions about prepositional particle meanings.

2 Particle meanings

2.1 Basic particle meanings and contexts

For the course of this article, we assume that each particle type has a restricted number of simple primary meanings, which we refer to as *basic meanings*. This is in accordance with Lindner (1983), who identifies a prototypical sense for the English verb particle *out* involving 'paths in the spatial domain'. Without a BV context, the basic particle meanings are underspecified first, and then resolved by contextual constraints provided by the BV. For example, the separation introduced by the particle *ab* in the context of the BV *nehmen* ('to take') evokes a change of state 'to take off/away', whereas in the context of the BV *schminken* ('to put on make up') it evokes a duration 'to remove make up' generated by a sequence of separations. However, not only the BV but also further context has to be taken into account, as there are ambiguous PVs with varying particle meaning contributions. For example, regarding the metaphorical meaning of *abschminken* in example (2), *ab* introduces only a single separation event, in contrast to the sequence of separation events in the literal PV reading.

Previous research has pointed out regularities in the interpretation of particle meanings associated with semantically coherent classes of base verbs, cf. Stiebels (1996); Lechler & Roßdeutscher (2009); Kliche (2011); Springorum (2011). For example, Direction and Contact represent two independent readings of *an*, among others: The PV in example (3) belongs to the Direction meaning class, suggesting that *an* assigns a direction to the BV, whereas in example (4) the PV carries a Contact particle meaning. In combination with a movement BV as in example (5), the particle again introduces a direction. In addition, the meaning of *anfahren* requires a decreasing distance, which results in a contact when maximal. Therefore, *anfahren* represents an example with meaning components from both classes, Direction and Contact. Examples (3–5) show that particle

senses vary in their complexity, and they also illustrate the limits of a hard class assignment.

(3) Karin schaut das Bild an.
Karin look the picture [an]
'Karin looks at the picture.'

(4) Karin klebt die Briefmarke an.
Karin stick the stamp [an]
'Karin sticks on the stamp.'

(5) Karin fährt die Laterne an.
Karin drive the lantern [an]
'Karin drives against the lantern.'

In addition, a classification of PVs should not only take lexical information into account. Sentimental connotations, associations to other sensory input, (nature) forces, and dimensionality are just as well involved in the process of sense development. For example, the metaphorical PV *abklappern* (lit. [ab]+'to clatter') illustrates that sensory information can be understood as a part of the PV meaning: *abklappern* creates an ideophone, which is mapped to the verb event, and leads to the meaning-shifted sense 'to pursue something successively', as illustrated by example (6). This perception-based meaning shift process is discussed in more detail by Springorum, Utt, et al. (2013).

(6) Sie klapperte die Geschäfte nach tollen Büchern ab.
she clattered the shops for great books [ab]
'She successively searched through the shops for great books.'

Particle meaning is not only influenced by its context, but also provides an influence on the meaning of the context. For example, participants in a sentence generation experiment relying on systematically created PV neologisms (neoPVs) were asked to generate sentences for neoPV types such as *antöten* ([an]+'to kill') and *abschlafen* ([ab]+'to sleep'), without being provided any context (Springorum, Schulte im Walde, et al. 2013). The participants did not only show considerable agreement regarding potential neoPV meanings, but also often agreed in their strategy of dealing with particle senses, in cases where the BV meaning did not fit to the PV senses, as in the case of *antöten*, where 'to kill' introduces an absolute change of state, and the generated sentences mainly suggested

an *an* meaning of Partial Affectedness, thus introducing quantification over event parts. This meaning typically cannot be applied to a verb with an absolute change of state, such as *töten*, but the participants obviously re-conceptualised the change-of-state BV *töten* as a process verb, which gradually approximates the final state of death. Often, adverbial specifications such as *fast tot* ('nearly dead') were added, which supported the above assumptions. The meaning components in the PV based on the BV were thus adjusted dependent on the particle meaning.

In sum, we define the meaning of a PV as either a direct composition of possible meaning components of particle and BV (if they are compatible), or alternatively as meaning-shifted particle and BV meaning components in strong interaction with the context. On the one hand, PVs can be assigned to discrete particle classes, based on semantically coherent groups of BVs, but on the other hand the classes need to be flexible, to allow semantic changes if necessary. At first these two alternative options might seem contradictory, but from a diachronic perspective they reflect two natural processes of sense development. For example, according to Waldron (1979) "new words should first be used in rather specialised senses and subsequently be generalised" and "when such words have once achieved general status we use them without reflection upon their former restricted or technical sense". In addition, "the reverse process, in which a general word is given a special meaning in a restricted context, is just as common". In this sense, the polysemy of particles is considered as a result of adjustment processes of basic meanings to recurring contextual conditions.

2.2 Spatial grounds of particle meanings

As we are focussing on PVs with prepositional particles, we assume that particles are spatially grounded, similar to preposition meanings. Prepositions indicate spatial fundamentals, as discussed by Herskovits (1986) and Dirven (1993), among others. They structure the physical space and determine "language-specific concepts built up in mental space". Simlarly, Gärdenfors (2004) claims that prepositions are "primarily spatial relations" and create "spatially structured mental representations", when used with non-locational words. In order to structure space, it has to be perceived through our senses, with vision representing the predominant human sense (Viberg 1983).

Furthermore, Jackendoff (1983) understands "perception as an interaction between environmental input and active principles in the mind, that impose structure on that input". He demonstrates his view by ambiguous pictures from the school of Gestalt psychology. Lakoff (1987) refers to the "spatialisation of form

hypothesis" by using the term *Image Schema*, which he defines as "schematic descriptions of meaning concepts". So perception of space cannot be separated from cognitive conceptualisation, and (meaning) concepts are often analogies of structures, to define space through perception. Although there are "significant differences between mental imagery and Image Schemas", according to Gibbs & Colston (1995) there is "good evidence that both spatial and visual representations exist for mental imagery".

We assume that prepositional particles –simlarly to prepositions– introduce relations to structure space and to add verb-related meaning components, such as aspectual or temporal modifications. These relations can be captured by Image Schemas as "dynamic analogue representations of spatial relations as movement in space" (Gibbs & Colston 1995) to describe aspects of PV meaning. Accordingly, earlier investigations connect (spatial) concepts with phrasal verbs. Going beyond the already mentioned work by Lindner (1983), Morgan (1997) provides an extension for metaphorical readings of some *out* phrasal verbs. From a didactic point of view, Side (1990) and Abreu & Vieira (2008) discuss the advantages of using Image Schemas in order to learn phrasal verbs. In a psycholinguistic setting, Richardson et al. (2001) carried out experiments to show that basic images can be related to spatial and abstract verbal meanings.

A semiotic perspective of schematic descriptions is provided by Frutiger (1987), who defines the essential task of a schema as description with the help of literally pictured elements, to divide objects into different parts, instead of only using words. Neurath (1983) developed the international picture language, which is based on pictographs instead of words, as a complementary system to natural language, because in some cases pictures or pictographs are better in transmitting information, as they are more universal than words. Such pictographs transmit conceptual information and therefore act on a universal level. They meet our demands to describe basic particle meaning components, both context- and domain-independent.

3 Experiment

This section presents the material, design, hypotheses and results of the experiment that collected human judgements on spatial aspects in particle meanings.

3.1 Material

3.1.1 Verb data

The German particle verbs for the experiment were generated systematically, based on a pre-selected set of base verbs and a pre-selected set of particles. We relied on base verbs from three different semantic domains, `Machines and Tools (MnT)`, `Force` and `Sound`, which differ regarding their degree of concreteness. Furthermore, Kövecses (2002) categorises `MnT` and `Force` domains as common source domains for metaphors. The verbs belonging to the `MnT` domain (such as *hämmern* ('to hammer') and *schaufeln* ('to dig')) are easy to imagine and represent very concrete BVs. In comparison, the verbs from the `Force` domain (such as *drücken* ('to press') and *quetschen* ('to squeeze')) are less concrete, as the force itself is not perceivable directly, but only through interactions of its concrete entities encoded in the verb arguments. The verbs from the `Sound` domain (such as *schreien* ('to cry') and *jaulen* ('to yowl')) represent intransitive verbs and define the most abstract source domain.

For each of the three domains we chose a total of ten base verbs that we thought as not obviously ambiguous among the three classes, cf. the Appendix. These 30 BVs were then systematically composed to PVs using nine different prepositional particles. We only took into account particles that cannot also be used in German prefix verbs: *ab, an, auf, aus, ein, mit, nach, vor, zu*. In this way, we obtained 300 verbs (30 selected BVs and 270 generated PVs) as target verbs for the experiment. Due to the systematic composition of the PVs, also PV neologisms (neoPVs) were part of this data set. As part of the experiment tasks, the experiment participants were thus asked to rate a PV as a neologism, such that our analyses can distinguish between existing PVs vs. PV neologisms. Approximately half of the PVs were rated as neoPVs (153 out of 270 PVs), cf. Section 3.2 below.

3.1.2 Concept images

Although there are many semantic analyses based on concepts and frequently illustrated by visual schemas or pictographs, as to our knowledge there is no general systematic standard available. We therefore decided to define visual representations for directional concepts from scratch. As source for inspiration, we relied on Dreyfuss' symbol sourcebook, a very detailed collection of various kinds of symbols from many different areas (Dreyfuss 1972), and on a more descriptive sign derivation in Frutiger (1987). We defined the set of directional pictographs as shown in Figure 1. The pictographs were intended to be as simple as possible,

to not distract from the actual information, but at the same time they should allow possibly alternative interpretations. We refer to our simplified pictographs as "concept images".

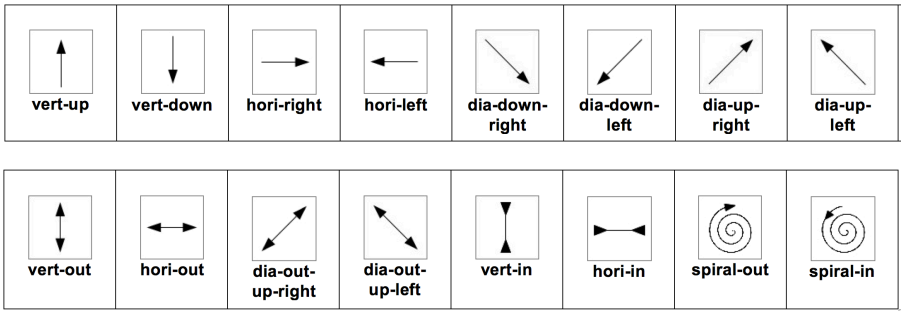


Figure 1: Set of concept images.

Although the number of directions in space is infinite, a simplified conceptual reduction into a two-dimensional setting is in many cases sufficient, because "the salient dimensions of the world reinforce the horizontal and the vertical" (Tversky 2011). We therefore included vertical arrows for upward and downward directions (***vert-up, vert-down***), horizontal right and left directions (***hori-right, hori-left***), and also the four diagonal directions (***dia-down-right, dia-down-left, dia-up-right, dia-up-left***). To represent single object-oriented center-periphery directions as expansion or constriction, lines with arrow heads at both ends are used. Whereas the outward-pointing arrow heads (***vert-out, hori-out, dia-out-up-right, dia-out-up-left***) correspond to expansion, the inward-pointing arrow heads (***vert-in, hori-in***) correspond to constriction. To distinguish between asymmetrical and uniform center-periphery directions, two arrows with concentric curved lines were added (***spiral-out, spiral-in***). The total set of concept images contains 16 pictograms.

3.2 Design

The experiment was performed as follows: The 300 verbs were distributed randomly over 6 lists with 50 verbs each. The random distribution was balanced for BVs vs. PVs, BV source domain, particle type and (non-)neologism[2], such that each file contained equal proportions of these.

[2]At this point, we did not yet have human ratings for PV neologisms, so we used a pre-categorisation which considered a PV as neoPV if it did not appear in the *SdeWaC* web corpus containing 880 million words (Faaß & Eckart 2013).

Each verb was judged by ≈20 participants, non-experts (mostly students on campus), without payment. They were presented a randomly ordered list of the target verbs (printed out or as a file), together with the concept images. For each verb, the participants were first asked to choose between one of the following statements, to check on whether they knew the PVs:

1. unknown and difficult to understand;
2. unknown but easy to understand;
3. infrequent usage but known;
4. frequent usage and known.

These ratings provided a participant-dependent categorisation of PVs (and also of BVs, but those were not relevant for us) into existing PVs vs. PV neologisms on a four-point scale.

Then the participants were asked to mark those concept images which fit to the meaning of the target verb. Multiple marks were allowed while we did not explicitly allow the participants to not select any concept image because we wanted to enforce a selection. However, we offered the participants to describe an alternative image if they decided that none of our concept images fits. In that way they would only fall back to not provide any selection if they really could not settle on a concept image.

3.3 Hypotheses

The main goal of our study was to investigate whether prepositional particles within German particle verbs can be associated with directional concepts, which are visually represented as concept images. As the basis for interpreting the experiment results, this section provides example-based and experience-driven hypotheses for the above-mentioned nine particle types regarding their most prevalent readings. Regarding the particles *ab, an* and *auf*, we in addition rely on detailed formal semantic analyses (Lechler & Roßdeutscher 2009; Kliche 2011; Springorum 2011).

Further than discussing the primary concepts as originating from the spatial domain, we also include time into the interpretation of space, as "knowledge of space frequently comes from motion in time, from exploring environments and piecing together the parts" (Tversky 2011). Furthermore, relying on Boroditsky (2001), who analyses time with the help of spatial metaphors, "concepts of space appear to be primary", concepts of time can be derived from concepts of space.

3.3.1 *ab*

ab has a basic meaning derived from the gravity force that causes objects to fall. This motion describes a DOWN directional meaning which may be represented by the concept image ***vert-down***. An example is the particle verb *ablaufen* ('to run down') in example (7) where the downward meaning can only be contributed by the particle and not by the BV *laufen* ('to run'). In contrast, example (8) with *absinken* ('to sink') the event of the BV *sinken* ('to sink') already introduces a downward direction. The difference between the BV and the PV meanings is that the BV events refer to an atelic continuous downward motion, as arising from the gravity force DOWN direction, while the PV event is resultative, so a direction is spanned between the pre-state and the resulting state of the object affected by the gravity force. That is, in example (8) the pre- and result states are the locations of the ship before and after the sinking motion. The PV meaning is thereby almost synonymous to the BV meaning, which only describes the downward motion of the ship, and in addition introduces a result state.

(7) Das Wasser läuft ab.
the water runs [ab]
'The water runs off.'

(8) Das Schiff sinkt ab.
the ship sink [ab]
'The ship sinks.'

The PV *abfallen* in example (9) also describes a downward direction with pre- and result states, regarding the button affected by gravity. Here we however find a further meaning component: the detachment of the button, a mereological part of the jacket, has to be caused by some force. This example 9 suggests that the particle *ab* may also contribute a SEPARATION meaning, which is –according to previous lexical semantic analyses– a productive reading for this particle (Kliche 2011). Often, it is not gravity but other intentional forces which are causing detachments, as in example (10) with the PV *abreißen* ('to pull off'). Here, the direction related to the force may even overwrite the basic downward direction of *ab*, which means that the particle only contributes the SEPARATION meaning component to the PV. The directions are explicitly specified through the semantics of the BV, through further contextual clues, or remain unspecified as in example (10). In addition to the gravity-dependent default direction described by ***vert-down***, a "neutral", gravity-independent horizontal direction described by ***hori-right*** or ***hori-left*** might therefore represent an alternative choice of concept image.

(9) Der Knopf fällt von der Jacke ab.
the button fall off the jacket [ab]
'The button falls off the jacket.'

(10) Karin reißt den Knopf von der Jacke ab.
Karin pull the button off the jacket [ab]
'Karin pulls the button off the jacket.'

The continuous variant of the discrete Separation reading is the Decrease of Proximity reading which occurs with motion verbs as in example (11). Here the alignment with the conceptual direction of time becomes obvious. Similarly, in sentences as in example (12) with *absitzen* ('to wait/endure', lit. 'to sit off') in an abstract context, the spatially grounded basic concept has to be transferred to available abstract dimensions, which are different from space. Regarding *absitzen*, the abstract context *seminar* belongs to the time domain, so that the direction of the particle *ab* can conceptually only align with the conceptual orientation of the time dimension. The conceptual direction is thereby spanned between the starting point and an iteration of separations of mereological parts, which are time intervals. This leads to a Progress reading which may be combined with a conceptually vertical value scale (Tversky 2011) to a Value Decrease meaning, as in example (13). Combining the Progress and the Value Decrease dimensions, ***dia-down-right*** is another concept to be expected for the particle *ab*. This idea is comparable to Talmy (2000)'s force dynamics, a conceptual notion of forces split up into different components and relations, which can be applied to various domains.

(11) Karin fährt (von Stuttgart) ab.
Karin drive from Stuttgart [ab]
'Karin departs (from Stuttgart).'

(12) Karin sitzt das Seminar ab.
Karin sits the seminar [ab]
'Karin endures the seminar.'

(13) Karin wertet (mit ihrer Kritik) alles ab.
Karin values with her criticism everything [ab]
'Karin devaluates everything (with her criticism).'

3.3.2 *an*

an introduces a direction which is force-independent in its primary meaning, as in example (14), repeated from example (3), with the PV *anschauen* ('to look at') derived from the perception BV *schauen* ('to look') in a spatial context. The direction of human sight –with a neutral head position which is horizontal by default– determines this conceptual direction. Given this, *an* can be represented by the concept images ***hori-right*** and ***hori-left***.

(14) Karin schaut das Bild an.
Karin look the picture [an]
'Karin looks at the picture.'

In contexts with forces, e.g. as derived from motion, the particle *an* contributes an Increase of Proximity reading in analogy to the Decrease of Proximity reading of *ab*. Its direction is aligned with the direction of the goal of the motion expressed by its object to which the proximity is increased, cf. example (15) in comparison to example (11). Due to this goal we expect the concept image ***hori-right*** with the right-pointing arrow, since the future in Western cultures is on a horizontal timeline conceptually located on the right.

(15) Karin fährt Stuttgart an.
Karin drive Stuttgart [an]
'Karin drives towards Stuttgart.'

If the argument represents a concrete object, as the *lantern* in example (16), repeated from example (5), the relation introduced by *an* can be understood as Maximal Proximity, such that there is a Contact situation in the result state of the verb. In addition, we find readings as in example (17) with *anhämmern* ('to attach by hammering'), where the particle *an* introduces a direction orthogonal to the vertical surface of a wall, again enforcing a Contact reading. In comparison, examples (18) and (19) refering to horizontal surfaces –where the direction of *an* needs to be vertical– are only semi-acceptable. This strengthens the assumption that the basic conceptual direction of *an* is horizontal, and that ***hori-right*** and ***hori-left*** will be selected as predominant representations for this reading.

(16) Karin fährt die Laterne an.
Karin drive the lantern [an]
'Karin drives against the lantern.'

(17) Karin hämmert das Bild an die Wand an.
Karin hammer the picture at the wall [an]
'Karin hammers the picture to the wall.'

(18) ?Karin hämmert das Bild an die Tischplatte an.
Karin hammer the picture at the table top [an]
'Karin hammers the picture to the table top.'

(19) ?Karin hämmert das Bild an die Decke an.
Karin hammer the picture at the ceiling [an]
'Karin hammers the picture to the ceiling.'

In example (20) with the PV *anfressen* ('to nibble') the particle *an* introduces a relation that identifies parts of the verbal object which are affected by the verb event. Here, the mouse nibbles only on some parts of the apple, scraping through the surface. Conceptually this is an extension of the MAXIMAL PROXIMITY reading, where the maximum is exceeded and results in a damaged surface of the direct object affected by the verb event. The meaning contributed by *an* is therefore a PARTIAL AFFECTEDNESS relation.

(20) Die Maus frisst den Apfel an.
the mouse nibble the apple [an]
'The mouse nibbles at the apple.'

In intransitive contexts with an abstract verb notion as in example (21) with the PV *anlaufen* ('to start') where the BV *laufen* ('to run') comes with its abstract and unspecific progress sense and therefore conceptually only provides the dimension of time, the particle *an* spans an abstract conceptual direction between the beginning of the time interval and an unspecified point later within this interval. In such cases, the conceptual direction of the particle is resolved to a meaning which refers to EVENT INITIATION.

(21) Der Motor läuft an.
the motor go [an]
'The motor starts.'

In contexts as in example (22) where the BV *heizen* ('to heat') provides a value dimension, the conceptual direction of *an* is not only associated with the time dimension of the verb event but also with the vertical-value heat dimension. This

means that the particle not only introduces the heating event initiation, but also a temperature rise along the timeline. This suggests that ***dia-up-right***, the synthesis of ***hori-right*** and ***vert-up*** is a suitable concept.

(22) Karin heizt den Ofen an.
Karin heat the oven [an]
'Karin heats the oven.'

3.3.3 *auf*

auf's basic meaning represents the upward direction Up, the opposite direction of the basic meaning of *ab* as derived from the directional alignment with the falling motion caused by gravity. That is, *auf*'s basic meaning is the direction derived from motions caused by forces which overcome gravity. This is the case in example (23), where the upward direction is a result of the gravity-countering shooting force.

(23) Das Wasser schießt auf.
The water shoot [auf]
'The water shoots up.'

Overcoming gravity often includes an elevation of an object, where a prominent position is more likely in the field of visual perception of an experiencer. Given this, the particle *auf* is also used to mark a Coming-into-Perception sense as in example (24), where startling birds suddenly become visually perceivable when they lift up from the ground.

(24) Karin schreckt die Vögel auf.
Karin scare the birds [auf]
'Karin startles the birds.'

The spatially derived basic Up meaning can also refer to a sudden increase of noise, volume or pitch, when resolved in a `Sound` source-domain context, as in example (25). This mapping of spatial height to a scale is very productive, and often the particle contributes an Increase meaning as in example (26). Therefore we expect the concept image ***vert-up*** to be associated with this particle.

(25) Karin schreit auf.
Karin cry [auf]
'Karin crys out.'

(26) Karin dreht die Musik auf.
Karin turn the music [auf]
'Karin turns up the music.'

If *auf* appears in contexts where it can only be applied to the time dimension, the spatially derived Up is conceptually spanned between beginning and end of the time interval of the BV event. In this interpretation of the directional concepts the particle covers the whole event time interval (in contrast to *an*'s Event Initiation interpretation, which only covers the first parts of the event time interval but says nothing about the endpoint), so that its semantic contribution is a Completeness reading as in example (27). The event duration is determined by the direct object, as in the consumption of a cookie in example (28). The scale adds a vertical value dimension to the horizontal time notion, measuring the progress and making ***dia-up-right*** a plausible concept image.

(27) Karin arbeitet die Aufgaben der letzen Woche auf.
Karin worked the tasks the last week [auf]
'Karin finishes off the tasks of the last week.'

(28) Karin isst den Keks auf.
Karin eats the cookie [auf]
'Karin eats up the cookie.'

3.3.4 *aus*

aus typically refers to an expansion in the spatial domain, as illustrated by example (29). The growth of an object may also be conceptualised as direction originating from a point within the object, so overall the concepts ***vert-out***, ***hori-out*** as well as ***dia-out-up-right***, ***dia-out-up-left*** and ***spiral-out*** are legitimised.

(29) Das Universum dehnt sich aus.
The universe expand itself [aus]
'The universe expands.'

From an object-extrinsic perspective the particle introduces a specified closed area –conceptually understood as a container– to distinguish between an inside and an outside. With the help of an imaginary container concept, it is possible

to relate our two-dimensional concept images to this particle meaning.[3] The concept image ***hori-right*** represents a plausible concept in order to describe the gravity-independent "default" direction pointing from an inside to an outside area. E.g., in example (30) the concept image ***hori-right*** may indicate the pulling direction from the bed moved out of its box, the imaginary container.

(30) Karin zieht das Schlafsofa aus.
Karin pull the sofa bed [aus]
'Karin opens the sofa bed.'

3.3.5 *ein*

ein can introduce a shrinking or constriction of an object, as in example (31), and therefore be related to the inward-orientated concepts ***vert-in, hori-in*** as well as ***spiral-in***. In analogy to the change from inside to outside described by *aus*, *ein* can also refer to a change from an outside to an inside area, as in example (32). This may be depicted with ***vert-down***, again refering to an imaginary conceptual container representing the transition direction from the outside area to an inside, e.g. through the default opening of a container at the top.

(31) Der Igel rollt sich ein.
the hedgehog roll itself [ein]
'The hedgehog rolls itself up.'

(32) Karin wirft eine Münze in den Automaten ein.
Karin throw a coin in the vending machine [ein]
'Karin throws a coin into the vending machine.'

3.3.6 *mit*

mit introduces a relation between two arguments of which one may be implicit, as in example (33). The particle does not provide additional information regarding these arguments, hence both symmetrical ***hori-in*** and ***hori-out*** concepts, which allow no inferences regarding an imbalance, are assumed possible representations for *mit*.

[3] A more appropriate notion of containers requires a spatial concept with a higher dimensional complexity and is thus going beyond the scope of the current study.

(33) Karin geht (mit ihrer Schwester) in das Schwimmbad mit.
Karin go with her sister in the pool [mit]
'Karin joins her sister to go to the pool.'

3.3.7 *nach/vor*

nach and *vor* introduce orderings in space which are gravity-independent and can therefore describe horizontal relations, suggesting ***hori-left*** and ***hori-right*** as their concepts. The main difference between *nach* and *vor* is their conceptual perspective on the one-dimensional ordering. *nach* focuses on something which can be conceptualised as following, as behind or as an end, cf. example (34), whereas *vor* focuses on a conceptual front or a beginning, as in example (35).

(34) Karin schmeißt ihrem Freund eine Zeitung nach.
Karin throw her boyfriend a newspaper [nach]
'Karin throws a newspaper after her boyfriend.'

(35) Karin drängelt sich vor.
Karin push herself [vor]
'Karin jumps the queue.'

3.3.8 *zu*

zu provides a gravity-independent direction in the spatial domain similar to *an*, and in addition introduces an assignment or an intention. The assignment can be concrete, as in example (36), or abstract, as in example (37), whereas the intention meaning is always abstract, so that the particle's direction also tends to be abstract, as in example (38). We predict that the particle always originates from the spatial domain, and that ***dia-up-right*** therefore represents a plausible concept for this P, because it is a synthesis of ***hori-right***, the default direction, and ***vert-up***, the goal representation. The fulfilment of an intention requires effort, i.e., a force, and therefore presupposes resistance. In analogy to *auf*'s counter-gravity direction, the direction introduced by *zu* is also a counter-direction facing resistance to reach the intended goal. Without further specification and with gravity as the default force to be overcome, the intention to reach a goal can conceptually be described with ***vert-up***.

(36) Karin fährt auf die Stadt zu.
Karin drive up the city [zu]
'Karin drives towards the city.'

(37) Karin ordnet die Telefonnummer Emelie zu.
Karin arrange the phone number Emelie [zu]
'Karin assigns the phone number to Emelie.'

(38) Karin schneidet den Stoff genau nach Plan zu.
Karin cut the fabric exactly after plan [zu]
'Karin cuts the fabric exactly according to the plan.'

3.4 Concept image selections

In this section, we present an overview of the actual selections of concept images by our experiment participants, before Section 4 discusses them in light of the hypotheses just introduced. The dataset is publicly available at http://www.ims.uni-stuttgart.de/data/pv-ci.

3.4.1 Dataset

As mentioned in Section 3.2, the 300 verbs were distributed randomly over 6 lists with 50 verbs each, and each list was judged by ≈20 non-experts. Given that participants might have refrained from judging a verb they did not know, the resulting distribution of the number of participant judgements over verb types differs slightly. Most of the verbs received between 16 and 20 judgements.

In total, we obtained judgements across 5,509 verb instances (including only those instances where at least one concept image had been chosen). Table 1 shows the number of concept images that were selected across verb instances. 3,192 (58%) of the target verb instances were assigned exactly one concept image; 1,556 (28%) received two concept images; 11% received three or four, and 2% were assigned between five and 16 concept images. Abstracting over target verbs to particle types, each of the nine particle types received between 540 and 560 judgements across concept images, i.e., we have a rather homogenous number of concept images across particle types.

Table 1: Number of selected concept images per verb instance.

No. of concept images	1	2	3	4	5	6	7	8	9	10–16
No. of verbs	3,192	1,556	456	178	72	31	11	7	1	5

Figure 2 shows the average ratings to which degree the target verbs were (un)known to the experiment participants (cf. Section 3.2). Setting a threshold in the middle of the scale 1–4 at 2.5 classifies 153 of the 300 target verbs as neologisms. All 30 base verbs were known to the participants and received an average rating >3.2. Figure 3 shows that the distribution of unknown vs. known PVs varies across the domains of their underlying BVs. PVs with `Force` and `Sound` BVs are more prominent among unknown PVs, while PVs with `MnT` BVs are more prominent among known PVs.

3.4.2 Concept image selection across particles

The heat map in Figure 4 shows the preferences for selected concept images across particle types, calculated as follows. For each annotated verb instance we determined the proportion of selection for each concept image. For example, if two concept images were chosen by a specific participant and for a specific verb instance, each of the two concept images received a proportion of 0.5, and all others received proportions of 0. These proportions were then averaged over all PV instances with the same particle type, across participants. The color red indicates strong preferences of a specific concept image selected for a specific particle type, the color blue indicates weak preferences. Overall, the average preferences range from 0.004 to 0.214.

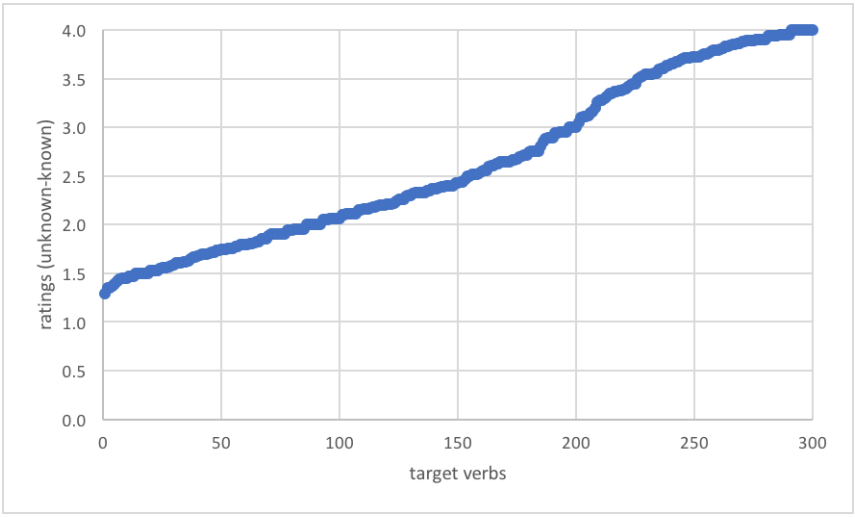


Figure 2: Ratings of unknown/known target verbs.

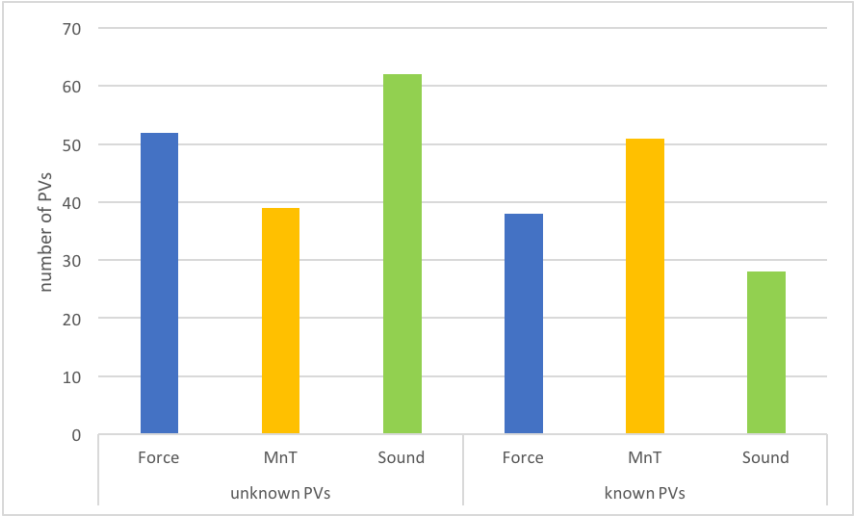


Figure 3: Unknown/known target particle verbs across domains.

	ab	an	auf	aus	ein	mit	nach	vor	zu
vert-up	0.016	0.046	0.136	0.047	0.023	0.029	0.018	0.051	0.017
vert-down	0.150	0.065	0.072	0.086	0.142	0.056	0.066	0.067	0.102
hori-right	0.048	0.138	0.032	0.051	0.057	0.102	0.167	0.158	0.076
hori-left	0.030	0.032	0.012	0.016	0.033	0.023	0.045	0.085	0.041
dia-down-right	0.165	0.088	0.033	0.088	0.106	0.039	0.079	0.074	0.089
dia-down-left	0.077	0.034	0.009	0.029	0.052	0.018	0.036	0.032	0.040
dia-up-right	0.055	0.107	0.214	0.107	0.052	0.071	0.067	0.108	0.052
dia-up-left	0.017	0.012	0.026	0.019	0.008	0.017	0.009	0.027	0.011
vert-out	0.049	0.050	0.069	0.070	0.046	0.069	0.051	0.045	0.051
hori-out	0.092	0.069	0.085	0.119	0.048	0.158	0.079	0.059	0.050
dia-out-up-right	0.027	0.027	0.046	0.049	0.018	0.053	0.032	0.039	0.012
dia-out-up-left	0.037	0.022	0.013	0.019	0.010	0.029	0.027	0.020	0.024
vert-in	0.030	0.047	0.052	0.027	0.064	0.047	0.045	0.031	0.085
hori-in	0.053	0.101	0.049	0.051	0.118	0.124	0.075	0.073	0.171
spiral-out	0.088	0.093	0.106	0.157	0.062	0.104	0.111	0.089	0.058
spiral-in	0.066	0.068	0.044	0.066	0.162	0.061	0.095	0.043	0.120

Figure 4: Concept image selection across particle types.

The heat map demonstrates that the particles exhibit clearly different concept image profiles. The particle *auf*, for example, achieved the overall strongest preference of 0.214 for the concept image ***dia-up-right***, and a preference of 0.136 for ***vert-up***. *ab* shows preferences of ≥0.150 for the concepts ***dia-down-right*** and ***vert-down***. *an*, *nach* and *vor* are associated most strongly with ***hori-right*** (preferences 0.138–0.167), *aus* with ***spiral-out*** (preference 0.157), *ein* with ***spiral-in*** and ***vert-down*** (preferences 0.162 and 0.142, respectively), *mit* with ***hori-out*** (preference 0.158), and *zu* with ***hori-in*** (preference 0.171).

3.4.3 Concept image selection across existing PVs and PV neologisms

The heat maps in Figure 5 specify the particle selections of concept images from Figure 4 regarding the participants' ratings of PV knowledge. That is, the upper plot in Figure 5 shows concept image preferences across particles for well-known PVs with an average rating ≥2.5, and the lower plot shows concept image preferences across particles for rather unknown PVs with an average rating <2.5.

While we expected to see more strongly associated concept images for particles in rather unknown PVs (refering to some predominant meaning contribution(s)), this is the case for the majority of particle types (e.g., ***dia-down-right*** for *ab*; ***hori-right*** and ***hori-in*** for *an*; ***spiral-in*** for *ein*; ***hori-right*** for *nach* and *vor*; ***hori-in*** for *zu*) but not for *auf, aus* and *mit*. The figure however indicates that the concept image selections are largely stable for well-known vs. unknown particle verbs, i.e., the strongest preferences of particle types regarding concept images show up in both heat maps.

3.4.4 Concept image selection across BV source domains

Figures 6 and 7 look into concept image selection across BV source domains. Figure 6 presents the average preferences of selected concept images per domain across all particle types. It shows that already the base verbs exhibit clearly different concept image profiles when taking into account the respective source domain. For `Force` BVs, the inward-pointing concept images ***hori-in*** (0.221) and ***vert-in*** (0.125) received the strongest preferences; for `MnT` BVs, the concept image ***vert-down*** (0.154) received the predominant amount of selections, followed by a set of concept images with preferences of ≈0.100–0.110: ***spiral-out***, ***vert-out***, ***dia-down-right*** and ***hori-out***, thus favouring downward- and outward-pointing arrow types while being rather flexible, i.e., with less strong overall preferences; for `Sound` BVs, the strongly favoured concept image is ***spiral-out*** (0.288), with a set of secondary selections for ***hori-out*** (0.134), ***spiral-in*** (0.109) and ***vert-out*** (0.097), thus favouring spiral-shaped and outward-directed arrows.

known PVs	ab	an	auf	aus	ein	mit	nach	vor	zu
vert-up	0.018	0.068	0.116	0.058	0.045	0.021	0.013	0.076	0.025
vert-down	0.149	0.064	0.067	0.114	0.140	0.044	0.045	0.074	0.115
hori-right	0.057	0.131	0.030	0.022	0.065	0.161	0.160	0.163	0.058
hori-left	0.037	0.027	0.015	0.008	0.023	0.013	0.028	0.108	0.036
dia-down-right	0.150	0.092	0.042	0.080	0.125	0.038	0.078	0.070	0.069
dia-down-left	0.063	0.024	0.009	0.034	0.050	0.007	0.033	0.025	0.046
dia-up-right	0.028	0.135	0.257	0.110	0.045	0.091	0.080	0.108	0.061
dia-up-left	0.018	0.014	0.032	0.012	0.006	0.033	0.006	0.044	0.008
vert-out	0.026	0.041	0.049	0.083	0.032	0.072	0.035	0.029	0.039
hori-out	0.134	0.071	0.065	0.072	0.044	0.169	0.062	0.075	0.051
dia-out-up-right	0.027	0.017	0.037	0.039	0.021	0.056	0.016	0.030	0.017
dia-out-up-left	0.036	0.017	0.014	0.016	0.010	0.031	0.022	0.025	0.027
vert-in	0.028	0.019	0.046	0.034	0.071	0.030	0.039	0.033	0.079
hori-in	0.063	0.069	0.059	0.078	0.127	0.077	0.060	0.042	0.159
spiral-out	0.095	0.121	0.111	0.162	0.071	0.099	0.184	0.064	0.068
spiral-in	0.071	0.090	0.050	0.076	0.124	0.058	0.138	0.033	0.142

unknown PVs	ab	an	auf	aus	ein	mit	nach	vor	zu
vert-up	0.014	0.020	0.159	0.040	0.010	0.033	0.021	0.034	0.008
vert-down	0.150	0.066	0.077	0.066	0.143	0.062	0.080	0.061	0.086
hori-right	0.042	0.148	0.033	0.070	0.052	0.072	0.171	0.154	0.098
hori-left	0.025	0.040	0.010	0.021	0.039	0.029	0.055	0.070	0.047
dia-down-right	0.176	0.082	0.023	0.093	0.093	0.039	0.079	0.077	0.112
dia-down-left	0.087	0.047	0.008	0.025	0.053	0.023	0.037	0.036	0.034
dia-up-right	0.073	0.073	0.164	0.105	0.056	0.061	0.058	0.107	0.041
dia-up-left	0.016	0.010	0.020	0.024	0.009	0.010	0.011	0.016	0.015
vert-out	0.066	0.061	0.092	0.060	0.055	0.067	0.061	0.057	0.065
hori-out	0.063	0.066	0.108	0.152	0.050	0.152	0.091	0.047	0.050
dia-out-up-right	0.027	0.039	0.057	0.055	0.015	0.052	0.042	0.045	0.007
dia-out-up-left	0.037	0.028	0.013	0.021	0.011	0.028	0.030	0.017	0.021
vert-in	0.031	0.081	0.058	0.023	0.059	0.055	0.048	0.030	0.091
hori-in	0.046	0.141	0.038	0.031	0.112	0.147	0.085	0.094	0.184
spiral-out	0.084	0.058	0.101	0.154	0.056	0.106	0.064	0.107	0.046
spiral-in	0.062	0.040	0.039	0.059	0.186	0.063	0.067	0.049	0.096

Figure 5: Concept image selection across particles and (un)known PVs.

Figure 7 demonstrates that the BV patterns across concept images are partly preserved and partly over-written when combining the BVs with specific particles. PVs composed of `Force` BVs and particles *an, ein, mit, zu* inherit the strong preference for ***hori-in***. Similarly, PVs composed of `Sound` BVs and particles *aus, ein, mit, nach, zu* inherit the strong preferences for spirals from the BVs, with *an, nach, vor* at the same time showing strong preferences for ***hori-right***. For PVs composed of `MnT` BVs, where already the concept image preferences for the BVs were less skewed than for the other two domains, it seems that also the respective PVs do not exhibit specific domain-dependent concept image preferences.

Across domains, the PVs with particles *ab, auf, nach, vor* appear to contribute rather constant meaning components: the most strongly selected concept images tend to be consistent across BV source domains and largely correspond

	Force	MnT	Sound
vert-up	0.023	0.052	0.053
vert-down	0.096	0.154	0.035
hori-right	0.061	0.035	0.020
hori-left	0.013	0.016	0.008
dia-down-right	0.059	0.100	0.024
dia-down-left	0.016	0.021	0.007
dia-up-right	0.029	0.064	0.063
dia-up-left	0.005	0.008	0.000
vert-out	0.054	0.111	0.097
hori-out	0.096	0.098	0.134
dia-out-up-right	0.028	0.056	0.050
dia-out-up-left	0.016	0.031	0.030
vert-in	0.125	0.013	0.045
hori-in	0.221	0.033	0.036
spiral-out	0.087	0.117	0.288
spiral-in	0.071	0.090	0.109

Figure 6: Concept image selection across base verbs, with reference to their domains.

to the overall strongest concept images in Figure 4. PVs with particles *an* and *auf* represent constants in a different way: in comparison to the other particle types, they seem to be more flexible in their meaning contribution, i.e., they do not show particularly strong preferences for specific concept images but similarly strong preferences for a range of concept images. Nevertheless, also these more constant particle meanings are influenced by the BV domains; for example, *ab* shows a strong preference for ***spiral-out*** when combined with `Sound` BVs; *an* shows a strong preference for ***hori-in*** when combined with `Force` BVs; *auf* shows a strong preference for ***vert-up*** when combined with `Sound` BVs and only a loose preference for ***dia-up-right*** when combined with `Force` BVs; *nach* and *vor* show strong preferences for spirals when combined with `Sound` BVs and no strong preferences when combined with `MnT` BVs.

FORCE

	ab	an	auf	aus	ein	mit	nach	vor	zu
vert-up	0.003	0.021	0.064	0.045	0.004	0.044	0.015	0.042	0.015
vert-down	0.148	0.068	0.098	0.066	0.129	0.075	0.065	0.096	0.091
hori-right	0.049	0.138	0.050	0.077	0.072	0.107	0.151	0.174	0.047
hori-left	0.056	0.047	0.019	0.019	0.047	0.034	0.058	0.090	0.047
dia-down-right	0.153	0.068	0.027	0.036	0.089	0.024	0.061	0.043	0.064
dia-down-left	0.071	0.035	0.013	0.011	0.041	0.020	0.039	0.032	0.014
dia-up-right	0.069	0.068	0.163	0.085	0.023	0.050	0.055	0.082	0.032
dia-up-left	0.020	0.016	0.029	0.028	0.004	0.004	0.019	0.034	0.017
vert-out	0.044	0.040	0.048	0.067	0.037	0.047	0.043	0.020	0.023
hori-out	0.092	0.048	0.100	0.173	0.041	0.108	0.059	0.054	0.023
dia-out-up-right	0.019	0.026	0.043	0.051	0.013	0.043	0.019	0.030	0.011
dia-out-up-left	0.028	0.030	0.013	0.021	0.003	0.029	0.023	0.015	0.033
vert-in	0.067	0.081	0.114	0.063	0.131	0.086	0.090	0.048	0.157
hori-in	0.104	0.211	0.124	0.098	0.213	0.230	0.158	0.167	0.286
spiral-out	0.035	0.047	0.048	0.103	0.051	0.051	0.066	0.036	0.049
spiral-in	0.042	0.055	0.049	0.057	0.103	0.047	0.080	0.037	0.090

MnT

	ab	an	auf	aus	ein	mit	nach	vor	zu
vert-up	0.028	0.070	0.144	0.071	0.060	0.024	0.016	0.055	0.026
vert-down	0.172	0.077	0.070	0.127	0.161	0.076	0.094	0.066	0.116
hori-right	0.044	0.097	0.023	0.026	0.048	0.117	0.152	0.125	0.058
hori-left	0.015	0.019	0.010	0.006	0.013	0.023	0.026	0.077	0.023
dia-down-right	0.166	0.134	0.043	0.109	0.166	0.061	0.106	0.105	0.099
dia-down-left	0.084	0.050	0.009	0.032	0.073	0.024	0.042	0.052	0.070
dia-up-right	0.038	0.136	0.239	0.126	0.049	0.093	0.077	0.117	0.061
dia-up-left	0.020	0.005	0.028	0.014	0.007	0.027	0.004	0.025	0.001
vert-out	0.066	0.049	0.067	0.095	0.051	0.093	0.063	0.049	0.071
hori-out	0.111	0.099	0.083	0.078	0.053	0.163	0.106	0.073	0.092
dia-out-up-right	0.033	0.017	0.071	0.056	0.029	0.066	0.043	0.051	0.019
dia-out-up-left	0.045	0.022	0.018	0.021	0.015	0.027	0.028	0.028	0.025
vert-in	0.006	0.019	0.006	0.004	0.028	0.015	0.020	0.020	0.049
hori-in	0.023	0.032	0.016	0.030	0.043	0.052	0.054	0.023	0.108
spiral-out	0.094	0.097	0.135	0.145	0.058	0.077	0.066	0.100	0.054
spiral-in	0.055	0.079	0.038	0.059	0.147	0.061	0.104	0.034	0.127

SOUND

	ab	an	auf	aus	ein	mit	nach	vor	zu
vert-up	0.017	0.049	0.201	0.024	0.005	0.017	0.023	0.057	0.011
vert-down	0.128	0.049	0.047	0.063	0.136	0.015	0.038	0.036	0.099
hori-right	0.051	0.181	0.022	0.047	0.050	0.079	0.199	0.174	0.125
hori-left	0.019	0.031	0.008	0.024	0.038	0.012	0.051	0.089	0.053
dia-down-right	0.177	0.061	0.029	0.125	0.060	0.030	0.069	0.074	0.103
dia-down-left	0.076	0.018	0.005	0.046	0.042	0.009	0.025	0.011	0.036
dia-up-right	0.057	0.119	0.238	0.111	0.086	0.071	0.068	0.124	0.063
dia-up-left	0.011	0.016	0.021	0.015	0.014	0.021	0.004	0.023	0.015
vert-out	0.038	0.061	0.094	0.045	0.051	0.066	0.045	0.069	0.059
hori-out	0.073	0.061	0.072	0.102	0.050	0.206	0.072	0.049	0.038
dia-out-up-right	0.029	0.037	0.024	0.038	0.011	0.050	0.033	0.036	0.007
dia-out-up-left	0.036	0.014	0.010	0.014	0.013	0.032	0.030	0.018	0.014
vert-in	0.017	0.040	0.038	0.012	0.029	0.037	0.025	0.025	0.046
hori-in	0.034	0.056	0.010	0.019	0.095	0.085	0.009	0.026	0.114
spiral-out	0.136	0.138	0.134	0.232	0.077	0.191	0.207	0.132	0.071
spiral-in	0.101	0.070	0.046	0.083	0.242	0.076	0.101	0.057	0.145

Figure 7: Concept image selection across particles and BV domains.

4 Discussion

In the remainder of this article, we refer the analyses in the previous section back to our hypotheses about particle meanings and particle concepts (Section 4.1) before we explore the role of the BV source domains (Section 4.2) and go into detailed meaning investigations regarding the particle *ab* (Section 4.3).

4.1 General analysis of particle concept hypotheses

The experiment participants associated *auf* and *ab* with the vertical arrows (***vert-up*** and ***vert-down***) and also with the corresponding diagonal versions pointing to the right: ***dia-up-right*** and ***dia-down-right***, as predicted. The respective diagonal arrows pointing to the left were not chosen, which is an indication for the involvement of the horizontal time dimension. The particle *an* was, as predicted, most strongly associated with the ***hori-right*** concept image; the additionally predicted ***dia-up-right*** concept image achieved a secondary preference.[4] *nach* and *vor* were strongly associated with the ***hori-right*** concept image, which again indicates a reference to the time dimension. Since most *nach* and *vor* readings have a temporal component, a derivation of the basic particle concept from the time domain instead of the space domain should therefore be considered as an explanation.

In the case of *aus*, the ***spiral-out*** concept image was selected most often. This can be explained by the strong association of the particle's prevalent meaning refering to a *container* Image Schema necessary for assigning a direction to *aus*. Since this experimental concept image setting consisted only of two-dimensional arrows, we can however only speculate about the relevance of the container representation. In contrast, *ein* was –in accordance with our assumptions– associated with ***vert-down*** (next to ***spiral-in***), although these directions also require the notion of a container. In order to conceptualise an outside area, as necessary for many *aus* PV readings, it might be sufficient to think of a single wall in order to distinguish between an outside and an inside area. This could explain why *ein* received –in contrast to *aus*, stronger preferences for the predicted concept images based on the constraint of the existence of an imaginary container.

The particle *mit* was most strongly associated with the ***hori-out*** concept image, in accordance with our assumptions. *zu* was not linked to ***dia-up-right***,

[4]Our results regarding *auf* and *an* are also in accordance with the insights of a lexical decision experiment presented by Frassinelli et al. (2017), which indicated that the particles have a predominant vertical/horizontal directionality, respectively.

which we considered as possible concept representation for the intentional readings with an abstract goal. The strongest selection was in favour of the double-arrow concept image ***hori-in***, followed by ***spiral-in***, thus suggesting that a different sense of the particle was more salient in the contexts of the selected BVs. For example, for the PVs *zuzerren* ('to drag until closed') and *zustopfen* ('to plug') the particle introduces a closure relation, which is connected to the also chosen ***vert-down***. However, the *zu*-PVs based on the abstract Sound BVs were also associated with these concept images, which at first sight does not fit to the concrete closure notion. Here, it seems to be more likely that the selection for ***spiral-in*** does not represent the particle meaning, but the meanings of the Sound BVs. Together with the choice of ***hori-in*** as in *zudröhnen* ([*zu*]+'to drone/get stoned'), this points to an interpretation of *zu* as an abstract closure, where the closure is understood as the impairment of auditory perception, as realised through the very dominant and constant sound provided by the BV *dröhnen*. In this interpretation, each arrow head of ***hori-in*** conceptually points to one ear.

4.2 Analysis of BV source domains

Figure 6 suggested that the BV source domains were associated with different preferences for concept images, although none of the BV classes is directional from a lexical semantic perspective. We believe that the associations between source domains and concept images thus indicate conceptual relations to directionality.

The MnT domain with its concrete BVs provides strong preferences for the concepts ***vert-down***, ***dia-down-right***, ***vert-out***, ***hori-out*** and ***spiral-out***. The associations with ***hori-out*** and ***spiral-out*** can be explained with the visually clearly defined and easily imaginable manners of movement of the BVs *schleifen* ('to sand'), *sägen* ('to saw'), *spitzen* ('to sharpen'), etc., whereas the associations to ***vert-down***, ***vert-out*** and ***dia-down-right*** can be traced back to the manners of movements of *hämmern* ('to hammer'), *graben* ('to dig'), *schaufeln* ('to shovel'), etc. However, the question arises why only the downward-pointing concept images were chosen and not the upward-oriented ones. We approach the question on a theoretical semantic basis. The BVs are denominal action verbs, either derived from an instrument (such as a shovel, a hammer, a fork) or from an intended result (such as a grave), and describe a repetitive motion. The involved motion has at least two changes of directions, marking the extreme points of the movement. The direct objects of MnT verbs typically refer to one of those extreme points, as in example (39), where *schaufeln* refers to the area beneath the ground which lies below our usual perceptual horizon. This idea corresponds to Lachmair

et al. (2016)'s research which shows that words trigger specific spatial locations. Other frequent arguments of *schaufeln*, such as *hole* and *soil*, also refer to such a "down" area, as in examples (40) and (41). Here, the motion is spanned between the initial position of the instrument and the position of the affected area. In the examples (39–41), the direction of the shovel motion is defined between the initial "up" location of the shovel and the "down" location of the ground, thus justifying the downward concept images over upward concept images.

(39) Karin schaufelt einen Graben.
Karin shovel a graben
'Karin shovels a graben.'

(40) Karin schaufelt ein Loch.
Karin shovel a hole
'Karin shovels a hole.'

(41) Karin schaufelt Erde.
Karin shovel soil
'Karin shovels soil.'

On the contrary, the Sound BVs, which are the most abstract verbs in this data set, were not linked to many of our simple directional concept images. They were mainly associated with the spirals, thus suggesting a mental mapping to the prototypical picture of a sound wave. That is, the underlying idea of the spiral as concept representation was a uniform expansion, which matches to the motion behaviour of sound waves. In addition, there was some preference for the double-headed arrows ***hori-out*** and ***vert-out*** as concept images for the BVs with a repetitive sound character. This can be attributed to the strongly prototypical manner of sound production actions, which are usually caused by an up-and-down motion as in drumming, or a left-to-right motion as in clapping. This means that the Sound BVs, which are not directional from a lexical semantic perspective, were analysed as conceptually directional. This clear-cut mapping between spiral and sound wave as well as between double-headed arrow and manner-of-production of repetitive sounds, allows distinguishing between the concept images triggered by the BVs and the concept images triggered by the particle, which provides insight into the composition process and explains the low compatibility between particle types and Sound BVs, as reflected in the high number of neoPVs in Figure 3.

The Force BVs describe events which are mainly defined through the interplay of two concrete arguments. In comparison to MnT verbs, the Force verbs are less concrete, but at the same time they are also less abstract than the Sound verbs. The importance of the arguments shows up in the preference for the concepts ***hori-in*** and ***vert-in***, which both have two arrow heads. The concept images are thereby similar to the vectors used in the schematic representations of forceful verbs by Zwarts (2010).

4.3 Analysis of particle *ab*

In the last part of our analyses we focus on concept image preferences regarding one specific particle type. We choose *ab*, the particle which is strongly associated with a downward direction.

Figure 8 shows the distribution over concept images for PVs with particle *ab* across BV source domains. In all three domains, the participants agreed on the two Down concepts (i.e., ***vert-down*** and ***dia-down-right***), although the PVs in the experiment were assigned to different lexical semantic classes by Kliche (2011).

	Force	MnT	Sound
vert-up	0.003	0.028	0.017
vert-down	0.148	0.172	0.128
hori-right	0.049	0.044	0.051
hori-left	0.056	0.015	0.019
dia-down-right	0.153	0.166	0.177
dia-down-left	0.071	0.084	0.076
dia-up-right	0.069	0.038	0.057
dia-up-left	0.020	0.020	0.011
vert-out	0.044	0.066	0.038
hori-out	0.092	0.111	0.073
dia-out-up-right	0.019	0.033	0.029
dia-out-up-left	0.028	0.045	0.036
vert-in	0.067	0.006	0.017
hori-in	0.104	0.023	0.034
spiral-out	0.035	0.094	0.136
spiral-in	0.042	0.055	0.101

Figure 8: Concept image selection for *ab* across BV domains.

Looking into specific PVs with strong preferences for the two Down concept images, an example instance of an unknown PV is represented by *abhämmern* ([ab]+'to hammer'), cf. example 42. We assume that this PV was understood as a separation performed by a hammering force. *abquetschen* ('to squeeze off') in

example (43) is an instance of a well-known PV where the particle is combined with a `Force` BV, describing a force that causes a SEPARATION. The well-known PV *abklingen* combines the particle with a `Sound` BV; literally, it describes that a sound fades away, but it is more common in its metaphorical reading of approaching the end of an event together with a value decrease, as in example (44). The approaching of the end of the storm can be conceptualised as decreasing intensity within both the value and the time dimensions, or can alternatively be interpreted only temporal, as a slowly ending process. However, in comparison to the previous examples no causer is involved, suggesting that the downward meaning is conceptually connected to *ab*, even if from a lexical semantic perspective only the result is expressed.

(42) Karin hämmert den rostigen Nagel ab.
Karin hammer the rusty nail [ab]
'Karin detaches the rusty nail by hammering.'

(43) Karin hat sich ihren Finger abgequetscht.
Karin has herself her finger [ab]+squeeze
'Karin has crushed her finger.'

(44) Der Sturm klingt ab.
The storm sound [ab]
'The storm is about to stop.'

The examples illustrate that even though the contexts are rather different, the meanings of the particle can in all cases be traced back to a downward direction, either causing or be caused by a separation, varying according to the constraints. We argue that the downward concept is not only the basic meaning component, but the prototypical reading for the particle *ab*.

5 Conclusion

In this article, we have shown that directional concepts, visually represented as arrow pictographs, can be applied to a systematically composed set of German particle verbs and their underlying base verbs. Furthermore, the selected concept images were mostly in accordance with the particle directions predicted on the basis of example sentences, lexical-semantic classifications and spatial experience, and largely stable for well-known vs. unknown particle verbs. Thus,

direction is a concept that should be taken into account as a part of the PV composition process and the contribution of the particle to the particle verb meaning.

Understanding potential particle fundamentals as concepts, instead of meanings, has the advantage that senses are not considered as discrete, static classifications requiring plenty of compromises or borderline cases. Concepts as basic components are flexible and can easily be adjusted to various contexts. Thereby, classes of similar contextual requirements trigger similar concept adjustments, and hence are assumed to enforce a specific particle sense.

Acknowledgements

The research was supported by the DFG Collaborative Research Centre SFB 732 (Sylvia Springorum, Sabine Schulte im Walde) and the DFG Heisenberg Fellowship SCHU-2580/1 (Sabine Schulte im Walde).

References

Abreu, Antonio & Sarah Barbieri Vieira. 2008. Learning phrasal verbs through image schemas: a new approach. In *From Language as System to Language as Use: Image Schemas in the Description and Explanation of Phrasal Verbs in English*.

Boroditsky, Lera. 2001. Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology* 43(1). 1–22.

Dirven, René. 1993. Dividing up physical and mental space into conceptual categories by means of English prepositions. In Zelinksy C. Wibbelt (ed.), *The Semantics of Prepositions – From Mental Processing to Natural Language Processing*, vol. 3, 73–98. Mouton de Gruyter.

Dreyfuss, Henry. 1972. *Symbol sourcebook: An authoritative guide to international graphic symbols*. John Wiley & Sons.

Faaß, Gertrud & Kerstin Eckart. 2013. SdeWaC – A corpus of parsable sentences from the web. In *Proceedings of the international conference of the German society for computational linguistics and language technology*, 61–68. Darmstadt, Germany.

Frassinelli, Diego, Alla Abrosimova, Sylvia Springorum & Sabine Schulte im Walde. 2017. *Meaning (mis-)match in the directionality of German particle verbs*. Poster at the 30th Annual CUNY Conference on Human Sentence Processing. MIT, Cambridge.

Frutiger, Adrian. 1987. *Der Mensch und seine Zeichen*. Marix Verlag.

Gärdenfors, Peter. 2004. *Conceptual spaces: The geometry of thought*. MIT Press.

Gibbs, Raymond W. & Herberg L. Colston. 1995. The cognitive psychological reality of image schemas and their transformations. *Cognitive Linguistics* 6(4). 347–378.

Herskovits, Anette. 1986. *Language and spatial cognition: An interdisciplinary study of the prepositions in English* (Studies in Natural Language Processing). London: Cambridge University Press.

Jackendoff, Ray. 1983. *Semantics and cognition*. Cambridge, MA, USA: MIT Press.

Kliche, Fritz. 2011. Semantic variants of German particle verbs with *"ab"*. *Leuvense Bijdragen* 97. 3–27.

Köper, Maximilian & Sabine Schulte im Walde. 2018. Analogies in complex verb meaning shifts: The effect of affect in semantic similarity models. In *Proceedings of the 16th annual conference of the north american chapter of the association for computational linguistics: human language technologies*, 150–156. New Orleans, LA, USA.

Kövecses, Zolzan. 2002. *Metaphor: A practical introduction*. New York: Oxford University Press.

Kratzer, Angelika. 2003. The event argument and the semantics of verbs. Manuscript, University of Massachusetts at Amherst.

Lachmair, Martin, Carolin Dudschig, Irmgard de la Vega & Barbara Kaup. 2016. Constructing meaning for up and down situated sentences: Is a sentence more than the sum of its words? *Language and Cognition*. 1–25.

Lakoff, George. 1987. *Women, fire and dangerous things: What categories reveal about the mind*. University of Chicago Press.

Lechler, Andrea & Antje Roßdeutscher. 2009. German particle verbs with *"auf"*. Reconstructing their composition in a DRT-based framework. *Linguistische Berichte* 220. 439–478.

Lindner, Susan. 1983. *A lexico-semantic analysis of English verb particle constructions with "out" and "up"*. Indiana University Linguistics Club.

Morgan, Pamela S. 1997. Figuring out *figure out*: Metaphor and the semantics of English verb-particle constructions. *Cognitive Linguistics* 8(4). 327–357.

Neurath, Otto. 1983. *An international encyclopedia of unified science*. Springer.

Richardson, Daniel C., Michael J. Spivey, Shimon Edelman & Adam J. Naples. 2001. "Language is Spatial": Experimental Evidence for Image Schemas of Concrete and Abstract Verbs. In *Proceedings of the annual meeting of the cognitive science society*.

Sag, Ivan A., Timothy Baldwin, Francis Bond, Ann Copestake & Dan Flickinger. 2002. Multiword expressions: A pain in the neck for NLP. In *Proceedings of the*

conference on intelligent text processing and computational linguistics. Mexico City, Mexico.

Schulte im Walde, Sabine, Maximilian Köper & Sylvia Springorum. 2018. Assessing meaning components in German complex verbs: A collection of source–target domains and directionality. In *Proceedings of the 7th joint conference on lexical and computational semantics*, 22–32. New Orleans, LA, USA.

Side, Richard. 1990. Phrasal verbs: Sorting them out. *ELT Journal* 44(2). 144–152.

Springorum, Sylvia. 2011. DRT-based analysis of the German verb particle *"an"*. *Leuvense Bijdragen* 97. 80–105.

Springorum, Sylvia, Sabine Schulte im Walde & Antje Roßdeutscher. 2013. Sentence generation and compositionality of systematic neologisms of German particle verbs. In *Proceedings of the 5th conference on quantitative investigations in theoretical linguistics*, 81–84. Leuven, Belgium.

Springorum, Sylvia, Jason Utt & Sabine Schulte im Walde. 2013. Regular meaning shifts in German particle verbs: A case study. In *Proceedings of the 10th international conference on computational semantics*, 228–239. Potsdam, Germany.

Stiebels, Barbara. 1996. *Lexikalische Argumente und Adjunkte. Zum semantischen Beitrag von verbalen Präfixen und Partikeln*. Berlin: Akademie Verlag.

Talmy, Leonard. 2000. *Toward a cognitive semantics*. Vol. 2. MIT Press.

Tversky, Barbara. 2011. Visualizing thought. *Topics in Cognitive Science* 3. 499–535.

Viberg, Ake. 1983. The verbs of perception: A typological study. *Linguistics* 21(1). 123–162.

Waldron, Ronald A. 1979. *Sense and sense development*. Andre Deutsch Ltd.

Zwarts, Joost. 2010. Forceful prepositions. In Vyvyan Evans & Paul Chilton (eds.), *Language, Cognition and Space: The State of the Art and New Directions*, 193–214. Equinox Publishing.

Appendix

Table 2: Selected 30 base verbs and their source domains. All these base verbs were systematically composed to a total of 270 particle verbs by prefixing them with the nine constituent particle types *ab, an, auf, aus, ein, mit, nach, vor, zu.*

Base verb	Source domain
biegen	Force
brechen	Force
brummen	Sound
donnern	Sound
drängen	Force
dröhnen	Sound
gabeln	Machines and Tools
graben	Machines and Tools
heulen	Sound
hämmern	Machines and Tools
jaulen	Sound
klappern	Sound
klingen	Sound
kämmen	Machines and Tools
pressen	Force
quetschen	Force
rattern	Sound
schalten	Machines and Tools
schaufeln	Machines and Tools
schleifen	Machines and Tools
schrauben	Machines and Tools
spitzen	Machines and Tools
stauen	Force
stopfen	Force
summen	Sound
sägen	Machines and Tools
wummern	Sound
zerren	Force
zwingen	Force
zwängen	Force

Chapter 2

Do semantic features capture a syntactic classification of compounds? Insights from compositional distributional semantics

Sandro Pezzelle
CIMeC – Center for Mind/Brain Sciences, University of Trento

Marco Marelli
Department of Psychology, University of Milano-Bicocca

Classifying compound words has been the ultimate goal of much research in formal linguistics. A popular, cross-linguistically applicable classification (Bisetto & Scalise 2005) distinguished three main types of compounds, namely Subordinate, Attributive, and Coordinate on the basis of the underlying *syntactic* relation between the compound elements. Similar tripartitions have been also proposed in cognitive psychology by work exploring conceptual combination. Focusing on the type of *semantic* interpretation assigned to novel combinations, three main classes have been traditionally described, namely Relation-linking, Property-mapping, and Hybrid or Conjunctive (see Wisniewski (1996)). Based on these commonalities, we conjecture that syntax-based compound types might be also explained by means of the semantic properties of the compound and its constituents. Using a compositional model of distributional semantics (cDSM), we show that both (a) the contribution of each constituent in determining the meaning of the compound and (b) the semantic similarity between the two constituent words are significant predictors of these classes. These findings suggest that the various compound types identified by syntactic criteria can be also predicted by means of semantic features. On the one hand, this confirms the validity of the proposed linguistic categorization. On the other hand, we bring further evidence proving the effectiveness of cDSMs in describing linguistic phenomena.

Sandro Pezzelle & Marco Marelli. 2019. Do semantic features capture a syntactic classification of compounds? Insights from compositional distributional semantics. In Sabine Schulte im Walde & Eva Smolka (eds.), *The role of constituents in multiword expressions: An interdisciplinary, cross-lingual perspective*, 35–56. Berlin: Language Science Press. DOI:??

1 Introduction

1.1 Classifying compounds

Compounding, namely the mechanism by which two independent words (e.g. *pet*, *food*) combine together to form a novel morphologically-complex word (e.g. *petfood*), is one of the most extensively covered topics in the literature of word formation.[1] On the theoretical level, many linguists have been particularly interested in classifying *compounds* according to various criteria, such as 'headedness' (roughly speaking, the position and the characteristics of the compound *head*, the dominant word in the compound, e.g. *food* in *petfood*) (Bloomfield 1933; Fabb 1998); the presence of a verb or a deverbal noun (Marchand 1969); the kind of underlying relation between the *constituent* words, either at a syntactic (Bloomfield 1933; Bally 1950; Lees 1960; Bisetto & Scalise 2005; Baroni et al. 2006; Dressler 2006; Scalise & Bisetto 2009) or at a semantic level (Levi 1978; Warren 1978; Fanselow 1981). Though different and pertaining to somehow diverse levels of analysis, these criteria have been traditionally explored and mixed together within the same classification frameworks (see among others Bauer 2001; Haspelmath 2002; Booij 2005). As a consequence, many influential proposals distinguish various classes of compounds on the basis of several overlapping properties, that often generate an inconvenient number of subclasses and special cases.

To overcome this issue, Bisetto & Scalise (2005) proposed a cross-linguistically (and nowadays widely accepted) classification framework based on a single, homogenous criterion, that is, the underlying syntactic relation between the compound constituents. Three main classes of compounds are isolated, namely Subordinate, Attributive, and Coordinate. To illustrate, the compound *doghouse* belongs to the Subordinate class, since the syntactic relation subtending *dog* and *house* is that of subordination. Indeed, the compound can be paraphrased as 'the house of the dog'. In contrast, *swordfish* is labeled as Attributive, given that the first constituent, *sword*, acts as an attribute of *fish* (a *swordfish* is 'a fish whose nose is shaped like a sword'). Finally, Coordinate compounds are formations like *comedy-drama*, where the first and second constituent are linked by the underlying conjunction 'and'.

1.2 From *word* combination to *conceptual* combination

Interestingly, a similar tripartition has been proposed in the cognitive psychology literature by works on *conceptual combination* (Wisniewski 1996; Costello &

[1]For a complete and exhaustive overview of compounding, see Lieber & Štekauer (2009).

Keane 2000), where the focus is on the type of interpretations provided by people to novel combinations. By analyzing the circumlocutions produced by speakers to interpret novel compounds like *zebra-horse*, in fact, three main classes have been traditionally isolated, namely Relation-linking, Property-mapping, and Hybrid or Conjunctive. The first class includes interpretations involving a relation between the two concepts, i.e. a *zebra-horse* is 'a horse that preys zebras'. In the second, a property of one concept is mapped to the other, i.e. a *zebra-horse* is 'a striped horse'. In the third, the novel concept is interpreted as a hybrid or conjunction of the constituent concepts, i.e. a *zebra-horse* is 'a creature having many properties of both horses and zebras'. Though the aim of these works is to study the various interpretations to novel conceptual combinations, without any interest in recognizing classes of *lexicalized* compound words, nevertheless the types they identify are reasonably comparable to the linguistic ones proposed by Bisetto & Scalise (2005). In particular, Relation-linking interpretations correspond to compounds included in the Subordinate class, Property-mapping to Attribute, and Hybrid/Conjunctive to Coordinate.

One notable difference is that the *linguistic* classification accounts for lexicalized (or familiar) compounds, whereas the *cognitive* one describes novel combinations which still lack a single, well-defined interpretation. However, we can easily assume that lexicalized compounds are the linguistic realization of a conceptual combination process, in a way that all compounds start out as novel formations and become lexicalized with usage in time (Gagné & Spalding 2006). Consistent with this claim is recent evidence showing that, in the processing of both novel and familiar compounds, an active combination of constituent meanings is routinely in place (Gagné & Spalding 2009; Ji et al. 2011; Marelli & Luzzatti 2012; Marelli et al. 2014). This would suggest that the difference between novel and familiar compounds is merely in their degree of lexicalization. While the former can still be interpreted by speakers in various ways, the latter have only one possible interpretation, that corresponds to a fixed syntactic relation between the constituent words in Bisetto & Scalise (2005).

The second important difference is that interpretations of novel combinations pertain to the conceptual level, namely they describe relations between the concepts being combined together. As such, the tripartition described above is essentially *semantic*. In contrast, the linguistic classification considered here is based on a purely *syntactic* criterion. Based on the commonalities highlighted above, however, it might be that the two levels of analysis are not mutually exclusive, but possibly related and somehow overlapping. Lexical semantic approaches corroborate this conjecture. Lieber (2009), for example, proposed that the different compound types identified by Bisetto & Scalise (2005) would depend, at least in

part, on the intrinsic semantic features of the compound constituents. Moreover, classifications of compounds based on taxonomies of semantic relations reveal a certain degree of overlap between the syntactic and semantic analysis (Levi 1978). For example, the semantic relation AND seems hardly distinguishable from the purely syntactic relation of coordination, which is subtended by the underlying conjunction 'and'.

1.3 Aim of the work

Based on this concurring evidence, we conjecture that various classes of compounds defined at the syntactic level may be also explained in terms of the semantic properties of the compounds and their constituents. In particular, our hypothesis is that measures quantifying the semantic role played by each constituent in contributing to the overall compound meaning, as well as the degree of semantic similarity between the constituents should be effective in predicting different classes. Moreover, we expect these semantic measures to be able to capture different, syntax-based classes without relying on other non-semantic compound properties. Crucially, we do not claim that the distinction is thus purely semantic, making superfluous any categorization focusing on the syntactic relation between the compound constituents. Rather, we believe that the theoretically motivated and widely accepted *discrete* classifications proposed by linguists can be also described in terms of the *continuous*, quantitative aspects of the meaning of compounds and their constituents. In other words, we expect the *quantitative* semantic properties to parallel the *qualitative* grammatical distinctions, thus demonstrating, at the same time, the effectiveness of our proposal and the validity of the linguistic theory.

We experiment with a dataset of English compounds for which annotation based on the classification by Bisetto & Scalise (2005) (Subordinate, Attributive, Coordinate) is available. To predict each class, we use several semantic variables such as the degree of similarity between the constituents and the individual contribution of each constituent word in determining the meaning of the whole compound. We quantify these measures by using a compositional model of distributional semantics (Baroni & Zamparelli 2010; Guevara 2010; Mitchell & Lapata 2010; Zanzotto et al. 2010), following recent evidence proving the effectiveness of this approach in modeling morphological processes such as composition and derivation (Marelli & Baroni 2015; Günther & Marelli 2016; Marelli et al. 2017).

1.4 Computational models of meaning

Based on the core notion that similar words occur in similar contexts (Harris 1954; Firth 1957), distributional semantic models (hence, DSMs) represent lexical meanings by means of vectors encoding the contexts in which words appear in a large corpus. The intuition is that words that occur in similar linguistic contexts (e.g., *cat* and *dog*) should be semantically more similar than words that do not. Typically, this geometric representation is used to quantify the degree of distributional similarity between two words. Given the corresponding vectors, the similarity is computed in terms of their geometric distance, typically the cosine of the angle (Turney & Pantel 2010). In particular, the closer two vectors in the semantic space (i.e., the space populated by all the linguistic vectors), the higher their similarity. Traditional DSMs, such as the pioneering Latent Semantic Analysis (LSA; Landauer & Dumais 1997), have been largely used to obtain quantitative estimates of important semantic variables such as the degree of conceptual or topical similarity between two words (Padó & Lapata 2007; Gagné & Spalding 2009; Kuperman 2009; Wang et al. 2014).

1.5 Distributional semantics and compounds

In the domain of compounds, distributional semantic approaches have been extensively applied to two main tasks: Noun-noun compound interpretation (Van de Cruys et al. 2013; Dima & Hinrichs 2015; Dima 2016; Shwartz & Dagan 2018; Fares et al. 2018) and compositionality prediction (Reddy et al. 2011; Schulte im Walde et al. 2013; Salehi et al. 2014; 2015; Cordeiro et al. 2016). The former task, usually tackled as a classification problem, aims at automatically predicting the *semantic* interpretation of the compound (i.e., the semantic relation tying the constituents). Given the compound *street protest*, for example, a system is trained to predict that the relation holding between the nouns is 'locative'. Several datasets of compounds annotated with different numbers of semantic relations have been released for the tasks (Ó Séaghdha 2007; Tratz & Hovy 2010), and various systems capitalizing on distributional representations (usually obtained with neural network architectures; see section 2.1) have been recently proposed. Overall, this approach has been proved to be successful in the task, though the performance is shown to be dependent on the number and granularity of semantic relations. As for the latter task, it is focused on predicting the degree of compositionality of a noun-noun compound, namely the extent to which the meaning of the whole depends on the meaning of the constituent words. Various datasets annotated with human judgments have been proposed through time (Reddy et al. 2011; Roller

et al. 2013; Farahmand et al. 2015), and extensive explorations of DSMs in the task have been carried out. Crucially for the purpose of this study, distributional measures of similarity obtained with compositional approaches were found to be highly predictive of human judgments in this task (Reddy et al. 2011; Schulte im Walde et al. 2013; Salehi et al. 2015; Cordeiro et al. 2016).

1.6 A compositional approach to compounds

Of great interest for the present work, Lynott & Ramscar (2001) were the first to employ distributional semantic models to study novel compounds (e.g. *zebra-horse*). In particular, the aim of that work was to test whether a measure of semantic similarity between compound constituents (quantified with LSA) was predictive of both (a) the ease of novel compound comprehension and (b) the distinction between Relation-linking and Property-mapping combinations. To do so, they experimented with novel compounds and their corresponding interpretations as provided by previous works on conceptual combination (Wisniewski & Love 1998; Gagné 2000). Overall, the model was shown to perform remarkably well in all the tasks. Lynott & Ramscar (2001), however, claimed that current distributional models like LSA were not capable of modeling the whole process of conceptual combination. Since they can only quantify the similarity between independent, free-standing words (e.g. *zebra* and *horse*), they are not informative at all about the relation between these words and the resulting compound. As such, they represent static, word-based models of lexical semantics which do not account for the potentially infinite linguistic productivity.

Compositional DSMs (hence, cDSMs) precisely tackle these issues. Aimed at accounting for the compositional nature of language (Baroni, Bernardi, et al. 2014), these models capitalize on DSM vectors and perform either simple (Mitchell & Lapata 2010) or more complex, theoretically inspired operations (Baroni & Zamparelli 2010; Guevara 2010; Zanzotto et al. 2010) to *compose* existing lexical entries. By exploiting simple operations (sum, multiplication) or being trained with distributional information about combinations that are already observed in the source corpus, these models can indeed be used to generate meaning representations for both novel and lexicalized formations. Recently, this approach was shown to be effective in modeling morphological processes such as derivation and compounding (Marelli & Baroni 2015; Günther & Marelli 2016; Marelli et al. 2017). Closely related to the present study, recent work (Günther & Marelli 2016; Marelli et al. 2017) exploited cDSMs to generate compositional representations of compounds. Marelli et al. (2017), in particular, explored whether a simple but effective regression-based compositional method (Guevara 2010) can capture the

variability in semantic relations between the constituents of novel compounds. This system was shown to be remarkably effective and flexible in capturing relational information. Based on this evidence, in the present work we employ the same model and test it in the task of predicting theoretically motivated, syntax-based classes of compounds.

2 Experiment

The present experiment investigates whether different, syntax-based classes of compound words (Subordinate, Attributive, and Coordinate) can be captured by means of semantic properties of the compound and its constituents. To quantify these properties, (a) we generate compositional representations of compounds and obtain similarity scores assessing the role of each constituent in contributing to the overall meaning; (b) we measure the degree of similarity between the first and second constituent.

A note on the terminology used in the paper. Until this point, we used the neutral terms 'first constituent' and 'second constituent' to refer to, respectively, *dog* and *house* in *doghouse*. As briefly mentioned in section 1.1, one constituent usually plays a dominant role compared to the other since it acts as the 'head' of underlying phrase. In this example, the head is clearly *house* (indeed, *doghouse* is 'the house of the dog'). Consistently, this element determines the syntactic category of the phrase and, semantically, it represents a hyperonym of the compound. By default, in English compounds the second constituent acts as the compound 'head', whereas the first acts as the compound 'modifier' (Bauer 2009). We stick with this arguably simplified terminology[2] and, from now on, we interchangeably use the terms 'first constituent' or 'modifier' to refer to the leftmost element, 'second constituent' or 'head' to refer to the rightmost one.

2.1 Semantic space

Following Baroni, Dinu, et al. (2014), who demonstrated that DSMs generated using feedforward neural networks models largely outperform traditional count-based architectures in many tasks, we built a state-of-the-art `CBOW` semantic space using the `word2vec` toolkit by Mikolov et al. (2013), with all the parameters that turned out to be best-predictive in Baroni, Dinu, et al. (2014). In particular, the

[2]Without going into much detail, it should be mentioned that this picture is indeed less straightforward than it may appear. For instance, in the English compound *singer-songwriter* the two constituents play a similar role, in a way that they could be both considered as the compound 'head' (and the compound as 'double-headed') (Bauer 2009)

vectors have 400 dimensions and were built using (a) a context window of 5 words to either side of the target word, (b) a subsampling procedure which penalizes high-frequency words in the training phase ($t = 1e^{-5}$), (c) 10 negative samples. The vectors were trained using a corpus of written English containing around 2.8-billion tokens (a concatenation of BNC, ukWaC, and a 2009-dump of Wikipedia), the same used in Baroni, Dinu, et al. (2014). To avoid sparsity effects, we experimented with the vectors corresponding to the 300K most frequent words in the corpus.

2.2 Materials

We experimented with a sample of the MorBoComp database including 163 English compounds. MorboComp is a large, multilingual database of compounds that has been developed to study compounding from a typological perspective.[3] Each compound in the database is richly annotated (i.e., it is provided with information about headedness, compound and constituents' grammatical category, compound structure, etc.) and, crucially for our purposes, it is classified as Subordinate (hence, SUB), Attributive (hence, ATT) or Coordinate (hence, CRD) on the basis of the classification and terminology proposed by Bisetto & Scalise (2005). To illustrate, *schoolteacher* is tagged as SUB, *keyword* as ATT, and *king-emperor* as CRD.

Consistently with the criteria outlined in Bisetto & Scalise (2005), the 163-item sample contained cases of both 'phrasal' compounds (*do-it-yourself illustration*, *around-the-world flight*) and 'neoclassical' formations (*bibliography*, *theology*). In addition, a handful of items labeled with OTH (i.e., Other) were found. Since this label was used by the annotators for either unresolved or idiosyncratic cases, however, we decided not to consider them in our investigation. Similarly, we removed neoclassical formations since their constituents can be affixes and suffixes rather than free-standing, independent words (e.g. *biblio-*). As a consequence, in our distributional semantics approach we could not have a vector representation for these items. Finally, additional 9 compounds were discarded since one of their constituents turned out not to be included in the 300K-vector semantic space. Specifically, 8 out of 9 of the missing items were first constituents of phrasal compounds, e.g. *all-goes-well* (*in all-goes-well atmosphere*) or *floor-of-a-birdcage* (*in floor-of-a-birdcage taste*), whereas in one case (*well-deserver*) the missing items was the second constituent (*deserver*). After this filtering process, our resulting dataset included 132 compounds (67 SUB, 49 ATT, 16 CRD), that we used for our experiment.

[3]For further details, see: http://morbocomp.sslmit.unibo.it/index.php?section=home

2.3 Generating composed representations

For each of the 132 compounds in the list, we generated a composed representation using the vectors described in section 2.1 and the compositional model by Guevara (2010). As previously mentioned, one of the main strengths of compositional DSMs is their ability to produce meaning representations also for combinations that are not attested in the source corpus. That is, given a novel or unattested compound, we are able to represent it as an independent vector on the basis of the meanings of its constituents (*zebra* and *horse*). This aspect was of crucial importance in our experiment, where 60 out of the 132 compounds extracted from MorBoComp turned out not to be present in the source semantic space. That is, almost half of the compounds were not among the 300K most frequent words in the corpus and, consequently, did not have a distributional representation. By using a compositional model capitalizing on the representations of the two constituents, however, we were able to overcome this limitation of traditional DSMs and generate a meaning representation for all the items, regardless of whether they had a 'static' semantic representation or not.

The method used in the present study, in particular, was implemented by Guevara (2010) to model compositionality as depending on the semantic relation instantiated in the syntactic structure. As such, it looks particularly suitable for the case of compounds, which embed a modifier-head structure. Indeed, previous work proved this model to be very effective in generating composed representations for compounds (Marelli et al. 2017). Technically, the composed representations are obtained with the combinatorial procedure depicted in Figure 1: given two vectors $\vec{u}$ and $\vec{v}$ each representing one of the constituent words, their composed representation can be computed as $\vec{c} = \mathbf{M}\vec{u} + \mathbf{H}\vec{v}$, where $\mathbf{M}$ and $\mathbf{H}$ are weight matrices estimated from training examples. These matrices are trained using least squares regression,[4] having the vectors of the constituents as independent words (*dog*, *house*) as inputs and the vectors of example compounds (*doghouse*) as outputs. The two matrices are thus optimized so that the similarity between the weighted sum of the two constituent vectors (the composed vector) and the compound vector extracted from the semantic space (the observed vector) is maximized. Or, in other words, the composed vector obtained by means of the compositional model is built in a way that closely approximates the original one.

[4] As reported by Guevara (2010), this method is commonly employed to approximate functions in problems of multivariate multiple regression with a small number of observations and a greater number of variables, that is a similar condition to the one involving high-dimensionality vectors representing word meanings and (relatively) limited data.

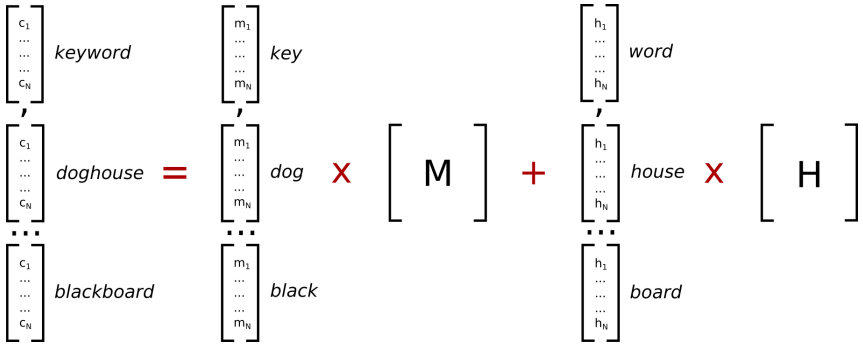


Figure 1: Representation of the training phase of the compositional method used in the study (adapted from Marelli et al. 2017).

In the present study, we trained the compositional model with a list of English noun-noun compounds extracted from the CELEX English Lexical Database (Baayen et al. 1995). By default, we treated all compounds as written in *solid* form, that is, without whitespaces or hyphens between the two constituents. When the solid compound was not found in our semantic space, we looked for it in its hyphenated form. The training set included 2174 triplets <modifier, head, compound>, none of which was also present in the dataset we obtained from MorBoComp. We then used the estimated weight matrices for generating composed representations for each of the 132 compounds in our sample.

2.4 Semantic variables

For each vector obtained compositionally, we computed four composition-based semantic measures, namely (1) similarity between the composed representation of the compound and its modifier (e.g. between *keyword* and *key*), (2) similarity between the composed representation of the compound and its head (e.g. between *keyword* and *word*), (3) neighborhood density, that is, the average cosine similarity between the composed vector and its top-10 nearest neighbor vectors in the semantic space (all these 3 measures have been introduced by Vecchi et al. 2011), and (4) entropy, that is a measure of vector quality firstly introduced by Lazaridou et al. (2013).

By operationalizing the similarity between the composed compound vector and either constituents, in particular, we aimed at quantifying the extent to which each single word contributes to the overall, compositionally-obtained meaning. Although operationalized in terms of the cosine of the angle between the

compound vector and either constituents (in the same way as standard DSMs do), indeed, these measures genuinely describe the morphological process itself rather than merely taking into account its starting and ending points. Based on these properties, such measures have been recently used in studies with compound words. For example, they have been shown to be effective in predicting meaningfulness ratings on novel combinations (Günther & Marelli 2016) and in capturing relational information in compounds (Marelli et al. 2017).

As far as neighborhood density and entropy are concerned, both of them have been proposed to provide information about the meaningfulness of vectors encoding new concepts. The rationale of the former is that meaningful vectors should live in a region of the semantic space that is densely populated by vectors representing many related concepts, while meaningless vectors should be way more isolated. For the latter, the intuition is that meaningful vectors should have a skewed distribution, with few dimensions (corresponding to the salient semantic features of the word) being highly activated, namely having large values. In contrast, meaningless vectors should have a more uniform distribution, which would be a proxy for a less defined, fuzzier meaning. As a consequence, entropy would be inversely correlated with meaningfulness.

A (5) fifth semantic but non-compositional measure was introduced following Lynott & Ramscar (2001), who employed Latent Semantic Analysis (LSA) to quantify the degree of similarity between the first and the second constituent of a compound. Here, we took the compound constituent vectors (e.g. the vectors of *key* and *word*) from the source semantic space (see section 2.1) and simply computed their cosine similarity. This measure might be helpful in distinguishing between different compound classes, based on the evidence that in both theoretical linguistics (see Lieber 2009) and conceptual combination literature (see Wisniewski 1996) this factor has been considered as explanatory of different classes/interpretations.

2.5 Non-semantic variables

In addition to the 5 semantic variables described above, we also included in our experiment a number of non-semantic control variables. For each compound and its constituent words we extracted word-form frequency from the source corpus (i.e., the number of times a word is encountered in the corpus in that exact form, regardless of its grammatical category). Compound frequency was calculated by summing the occurrences of the given compound in both solid and hyphenated orthographic form (*blackboard* and *black-board*, respectively). All frequency values, namely (6) compound frequency, (7) modifier frequency and (8) head fre-

Table 1: Mean and standard deviation of all the predictors included in the experiment. MCsim: modifier-compound similarity. HCsim: head-compound similarity. MHsim: modifier-head similarity. Comp length: compound length. Comp freq: compound frequency. Mod freq: modifier frequency. Head freq: head frequency.

Predictor	SUB		ATT		CRD		Total	
	mean	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*
MCsim	0.21	0.10	0.18	0.09	0.24	0.13	0.20	0.10
HCsim	0.25	0.11	0.21	0.10	0.24	0.10	0.23	0.11
MHsim	0.13	0.10	0.14	0.09	0.34	0.15	0.16	0.12
Density	0.41	0.07	0.40	0.09	0.39	0.08	0.40	0.08
Entropy	4.50	0.05	4.99	0.06	5.01	0.05	4.50	0.06
Comp length	10.40	2.54	10.90	3.37	10.90	2.94	10.70	2.90
Comp freq	1.87	1.17	2.25	1.38	2.31	0.72	2.06	1.22
Mod freq	5.26	0.86	5.31	1.38	5.32	0.48	5.29	1.05
Head freq	4.97	0.85	4.92	0.89	5.12	0.65	4.97	0.84
PMI	3.53	4.00	4.83	4.80	4.34	3.47	4.11	4.27
Num items	67		49		16		132	

quency were subsequently log-transformed following standard practice in psycholinguistics (Brysbaert et al. 2018).

In addition, we computed (9) Pointwise Mutual Information (PMI) between the constituents as a measure of compound lexicalization. This largely-used association measure (Church & Hanks 1990) compares the probability of co-occurrence of two words in the source corpus with the probability of the two words to co-occur by chance. To illustrate, although the word pair <the apple> is likely much more frequent than <apple juice>, the PMI of the latter will be higher, since the determiner *the* is likely to co-occur very frequently with any noun in the corpus, thus being less informative compared to the pair <apple juice>, whose mutual association is intuitively strong. In particular, the higher the degree of lexical association between two words, the higher the PMI value.

Finally, we included (10) compound length measured as the number of characters making up the string (e.g., *blackboard* has length 10). When present, hyphens were not counted. Descriptive statistics including mean values and standard deviations for all the predictors used in the present experiment are reported in Table 1.

2.6 Data analysis

Our hypothesis is that various, syntax-based classes of compounds might be predicted on the basis of semantic features. If this is correct, our semantic variables will turn out to be reliable predictors of one class over the others. In order to test our hypothesis, we included all the predictors reported in Table 1 in a series of logit regression models that individually estimated the probability of one class over the other. That is, we tested three separate models in the task of predicting one compound type against each of the others: (1) ATT vs SUB, (2) ATT vs CRD, (3) CRD vs SUB.

All analyses were carried out within the R statistical computing environment. We adopted a backward procedure to progressively simplify each statistical model. Starting from a full-factorial model including all the independent variables, predictors were removed one by one when their absence did not significantly lower the overall model fit. At each step, the removal procedure was attempted for the predictor with the largest p-value. The contribution of each parameter to be removed was checked with a goodness-of-fit chi-square test. Finally, atypical outliers were identified and removed using as a criterion 2.5 standard deviation of the residual errors.

3 Results

In order to report results in a readable format, we discuss each model in a separate subsection by also providing a table summarizing the results. In the leftmost part, each table reports the list of variables included in the full-factorial version of each model. In the central part, model-simplification procedure (*Removal order*), chi-square goodness-of-fit test (*Chi-square*) and its results (p) are reported. The rightmost part shows the effects of the variables included in the final model.

3.1 ATT vs SUB

The first model, testing ATT (*halfprice*) against SUB (*bus-stop*) compounds, reliably distinguishes the two classes on semantic bases. As shown in Table 2, SUB is predicted against ATT by the higher semantic similarity between the compound and either the modifier (p=.0182) or the head (p=.0355). That is, the meaning of SUB compounds such as *bus-stop* is found to be more strongly determined by the individual meanings of its constituents compared to ATT compounds like *half-price*, since both the modifier and the head contribute to the overall meaning to a greater extent than either constituents of ATT compounds do. Therefore, the

Table 2: Results of the logit model opposing ATT (1) to SUB (0).

Parameter	Chi-square	p	Removal order	Estimate	z-value	p
(Intercept)	–	–	Not removed	−0.8362	−0.741	0.4587
MCsim	–	–	Not removed	−6.4337	−2.361	0.0182
HCsim	–	–	Not removed	−5.5904	−2.102	0.0355
Density	–	–	Not removed	7.6332	1.916	0.0553
PMI	–	–	Not removed	0.0878	1.885	0.0594
MHsim	2.5275	0.1119	6			
Head freq	0.7637	0.3822	5			
Mod freq	0.4882	0.4847	4			
Comp length	0.8206	0.365	3			
Comp freq	0.0745	0.7848	2			
Entropy	0.005	0.9433	1			

higher the similarity between the compound and either constituents, the higher the probability to have a SUB rather than an ATT compound.

It should be noted that frequency measures, entropy, compound length and the similarity between the two constituents were progressively removed from the model. That is, their effects do not contribute to predict one class over the other. The remaining variables, namely PMI and neighborhood density, are instead included in the final model, even though their effect is only partially reliable (p>.05). Both these measures, anyway, indicate that higher values of both PMI and density are more likely to predict ATT rather than SUB compounds.

3.2 ATT vs CRD

The second model tests ATT (*halfprice*) against CRD (*comedy-drama*) compounds. As reported in Table 3, our model reliably distinguishes between the two classes on the basis of a single, highly significant semantic predictor, namely the semantic similarity between the compound constituents (p=.0002). In particular, the higher the similarity between the modifier and the head of a compound, the higher the probability of having a CRD, rather than an ATT compound. All other variables have been progressively removed from the final model since none of them significantly contribute to the overall goodness fit.

Table 3: Results of the logit model opposing ATT (1) to CRD (0).

Parameter	Chi-square	p	Removal order	Estimate	z-value	p
(Intercept)	–	–	Not removed	5.667	4.094	0.0001
MHsim	–	–	Not removed	−18.182	−3.667	0.0002
Comp length	2.0539	0.1518	9			
Comp freq	1.2866	0.2567	8			
Head freq	0.77	0.3802	7			
HCsim	0.8033	0.3701	6			
MCsim	0.5967	0.4398	5			
Mod freq	0.7654	0.3816	4			
PMI	0.5182	0.4716	3			
Entropy	0.1344	0.7139	2			
Density	0.0346	0.8524	1			

3.3 CRD vs SUB

The third model opposes CRD (*comedy-drama*) to SUB (*bus-stop*) compounds. Also in this case, the model reliably distinguishes between one class and the other on semantic bases. In particular, CRD is predicted over SUB by the degree of semantic similarity between the two constituents (p=.0024). The greater the similarity between the modifier and the head of a compound, the higher the probability of having a CRD rather than a SUB compound. Also, SUB is predicted over

Table 4: Results of the logit model opposing CRD (1) to SUB (0).

Parameter	Chi-square	p	Removal order	Estimate	z-value	p
(Intercept)	–	–	Not removed	−0.0843	−0.037	0.9707
MHsim	–	–	Not removed	36.3465	3.03	0.0024
HCsim	–	–	Not removed	−25.2847	−2.44	0.0146
Comp length	–	–	Not removed	−0.5323	−1.996	0.0459
PMI	0.3016	0.5828	7			
Comp freq	0.3286	0.5665	6			
MCsim	0.2825	0.595	5			
Density	0.1055	0.7453	4			
Entropy	0.2244	0.6357	3			
Mod freq	0.1236	0.7251	2			
Head freq	0.0084	0.9267	1			

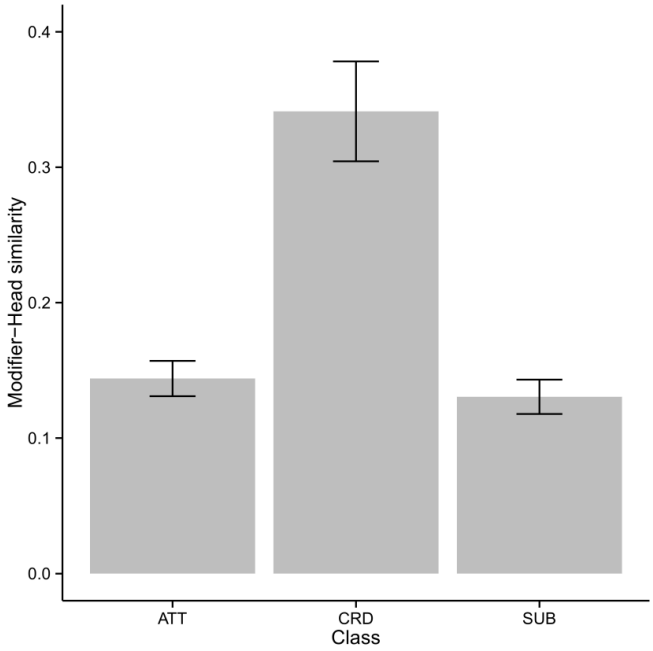


Figure 2: Similarity between modifier and head is predictive of CRD over both ATT and SUB.

CRD by the degree of similarity between the compound and its head (p=.0146). That is, the head constituent contributes more to the overall meaning of SUB compounds (e.g., *stop* in *bus-stop*) than CRD compounds (e.g., *drama* in *comedy-drama*).

In addition to these semantic variables, compound length turns out to be also predictive of one class over the other. As reported in Table 4, in fact, its effect is reliable (p=.0459) and it indicates that longer compounds are more likely to be SUB than CRD. All other parameters were instead progressively removed.

3.4 Overall results

Taken together, these results indicate that the degree of semantic similarity between the compound constituents (i.e. the modifier and the head) is a highly reliable predictor of CRD against both other classes. As shown in the barplot in Figure 2, the higher the similarity between the constituents, the more a compound is likely to be CRD rather than either ATT (p=.0002) or SUB (p=.0024).

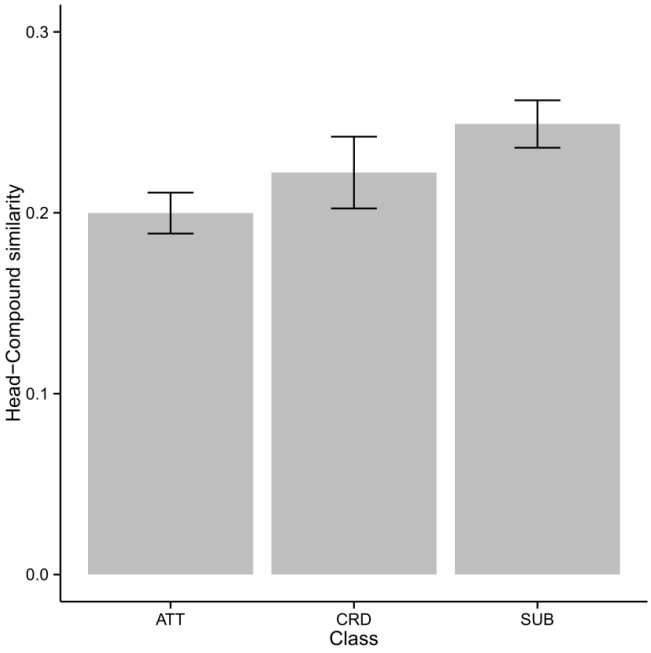


Figure 3: Similarity between the compound and its head is predictive of SUB over both ATT and CRD.

Moreover, the semantic similarity between the compound and its head is a predictive measure of SUB over both other types, as shown in Figure 3. That is, the more the head contributes to the meaning of the overall compound, the more the compound is likely to be SUB rather than either ATT (p=.0355) or CRD (p=.0146).

In order to evaluate the predictive power of each model, we further computed the accuracy with which the items under investigation were correctly assigned to the correct classes. First, we obtained the classes predicted by each logit model. Second, we computed the accuracy of each model by dividing the number of correctly predicted items by the total number of items included in the final model. As a comparison, for each model we also computed the accuracy of a majority baseline obtained by simply dividing the number of cases of the majority class by the total number of cases involved. As reported in Table 5, the best predictive model turned out to be the one opposing CRD vs SUB (.90 accuracy) followed by ATT vs CRD (.85) and ATT vs SUB (.61). These numbers were in line with the pattern of accuracies obtained by the majority baseline, which is sensible to the low number of CRD cases and therefore outputs higher scores for comparisons involving

Table 5: From left to right: overall accuracy of each model as compared to the accuracy of the majority baseline, correctly predicted cases, missed cases. In brackets we report the correct class.

Model	Accuracy (baseline)	Predicted cases	Missed cases
ATT-SUB	0.61 (0.58)	*able-bodied (ATT)* *long-awaited (ATT)* *schoolteacher (SUB)* *racingclub (SUB)*	*commonroom (ATT)* *ironcurtain (ATT)* *underbody (SUB)* *apronstring (SUB)*
ATT-CRD	0.85 (0.75)	*best-equipped (ATT)* *keyword (ATT)* *father-daughter (CRD)* *comedy-drama (CRD)*	*bodypolitic (ATT)* *highschool (ATT)* *typewrite (CRD)* *subject-verb (CRD)*
CRD-SUB	0.90 (0.81)	*comedy-drama (CRD)* *blue-black (CRD)* *bus-stop (SUB)* *cutthroat (SUB)*	*schoolteacher (SUB)* *sleepwalk (SUB)* *typewrite (CRD)* *subject-verb (CRD)*

this class. Though our models always outperformed the baselines, the increase was sensibly lower in ATT vs SUB (+3%) compared to both CRD vs SUB (+9%) and ATT vs CRD (+10%). The limited number of items do not allow us to make any statistically reliable claim on the performance of the classifier. However, our focus is on testing whether the membership in a compound class is affected by a set of theoretically-relevant variables rather than proposing an effective classification algorithm. In this light, our results provided evidence for the effectiveness of these models. At the same time, they suggested that experimenting with more data would be desirable to further validate their power.

Besides accuracies, Table 5 reports some correctly-predicted and missed cases for each of the models.

4 Discussion

The present study investigated whether various, syntax-based classes of compounds (Subordinate, Attributive, Coordinate) can be described in terms of the quantitative, continuous properties of the meaning of the compounds and words belonging to them. To obtain these semantic measures, we generated cDSM representations for a list of compounds for which such classification was available.

By running a series of logit models including both semantic and non-semantic factors as independent variables, we showed that our models are able to reliably capture different classes by means of semantic features.

4.1 On the modifier-head similarity

In particular, we showed that Coordinate compounds like *comedy-drama* are predicted over either Subordinate (*busstop*) or Attributive (*halfprice*) by the higher semantic similarity between the head and the modifier. This finding is consistent with previous evidence from both theoretical linguistics and psychology. Within the lexical semantics approach, Lieber (2009) indeed proposed that Coordinate compounds are generated when the two constituents share almost identical 'bodies' and 'skeletons', that is, when the words to be combined have highly similar meanings.

Also, our finding is in line with several theories of conceptual combination, according to which Hybrid or Conjunctive interpretations would be produced by people for novel combinations which involve highly similar concepts, e.g. *moose-elephant* (see among others Wisniewski 1996). Accordingly, and consistently with our results, Relation-linking interpretations (roughly equivalent to Subordinate compounds) would be instead produced for semantically highly dissimilar pairs, e.g. *apartment-dog*. Since in our model the similarity between the constituents also distinguishes between Coordinate (Hybrid/Conjunctive) and Attributive compounds (Property-mapping), we propose that this result is consistent with the graded description proposed in many conceptual combination theories, where the difference between Property-mapping and Hybrid combinations would be due to an increasing number of both 'commonalities' and 'alignable differences' between the concepts to be combined (Wisniewski 1996).

4.2 On the semantic role of compound constituents

Second, we showed that Subordinate compounds are predicted against Attributive on the basis of the higher similarity between the compound and either constituents. That is, in compositionally-obtained Subordinate compounds both the modifier and the head contribute to a greater extent to the overall meaning than in Attributive ones. Moreover, the similarity between the compound and its head is a reliable predictor of Subordinate over both other classes.

First of all, these findings are again consistent with the lexical semantics literature (Scalise et al. 2005; Lieber 2009). In it, Subordinate compounds are typically characterized by a structure in which the head *selects* its argument. Therefore,

the head contributes more to the overall meaning in this kind of compounds compared to the other classes, where a formal relation between the elements is absent. Also, these results are consistent with the different mechanisms proposed in the conceptual combination literature for Relation-based interpretations (capitalizing on a 'slot-filling' procedure) and Property-based ones (where an 'alignment' process is routinely carried out) (Wisniewski & Gentner 1991; Wisniewski 1996). In a nutshell, the slot-filling procedure would imply a bigger role of the compound *head* compared to the other competing mechanism since, during combination, the head would be just *filled* in one of its 'slots' by the modifier concept.

Interestingly, these findings are also consistent with evidence from embodied cognition (Louwerse 2008). In particular, the embodied conceptual combination theory (ECCo) by Lynott & Connell (2009) proposes that the great majority of relational interpretations (corresponding to Subordinate compounds) are 'non-destructive', namely, they result from the combination of constituent concepts that are left intact during the meshing of their 'affordances'. To illustrate, in this approach the compound *picture book* (i.e. 'a book that has pictures') is non-destructive, since the *pictures* in question are still intact entities in the pages of the *book*. Simplifying somewhat, the combinatorial procedure that leads to Relation-based interpretations (Subordinate) does not heavily modify the meaning of the original constituents, whereas Property-mapping ones (Attributive) are almost always destructive, that is, they involve the 'destruction' of (part of) the constituent concepts. Using an example from Lynott & Connell (2009), the compound *icicle fingers* would reduce *icicle* to a representation of 'coldness' and 'stiffness'. At the same time, the representation of the head (*fingers*) would be switched toward a more figurative, metaphorical meaning, less similar to its prototypical representation (see also, e.g., *iron curtain*). In this light, the similarity between either constituents as independent words and the compound will be generally higher in Relation-based (Subordinate) compared to Property-based interpretations (Attributive), given that the combinatorial procedure of the former type does not heavily modify the meaning of the original constituents. More in general, this observation provides indirect evidence that meaning representations extracted by texts via distributional semantics models can encode grounded information, at least to some extent (Louwerse 2011).

4.3 On attributive compounds

Third, compositionally-derived Attributive compounds are characterized by both a weaker contribution of the constituents in determining the overall meaning

compared to Subordinate and a lower similarity between the constituents compared to Coordinate. This pattern of results is again consistent with Lieber (2009), who proposes that Attributive formations emerge when the semantic features of the constituents are too disparate to be interpreted in a Coordinative way and lack the argument structure that is typical of Subordinate compounds. Accordingly, Attributive compounds would represent a last-resort strategy used when the typical semantic features of the other classes are not satisfied (Lieber 2009). In line with such a description of Attributive compounds emerging *in absentia* of discriminative features is evidence from conceptual combination showing that acceptability judgements for Property-based (Attributive) interpretations to novel compounds (e.g., a *whale boat* is 'a large boat') are slower compared to Relation-based (Subordinate) interpretations (e.g. a *whale boat* is 'a boat for hunting whales') (Gagné 2000). According to Gagné & Spalding (2015), indeed, this would suggest that Relation-based interpretations are the product of an initial compositional process that, in the absence of the features that lead to either a relational interpretation (Subordinate) or a coordinate interpretation (Coordinate), leads to Property-mapping interpretations.

4.4 On the methodology

On the methodological level, it should be mentioned that we used a compositional model to generate representations for a list of compounds whose constituents were either nouns, verbs, adverbs, adjectives, etc. even though, in the training phase, only noun-noun compounds from CELEX were used. This could have represented a weakness for the system, causing the model to be biased toward noun-noun combinations. By looking at the results, however, we observed a similar, remarkably good performance of the model in all items, regardless of the grammatical category of the constituents. This is also clear by inspecting the examples in Table 5, where it can be noted that the parts-of-speech are almost uniformly distributed. However, it might be still possible that a richer training set might lead to even better results, perhaps achieving a better performance in generating meaning representations for less systematic, more opaque compounds. Indeed, we hypothesize that the lower accuracy obtained by the model opposing Attributive vs Subordinate compared to the others might be possibly due to this issue. Finally, we believe that the effectiveness of such an approach might be further validated by testing it on a larger (and possibly balanced with respect to compound type) set of annotated compounds. This, on the one hand, would strengthen the predictive power on the prediction task. On the other hand, it

would allow more extensive, fine-grained analyses on the successes and failures of the models. We plan to further investigate this issue in future work.

4.5 On the effectiveness of cDSMs in predicting compound relations

The effectiveness of our approach in the proposed task is in line with previous work showing that compositional models of distributional semantics are successful in capturing relational information between the constituents of a compound. In particular, our task is related to that of predicting compound semantic interpretation (see section 1.5), where compositionally-obtained representations have been used to assign the correct semantic relation to noun-noun expressions. By experimenting with a number of cDSMs (including the one adopted in this study by Guevara 2010), for example, Dima (2016) obtained results comparable to state-of-the-art in 2 popular datasets (Ó Séaghdha 2007; Tratz & Hovy 2010). Compared to SoA methods, however, Dima (2016) only exploited information from word embeddings, thus proving the effectiveness of both distributed representations and compositional methods. In quantitative terms, our results are not directly comparable due to both the different experimental setting (we did not tackle the task as a classification problem) and the number of relations involved (3 vs either 6 or 43). Moreover, our results cannot be compared with previous work since, to our knowledge, we are the first in proposing this task. However, these studies jointly show that compositional representations are successful in predicting compound relations defined on either *semantic* or *syntactic* bases.

4.6 Final remarks

In conclusion, this study suggests that different compound types identified on syntactic bases can be also defined in terms of continuous, quantitative features of the meaning of the compound and its constituents. We believe that discrete and continuous approaches are two faces of the same coin, the former representing a theoretically motivated, cross-linguistically valuable framework aimed at describing complex linguistic phenomena, the latter providing an interesting way to quantitatively test them. As indicated by our results, compositional models of distributional semantics represent a flexible and powerful way to capture many of these phenomena.

Acknowledgements

We are grateful to Marco Baroni and Laura Vanelli for their valuable feedback during the early stages of the project. We thank Sergio Scalise for providing the MorBoComp subset used in the experiment. We are also grateful to the participants of the First Quantitative Morphology Meeting (Belgrade, July 2015) for the helpful questions and discussion.

References

Baayen, R Harald, Richard Piepenbrock & Léon Gulikers. 1995. *The CELEX lexical database (CD-ROM)*. University of Pennsylvania, Philadelphia, PA: Linguistic Data Consortium.

Bally, Charles. 1950. *Linguistique générale et linguistique française. 3me éd. conforme à la 2me*. Francke.

Baroni, Marco, Raffaela Bernardi & Roberto Zamparelli. 2014. Frege in Space: A Program of Compositional Distributional Semantics. *Linguistic Issues in Language Technology* 9.

Baroni, Marco, Georgiana Dinu & Germán Kruszewski. 2014. Don't count, predict! A systematic comparison of context-counting vs context-predicting semantic vectors. In *Proceedings of the 52nd annual meeting of the association for computational linguistics*, vol. 1, 238–247.

Baroni, Marco, Emiliano Guevara & Vito Pirrelli. 2006. Sulla tipologia dei composti N+N in italiano: principi categoriali ed evidenza distribuzionale a confronto. *Atti del convegno SLI*. 21–23.

Baroni, Marco & Roberto Zamparelli. 2010. Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. In *Proceedings of the 2010 conference on empirical methods in natural language processing*, 1183–1193.

Bauer, Laurie. 2001. Compounding. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals*. Berlin & New York: De Gruyter.

Bauer, Laurie. 2009. Typology of compounds. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford Handbook of Compounding*, chap. 17, 343–356. New York: Oxford University Press.

Bisetto, Antonietta & Sergio Scalise. 2005. The classification of compounds. *Lingue e linguaggio* 4(2). 319–332.

Bloomfield, L. 1933. *Language*. University of Chicago Press.

Booij, Geert. 2005. *The Grammar of Words: An Introduction to Linguistic Morphology* (Oxford linguistics). Oxford: Oxford University Press.

Brysbaert, Marc, Paweł Mandera & Emmanuel Keuleers. 2018. The word frequency effect in word processing: an updated review. *Current Directions in Psychological Science* 27(1). 45–50.

Church, Kenneth Ward & Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational linguistics* 16(1). 22–29.

Cordeiro, Silvio, Carlos Ramisch, Marco Idiart & Aline Villavicencio. 2016. Predicting the compositionality of nominal compounds: giving word embeddings a hard time. In *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 1: long papers)*, vol. 1, 1986–1997.

Costello, Fintan J & Mark T Keane. 2000. Efficient creativity: Constraint-guided conceptual combination. *Cognitive Science* 24(2). 299–349.

Dima, Corina. 2016. On the compositionality and semantic interpretation of English noun compounds. In *Proceedings of the 1st workshop on representation learning for nlp*, 27–39.

Dima, Corina & Erhard Hinrichs. 2015. Automatic noun compound interpretation using deep neural networks and word embeddings. In *Proceedings of the 11th international conference on computational semantics*, 173–183.

Dressler, Wolfgang U. 2006. Compound types. In Gary Libben & Gonia Jarema (eds.), *The representation and processing of compound words*, 23–44. Oxford: Oxford University Press Oxford.

Fabb, Nigel. 1998. Compounding. In A Spencer & A M Zwicky (eds.), *Handbook of morphology*, 66–83. Oxford: Blackwell.

Fanselow, Gisbert. 1981. Zur Syntax und Semantik der Nominalkomposition. *Tübingen: Niemeyer*.

Farahmand, Meghdad, Aaron Smith & Joakim Nivre. 2015. A multiword expression data set: annotating non-compositionality and conventionalization for english noun compounds. In *Proceedings of the 11th workshop on multiword expressions*, 29–33.

Fares, Murhaf, Stephan Oepen & Erik Velldal. 2018. Transfer and Multi-Task Learning for Noun-Noun Compound Interpretation. *arXiv preprint arXiv:1809.06748*.

Firth, John Rupert. 1957. *Papers in Linguistics, 1934-1951*. Oxford, UK: Oxford University Press.

Gagné, Christina L. 2000. Relation-based combinations versus property-based combinations: A test of the CARIN theory and the dual-process theory of conceptual combination. *Journal of Memory and Language* 42(3). 365–389.

Gagné, Christina L & Thomas L Spalding. 2006. Conceptual combination: Implications for the mental lexicon. In *The representation and processing of compound words*, chap. 7, 145–168. New York: Oxford University Press Oxford.

Gagné, Christina L & Thomas L Spalding. 2009. Constituent integration during the processing of compound words: Does it involve the use of relational structures? *Journal of Memory and Language* 60(1). 20–35.

Gagné, Christina L & Thomas L Spalding. 2015. Semantics, Concepts, and Metacognition: Attributing Properties and Meanings to Complex Concepts. In *Semantics of complex words*, 9–25. Springer.

Guevara, Emiliano. 2010. A regression model of adjective-noun compositionality in distributional semantics. In *Proceedings of the 2010 workshop on geometrical models of natural language semantics*, 33–37.

Günther, Fritz & Marco Marelli. 2016. Understanding Karma Police: The Perceived Plausibility of Noun Compounds as Predicted by Distributional Models of Semantic Representation. *PloS one* 11(10). e0163200.

Harris, Zellig S. 1954. Distributional structure. *Word*.

Haspelmath, Martin. 2002. *Understanding Morphology*. London: Edward Arnold.

Ji, Hongbo, Christina L Gagné & Thomas L Spalding. 2011. Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. *Journal of Memory and Language* 65(4). 406–430.

Kuperman, Victor. 2009. Revisiting semantic transparency in English compound words. In *6th international morphological processing conference*. Turku, Finland.

Landauer, Thomas K & Susan T Dumais. 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review* 104(2). 211.

Lazaridou, Angeliki, Eva Maria Vecchi & Marco Baroni. 2013. Fish transporters and miracle homes: How compositional distributional semantics can help NP parsing. In *Proceedings of the 2013 conference on empirical methods in natural language processing*.

Lees, Robert B. 1960. The Grammar of English Nominalizations. *Folklore & Linguistics. Bloomington: Research Center in Anthropology*.

Levi, Judith N. 1978. *The syntax and semantics of complex nominals*. New York: Academic Press.

Lieber, Rochelle. 2009. A lexical semantic approach to compounding. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford Handbook of Compounding*, chap. 5, 78–104. New York: Oxford University Press.

Lieber, Rochelle & Pavol Štekauer (eds.). 2009. *The Oxford Handbook of Compounding*. New York: Oxford University Press.

Louwerse, Max M. 2008. Embodied relations are encoded in language. *Psychonomic Bulletin & Review* 15(4). 838–844.

Louwerse, Max M. 2011. Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science* 3(2). 273–302.

Lynott, Dermot & Louise Connell. 2009. Embodied conceptual combination. *Embodied and grounded cognition*. 79.

Lynott, Dermot & Michael Ramscar. 2001. Can we model conceptual combination using distributional information? In *Proceedings of the 12th Irish conference on artificial intelligence and cognitive science*, 1–10.

Marchand, Hans. 1969. *The categories and types of present-day English word-formation: A synchronic-diachronic approach*. Beck.

Marelli, Marco & Marco Baroni. 2015. Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological review* 122(3). 485.

Marelli, Marco, Georgiana Dinu, Roberto Zamparelli & Marco Baroni. 2014. Picking buttercups and eating butter cups: Spelling alternations, semantic relatedness, and their consequences for compound processing. *Applied Psycholinguistics*. 1–19.

Marelli, Marco, Christina L Gagné & Thomas L Spalding. 2017. Compounding as Abstract Operation in Semantic Space: Investigating relational effects through a large-scale, data-driven computational model. *Cognition* 166. 207–224.

Marelli, Marco & Claudio Luzzatti. 2012. Frequency effects in the processing of Italian nominal compounds: Modulation of headedness and semantic transparency. *Journal of Memory and Language* 66(4). 644–664.

Mikolov, Tomas, Kai Chen, Greg Corrado & Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. http://arxiv.org/abs/1301.3781/.

Mitchell, Jeff & Mirella Lapata. 2010. Composition in distributional models of semantics. *Cognitive science* 34(8). 1388–1429.

Ó Séaghdha, Diarmuid. 2007. Annotating and learning compound noun semantics. In *Proceedings of the 45th annual meeting of the acl: student research workshop*, 73–78.

Padó, Sebastian & Mirella Lapata. 2007. Dependency-based construction of semantic space models. *Computational Linguistics* 33(2). 161–199.

Reddy, Siva, Diana McCarthy & Suresh Manandhar. 2011. An empirical study on compositionality in compound nouns. In *Proceedings of 5th international joint conference on natural language processing*, 210–218.

Roller, Stephen, Sabine Schulte im Walde & Silke Scheible. 2013. The (un) expected effects of applying standard cleansing models to human ratings on compositionality. In *Proceedings of the 9th workshop on multiword expressions*, 32–41.

Salehi, Bahar, Paul Cook & Timothy Baldwin. 2014. Using distributional similarity of multi-way translations to predict multiword expression compositionality. In *Proceedings of the 14th conference of the european chapter of the association for computational linguistics*, 472–481.

Salehi, Bahar, Paul Cook & Timothy Baldwin. 2015. A word embedding approach to predicting the compositionality of multiword expressions. In *Proceedings of the 2015 conference of the north american chapter of the association for computational linguistics: human language technologies*, 977–983.

Scalise, Sergio & Antonietta Bisetto. 2009. The classification of compound. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford Handbook of Compounding*, chap. 3, 34–53. New York: Oxford University Press.

Scalise, Sergio, Antonietta Bisetto & Emiliano Guevara. 2005. Selection in Compounding and Derivation: Morphology and Its Demarcations. In Wolfgang U. Dressler, Dieter Kastovsky, Oskar E. Pfeiffer & Franz Rainer (eds.), *Morphology and its demarcations*, chap. 14, 133–150. Amsterdam: John Benjamins Publishing Company.

Schulte im Walde, Sabine, Stefan Müller & Stefan Roller. 2013. Exploring vector space models to predict the compositionality of German noun-noun compounds. In *2nd joint conference on lexical and computational semantics, volume 1: proceedings of the main conference and the shared task: semantic textual similarity*, vol. 1, 255–265.

Shwartz, Vered & Ido Dagan. 2018. Paraphrase to explicate: revealing implicit noun-compound relations. In *Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: long papers)*, 1200–1211. Melbourne, Australia. http://aclweb.org/anthology/P18-1111.

Tratz, Stephen & Eduard Hovy. 2010. A taxonomy, dataset, and classifier for automatic noun compound interpretation. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, 678–687.

Turney, Peter D & Patrick Pantel. 2010. From frequency to meaning: vector space models of semantics. *Journal of artificial intelligence research* 37(1). 141–188.

Van de Cruys, Tim, Stergos Afantenos & Philippe Muller. 2013. MELODI: A supervised distributional approach for free paraphrasing of noun compounds. In *2nd joint conference on lexical and computational semantics, volume 2: proceedings of the 7th international workshop on semantic evaluation*, vol. 2, 144–147.

Vecchi, Eva Maria, Marco Baroni & Roberto Zamparelli. 2011. (linear) maps of the impossible: capturing semantic anomalies in distributional space. In *Proceedings of the workshop on distributional semantics and compositionality*, 1–9.

Wang, Hsueh-Cheng, Li-Chuan Hsu, Yi-Min Tien & Marc Pomplun. 2014. Predicting raters' transparency judgments of English and Chinese morphological constituents using latent semantic analysis. *Behavior research methods* 46(1). 284–306.

Warren, Beatrice. 1978. Semantic patterns of noun-noun compounds. *Acta Universitatis Gothoburgensis. Gothenburg Studies in English Goteborg* 41. 1–266.

Wisniewski, Edward J. 1996. Construal and similarity in conceptual combination. *Journal of Memory and Language* 35(3). 434–453.

Wisniewski, Edward J & Dedre Gentner. 1991. On the combinatorial semantics of noun pairs: minor and major adjustments to meaning. *Advances in Psychology* 77. 241–284.

Wisniewski, Edward J & Bradley C. Love. 1998. Relations versus properties in conceptual combination. *Journal of Memory and Language* 38(2). 177–202.

Zanzotto, Fabio Massimo, Ioannis Korkontzelos, Francesca Fallucchi & Suresh Manandhar. 2010. Estimating linear models for compositional distributional semantics. In *Proceedings of the 23rd international conference on computational linguistics*, 1263–1271.

Chapter 3

Compositionality in English deverbal compounds: The role of the head

Gianina Iordăchioaia
University of Stuttgart

Lonneke van der Plas
University of Malta

Glorianna Jagfeld
University of Stuttgart

This paper is concerned with the compositionality of deverbal compounds such as *budget assessment* in English. We present an interdisciplinary study on how the morphosyntactic properties of the deverbal noun head (e.g., *assessment*) can predict the interpretation of the compound, as mediated by the syntactic-semantic relationship between the non-head (e.g., *budget*) and the head. We start with Grimshaw's (1990) observation that deverbal nouns are ambiguous between compositionally interpreted Argument Structure Nominals, which inherit verbal structure and realize arguments (e.g., *the* ***assessment*** *of the budget by the government*), and more lexicalized Result Nominals, which preserve no verbal properties or arguments (e.g., *The* ***assessment*** *is on the table*). Our hypothesis is that deverbal compounds with Argument Structure Nominal heads are fully compositional and, in our system, more easily predictable than those headed by Result Nominals, since their compositional make-up triggers an (unambiguous) object interpretation of the non-heads.
Linguistic evidence gathered from corpora and human annotations, and evaluated with machine learning techniques supports this hypothesis. At the same time, it raises interesting discussion points on how different properties of the head contribute to the interpretation of the deverbal compound.

Gianina Iordăchioaia, Lonneke van der Plas & Glorianna Jagfeld. 2019. Compositionality in English deverbal compounds: The role of the head. In Sabine Schulte im Walde & Eva Smolka (eds.), *The role of constituents in multiword expressions: An interdisciplinary, cross-lingual perspective*, 57–98. Berlin: Language Science Press. DOI:??

1 Introduction

This paper contributes a study on how constituents influence the compositionality of multiword expressions from the perspective of deverbal compounds in English with a focus on the role of their head nouns.

1.1 Deverbal compounds (DCs)

DCs are noun-noun compounds, whose head is deverbal, as illustrated in (1). In contrast to root compounds (RCs) (see (2)), whose head nouns are typically simple (non-derived), DCs usually receive an interpretation in which the non-head establishes a syntactic-semantic relationship with the verb from which the deverbal noun is derived (i.e., as a direct object, subject or other argument/adjunct). RCs often receive a fixed interpretation (see (2a)) or one depending on the immediate context (see (2b)). *Tomato bag* in (2b) may refer to a bag of tomatoes, a bag having the shape or color of a tomato, or any other connection between a bag and tomatoes mentioned in previous context. The same holds for *jelly bottle*.

(1) a. *budget* assess**ment** – to assess (a) *budget*(s) (Object)
b. *police* question**ing** – *police* questions sb. (Subject)
c. *college* educat**ion** – to educate sb. in *college* (Adjunct)

(2) a. train station, bookstore
b. tomato bag, jelly bottle

Nominal DCs may be headed by deverbal nouns built with a variety of suffixes, including those that form participant-denoting nominals, as in (3) (see Lieber 2016: 73). For reasons that will be given in §3.2, we concentrate here on DCs headed by eventive deverbal nominals as in (1), formed by means of the suffixes *-al*, *-ance*, *-(at)ion*, *-ing*, and *-ment*.

(3) a. dog train**er**, flight attend**ant**
b. bank employ**ee**, award nomin**ee**

1.2 Argument Structure Nominals and Result Nominals

Grimshaw (1990) points out that the majority of deverbal nouns exhibit an ambiguity between an Argument Structure Nominal (ASN) reading, which perfectly

mirrors the corresponding verb phrase with its argument structure, and a Result Nominal reading (RN), which is more lexicalized and departs from the base verb at various degrees.[1] The crucial difference between the two originates in the availability of verbal event structure, which enforces and constrains argument realization in ASNs (see (6) below), and its absence in RNs. The examples in (4) illustrate the two readings, building on Grimshaw (1990: 49).

(4) a. The **examination**/exam was [on the table /in the bag]. (RN)
b. The **examination**/*exam *of the patients* took a long time. (ASN)
c. * The **examination** *of the patients* was [on the table /in the bag]. (ASN)

In the absence of the object argument *of the patients*, the noun *examination* receives a RN reading, in which, similarly to *exam*, it denotes a concrete entity, which can lie on a table or be in a bag (see (4a)). When the argument is realized, the synonymy with *exam* is lost, and the noun behaves like a nominalized verb, expressing an event, which can take a long time (see (4b)), but cannot be on a table or in a bag (see (4c)). In combination with *exam*, the phrase *of the patients* in (4b) could receive a possessive interpretation, i.e., the exam that belongs to the patients, but not that of an object argument of an examining event, since *exam* lacks such a reading. A similar interpretation would be possible in (4c) with *examination* on its RN reading.[2]

1.3 Compositionality and transparency in deverbal compounds

Compositionality is a long debated issue in theoretical linguistics, and its first formalization is offered in Montague's (1970) *Universal Grammar*. A simple formulation of the principle of compositionality in this tradition is given in (5).

(5) The Principle of Compositionality (PoC, Partee 1984: 281)
The meaning of an expression is a function of the meanings of its parts and of the way they are syntactically combined.

According to the PoC in (5), the interpretation of a complex expression relies on the individual meanings of its parts and their syntactic combination. Leaving technical details aside, an expression like *to kick the bucket* will be interpreted compositionally from the meanings of the verb *to kick* and of the noun phrase

[1] For the sake of simplicity, we leave aside Grimshaw's third possible reading of deverbal nouns as Simple Event Nominals, since, from the perspective of the properties we consider here, they pattern with RNs and contrast with ASNs in similar ways.
[2] In her examples, Grimshaw strictly uses *of the patients* on its argument interpretation.

the bucket, via a verb–direct object syntactic relationship and the corresponding semantic relation. On this compositional reading, this expression is semantically transparent both with respect to the meanings of the parts and the syntactic-semantic relationship: the object *the bucket* is semantically interpreted as a patient of the kicking. However, *to kick the bucket* also has the idiomatic reading *to die*, on which neither the meanings of the two parts, nor any syntactic relationship between them can be compositionally retrieved. There is nothing particular about kicking or buckets or the verb–direct object relationship between them to be found in the meaning of *to die*. This reading is non-compositional and opaque.

Some idiomatic expressions, however, may be partially compositional. For instance, in *to spill the beans* 'to divulge a secret', the verb–object relationship is preserved in the idiom meaning and, while the object *beans* is lexico-semantically unrelated to *secret*, the verb *to spill* shares lexical semantic properties with *to divulge* (i.e., 'to let out'), which can be viewed as its figurative meaning. In this expression, the non-head is opaque, the head is partially transparent, and the relationship is compositional and transparent. The head is only partially transparent because it is ambiguous and the meaning *divulge* is not its basic meaning.

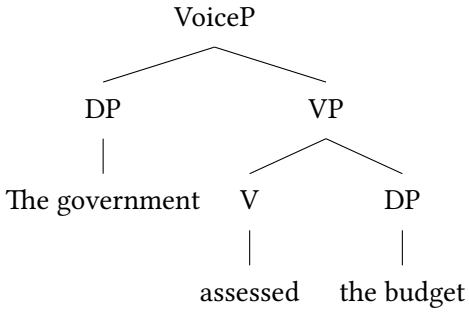
Deverbal compounds offer another pattern of expressions that are not fully compositional – yet, one different from the idioms above. The interpretation of DCs usually relies on a syntactic-semantic relationship between the base verb of their head noun and their non-heads, as shown in (1). Unlike in the corresponding verbal phrases, however, the syntactic relationship is not overt in DCs: e.g., *budget* in the DC *budget assessment* is not marked with accusative case as in the corresponding verb phrase in (1a), and *police* in *police questioning* is not marked by nominative case in (1b).[3] In the absence of overt marking, it is often unclear how to interpret the non-head of a DC, as, for instance, in *police killing*, where *police* could be either the object or the subject of *kill*. The indeterminacy of the syntactic relationship leads to ambiguity, which reduces the transparency of DCs from the perspective of syntactic compositionality, even though the meanings of the parts are transparent (by contrast with *beans* in *to spill the beans* or *kick* and *bucket* in the idiom *to kick the bucket*).

Yet, following the PoC and the compositional make-up of a sentence, if a particular DC is built up compositionally in parallel to the corresponding verbal phrase, then an object interpretation of the non-head is expected. This is the thesis we will follow and support here. But why does an object non-head indicate compositionality and, e.g., a subject does not? The reason follows from simple

[3]In a morphologically poor language like English, case marking comes from the syntactic position of the noun phrase, which is also missing in DCs, given their fixed word order.

sentence structure. Transitive verbs form immediate constituents with their direct objects but not with their subjects, which is why in sentence structure we first form a VP from the verb and its object, and the subject attaches afterwards, usually under a different projection such as VoiceP (or little vP), as in (6) (see Chomsky (1995) and Kratzer (1996) for a discussion on the differences between objects and subjects with respect to the event structure of verbs).

(6)



A DC based on the construction in (6) contains *two* nouns: one is the head derived from the verb and the other is the non-head. The latter can realize only one of the two arguments of the verb. Given the hierarchical structure in (6), this must be the object: see *budget assessment*. Nothing prevents the original subject to be realized as a non-head (e.g., *government assessment*). In that case, however, the DC does not follow the compositional make-up in (6), since the object is missing and the subject cannot form a constituent with the transitive verb alone. Such a DC will be interpreted by means of world knowledge, similarly to RCs as in (2). From this perspective, the subject behaves just like an adjunct/modifier, since it does not play any role in the compositional make-up of the DC.[4]

The importance of compositionality in language use is undebatable: without recursive compositional rules, speakers would not be able to produce and understand infinitely many sentences (Dowty (2007)). That compositionality in DCs imposes an object interpretation, as predicted by the structure in (6), is supported by the fact that the default reading of a possibly ambiguous DC like *police killing* is that with an object non-head; the subject reading becomes available if established by a particular context, as, e.g., recent discussion in the U.S. about police killing unarmed civilians. Similarly, out of context, *student evaluation* also receives an object interpretation. The subject reading is brought about by a particular social environment in which people talk about students evaluating their

[4]Indirect objects are included here as well, since they also attach to the verb after the direct object does: see Larson (1988).

teachers. Moreover, as shown in the linguistic literature (e.g., Grimshaw (1990); Borer (2013); Iordăchioaia et al. (2017)), if a DC type is compositionally derived from a VP, it should also be fully productive: that is, any verb–object combination should be able to form a compositional DC, which is confirmed, for instance, by (7b). By contrast, not any subject–verb combination can form a DC: the non-heads in (7c) may at best receive a peculiar object interpretation, but not the subject reading of the corresponding sentence in (7a).

(7) a. A *boy*/*girl* broke the **window**/**pen**.
b. **window** breaking, **pen** breaking
c. * *boy* breaking, **girl* breaking

To summarize, ambiguous DCs as in (8) below are partially opaque, as the relationship between the two nouns is not explicit, and may receive several interpretations. However, if the DC is interpreted compositionally (in parallel to the verbal construction), it will be fully transparent and involve an object reading. The task remains to find independent evidence for the compositionality of a DC. In this respect, we will follow Grimshaw's (1990) distinction from §1.2 concerning the head nouns of DCs, as specified below in §1.5 and §4.1.

(8) a. policy/police/radio announcement (Object/Subject/Adjunct)
b. marketing approval, security assistance (Object/Subject/Adjunct)

1.4 Terminology

Before we introduce our research program, a few terminological clarifications are in place. The term *compositionality* is often used without particular focus on the syntactic-semantic relationship between the parts of the complex expression, an aspect that is of crucial importance in our study. Natural Language Processing literature on (root) noun-noun compounds, for instance, occasionally speaks of *compositionality ratings*, in which annotators evaluate how accessible the lexical meaning of the two nouns is in the overall meaning of the compound (see §2.2.2 for details and references). This notion of compositionality is similar to what we call below *lexico-semantic transparency*.

A notion of *compositionality* that is closer to ours appears in some Distributional Semantics (DS) approaches, which, in view of the PoC in (5), seek to identify linguistically-informed composite functions to combine the individual parts of complex expressions (e.g., Marelli & Baroni (2015); Baroni & Zamparelli (2010)). Like us, these authors take a closer look at the relationship between the parts;

however, their focus is more on the technical implementation (i.e., the DS correspondent of function application from theoretical linguistics) rather than on the linguistically relevant constraints that are at play. Although we share the interest in the relationship between the parts with this literature, we are not concerned with the technical details of the function, but with how this relationship interacts with other morphosyntactic properties of the head, as explained in §1.5.

We use the terminology as follows: *compositional* refers to DCs that encode the structure in (6). Some may call this 'syntactic compositionality'. The term *transparent* is broader and allows two specifications. First, *lexico-semantically transparent* characterizes the compounds whose parts are semantically fully recoverable from the compound meaning. These include all DCs as in (1) and (8), as well as some RCs as those in (2).[5] Second, what we would call *compositionally transparent* applies to DCs that, besides being lexico-semantically transparent, also follow the structure in (6). These correspond to our *compositional DCs*, since all the DCs we consider here are lexico-semantically transparent.

1.5 Our contribution

We start with the assumption that an important source of ambiguity in DCs such as in (8) is the ambiguity of their deverbal head nouns as in (4) and the correlated ambiguous relationship that they establish with the non-heads. The non-head is entirely transparent in DCs: its lexical semantics is present in the DC meaning, and, as an argument or adjunct, it brings no syntactic constraints to influence its syntactic-semantic relationship with the head noun. By contrast, the head noun is more complex. Its lexical semantics is also visible in the DC; yet, following Grimshaw's distinction in (4), its ambiguity between ASN and RN readings has a great impact on its syntactic-semantic relationship with the non-head. As perfect transpositions of verb phrases, ASNs follow the compositional structure in (6) and require objects to be realized first. RNs maintain only remote lexical connections to the verb base and do not inherit their compositional structure. Thus, RNs impose no syntactic requirements on the non-heads and are compatible with any syntactic-semantic relationship allowed by their lexical semantics.

Following this reasoning, our hypothesis is that DCs with ASN heads will obey the constituent structure in (6) and realize only objects as non-heads. These DCs will be both compositional and lexico-semantically transparent. DCs whose heads are RNs do not respect this structural condition and allow any interpretation that a context or world knowledge provide – whether related to the base

[5]Other RCs like *hogwash* are substantially less transparent: see the previous literature in §2.2.

verb or not (cf. *police building* 'building that hosts the police department'). In this respect, DCs headed by RNs are semantically similar to RCs; their deverbal morphology is irrelevant for their interpretation, since they are lexicalized. Such DCs are lexico-semantically transparent, but they are not fully compositional.

To test this hypothesis, we use a series of morphosyntactic properties that Grimshaw argued to be ASN-specific (see §4.1) and check their presence in the behavior of DC heads, on the basis of evidence from a large corpus of naturally occurring text. Since it is not a given fact that the ASN-features defined by Grimshaw can be reliably informed by corpora, we also gathered human judgments on ASN-hood – namely, we asked annotators to indicate in how far the deverbal head refers to a process (or verbal event). By asking annotators directly for their judgments, we try to get an estimate for the latent variable that underlies the ASN properties defined by Grimshaw. We use these different types of data as features in a logistic regression classifier, by which we aim to predict the syntactic-semantic relation between the head and the non-head. These results are compared with the manually annotated interpretation of DCs.

Given our hypothesis and methodology, we expect that the ASN-features extracted from the corpus, as well as that based on human judgments, will point to an object interpretation of the DC (as predicted by (6)) and will have high predictive power in determining whether the DC's non-head is an object or not. A high predictive power of the features will additionally show us that compositionality is an important aspect in the disambiguation of DCs.

First of all, our results indicate that all the ASN-hood features have predictive power above the chance level when tested individually and together. The most stable individual features point to an object interpretation, as expected under our hypothesis. Second, the ablation experiments show that many features overlap in the identification of ASN-hood, inviting to theoretical reflection on the individual contribution of these features. Third, the best feature is the manual annotation of ASN-hood, which confirms the importance of this property for interpreting DCs; it also indicates that either the morphosyntactic features are comparably weaker or our corpus did not offer enough material for better results. Fourth, some weaker features raise stimulating questions especially relevant for linguistic investigation.

Our study investigates transparency strictly from the perspective of the compositional structure in (6). The degree of (lexico-semantic) transparency of DCs that do not receive such a verb-related compositional interpretation (i.e., those headed by RNs) goes beyond the scope of our present study and must be left for a future endeavor. As mentioned above, the role of world knowledge and context is essential for such DCs. Therefore, such an investigation would need to employ

a different methodology, more similar to that pursued in several computational studies as presented in §2.2.2. We also do not aim to measure speaker intuitions about the transparency degrees of DCs (as done in some of these computational approaches), although it would be interesting to compare such ratings with our relation-based annotations in the future. Our present study conceptually differs from these computational approaches, as it addresses the transparency of DCs from a structural perspective. We use insights from theoretical linguistics on the morphosyntactic properties of the deverbal noun heads of DCs and general principles of syntax-semantics mapping, and test these theoretical hypotheses with corpus-based and computational methods.

We start with an overview of relevant previous studies from Theoretical Linguistics (TL) and Natural Language Processing (NLP) in §2. Sections §3 and §4 describe our data collection and methodology; §5 presents our experiments, followed by a discussion in §6. We draw our conclusions in §7.

2 Previous literature

In §2.1 we introduce the main theoretical concepts that have guided our investigation and briefly refer to previous analyses of DCs relevant to our assumptions. §2.2 presents the NLP literature on deverbal and root noun-noun compounds and the extent to which these studies can be compared with ours.

2.1 Theoretical approaches to DCs

Deverbal compounds have been at the forefront of theoretical linguistics since the early days of generative grammar. Especially beginning with the 70s, after Chomsky's (1970) *Remarks on nominalization*, the theme of the theoretical debate has been whether word formation is part of the syntax or the lexicon. Syntactic approaches have argued that DCs behave systematically enough to be accounted for by syntactic rules (e.g., Roeper & Siegel (1978); Ackema & Neeleman (2004)); lexicalist approaches have pointed out peculiar properties of DCs, which would require their analysis as part of the lexicon (e.g., Selkirk (1982); Lieber (2004)).

The syntax vs. lexicon debate is relevant for our study in so far as recognizing a syntactic component in DCs leads to their compositional analysis, while specifying lexical rules for them suggests that they are like RCs and lack a systematic morphosyntax that preserves phrase-like compositionality. Meanwhile, both theoretical trends have argued for both kinds of analysis of DCs, and we will abstract away from the type of framework to focus on the properties of DCs.

Noteworthily, in theoretical studies the problem of compositionality in DCs is not addressed with respect to the contribution of the two individual nouns as in recent NLP studies (see §2.2). If available at all, implications on compositionality come indirectly from the claims on the make-up of DCs and the structural relationship between their parts as in (6) (see §2.1.2).

2.1.1 Morphosyntactic properties of ASNs

In support of the contrast illustrated in (4), Grimshaw (1990) argues that deverbal nouns in their ASN reading exhibit a special morphosyntactic behavior, which is not shared by RNs. Table 1 is a summary of the main contrastive properties of ASNs (vs. RNs) from Grimshaw (1990) that are relevant for our study, adapted from Alexiadou & Grimshaw (2008: 3). These properties are positively specified for ASNs only, since RNs behave like non-derived lexical nouns and do not present any such particularities. The reasoning is that ASNs have verbal properties (i.e., event structure as in (6)), which will impose restrictions on their nominal behavior (e.g., must appear in the singular) or make them compatible with verb-specific modifiers (e.g., aspectual adverbials).

Table 1: Morphosyntactic properties of ASNs vs. RNs

	Morphosyntactic property	ASN	RN
i.	Obligatory object arguments realized as *of*-phrases	Yes	No
ii.	Agent-oriented modifiers (*deliberate*, *intentional*, *careful*)	Yes	No
iii.	*By*-phrases are (subject) arguments	Yes	No
iv.	Aspectual *in*/*for-X-time* adverbials	Yes	No
v.	*Frequent*, *constant* appear with singular	Yes	No
vi.	Must appear in the singular	Yes	No

The realization of object arguments is a necessary and sufficient condition for ASNs. It indicates the presence of verbal event structure, which associates with the other ASN-properties. However, the morphosyntactic means to introduce an object argument in nominals is an *of*-phrase, which may also express possession. Given this ambiguity, using an *of*-phrase in combination with other ASN-properties is more reliable. For instance, in (4b), the predicate *took a long time* requires an event as a subject, which shows that *the examination of the patients* is an ASN, while *the exam of the patients* is not. As mentioned above, in the latter case *of the patients* expresses a possessor of the entity *exam*.

Agent-oriented adjectives like *deliberate, intentional, careful* are also taken by Grimshaw (1990: 51–52) to depict ASNs. Like *of*-phrases, possessive marking is ambiguous between expressing subject arguments, as in (9b), and possessive modifiers, as in (9c)). Agentive modifiers, however, require verbal event structure with a subject (agent) argument, which cannot be available in the absence of the object argument in (9a) and (9c)) (cf. the hierarchy in (6)). The contrast between (9a) and (9b) shows that the possessive *the instructor's* cannot introduce the subject argument, if the object argument is not realized.

(9) a. * The instructor's *intentional/deliberate* examination took a long time.
b. The instructor's *intentional/deliberate* examination *of the papers* took a long time. (ASN)
c. the instructor's (**intentional/*deliberate*) book

In ASNs, *by*-phrases have a function similar to that of the possessive in (9b): they introduce the subject argument. Yet, like the possessive and *of*-phrases, *by*-phrases may also introduce modifiers. In (10a), the *by*-phrase acts as a modifier of the lexical noun *book*, which has no event structure. In (10b), however, it introduces the subject argument of an ASN, the same way the possessive does in (9b). (10c) is ungrammatical, because the agent-oriented modifiers *intentional/deliberate* require a subject argument, which the *by*-phrase cannot introduce in the absence of event structure and the object: (10c) parallels (9a).

(10) a. a book *by Chomsky*
b. The intentional/deliberate examination *of the papers by the instructor* took a long time. (ASN)
c. * The intentional/deliberate examination *by the instructor* took a long time.

Given the verbal event structure and the correlated aspectual properties of ASNs, they are expected to allow aspectual adverbials and to obey the aspectual restrictions of their base verbs. In (11a), the telic verb *destroy* allows *in-* but not *for*-adverbials. The correlated ASN in (11b) exhibits the same constraint. By contrast, simple nouns that lexically denote events such as *trip, process* are incompatible with such modifiers in (11c), although they occupy time, as shown by (11d). The latter pattern with RNs (Grimshaw 1990: 58–59).

(11) a. The bombing destroyed the city *in/*for only 2 days*.
b. The total destruction of the city *in/*for only 2 days* appalled everyone.

c. * The process/John's trip *in/for 5 hours*
d. The process/John's trip *took 5 hours.*

Finally, Grimshaw argues that, due to their verbal structure, ASNs, in general, disallow plural marking, and when plural is available it indicates a RN reading. This is illustrated in (12) from Grimshaw (1990: 54). Related to this and the aspectual contrast in (11), Grimshaw notes that aspectual modifiers like *constant, frequent* will combine with a singular ASN, but with a plural RN. These modifiers require habitual/iterative aspect, which is made available by the event structure of ASNs, but not by the lexicalized RNs. The latter need the plural to contribute the iterative meaning: see (13a)/(13c)–(13b).

(12) a. The *assignments* were long. (RN)
b. * The *assignments of the problems* took a long time. (ASN)
c. The *assignment of that problem* always causes problems. (ASN)

(13) a. * The *constant* assignment is to be avoided. (RN)
b. The *constant* assignment of unsolvable problems is to be avoided. (ASN)
c. The *constant* assignments were avoided by students. (RN)

In (9) to (13), the contrasts between ASNs and RNs are clear. Yet, depending on the lexical semantics of the individual nouns, the application of these tests may exhibit quite a bit of variation, which led many to challenge Grimshaw's generalizations. For instance, Alexiadou et al. (2010) show that in some languages, ASNs may pluralize provided particular aspectual properties, while Grimm & McNally (2013) and Lieber (2016) challenge some of Grimshaw's claims with counterexamples attested in corpora. However, a general tendency of ASNs to exhibit the properties in Table 1 cannot be denied. At least so far, no corpus study has offered a quantitative analysis to prove that these properties are irrelevant for ASNs. From this perspective, our study can also be viewed as testing the relevance of these properties on the basis of deverbal compounds, which, according to Grimshaw, are headed by ASNs (see §2.1.2).

2.1.2 Deverbal compounds between ASNs and RNs: Grimshaw (1990)

Let us now consider DCs from the perspective of the documented ASN vs. RN contrast. We focus on Grimshaw's analysis of DCs and on Borer (2013), the latter of which reviews Grimshaw's arguments to support an opposite position.

In her study of nominalization, Grimshaw (1990) argues that the heads of DCs (i.e., her *synthetic compounds*) are ASNs. Her reasoning relies on the observation that DCs obey argument structure constraints in the realization of their non-heads. In her model of argument realization, she proposes a hierarchy of argument roles as in (14), such that the lower arguments (from right to left) must be realized syntactically before the higher ones. This means that the theme, i.e., the syntactic direct object, must be realized before the goal (indirect object) and the agent (subject). This thematic hierarchy reminds us of the constituent structure of verb phrases in (6).

(14) Agent (subject) > Goal (indirect object) > Theme (direct object)

Grimshaw argues that DCs obey the hierarchy in (14), since they disallow non-heads that realize other arguments than the theme (object). (15) repeats two of her examples. Her explanation is that, when occurring in DCs, deverbal nouns such as *giving* and *reading* are disambiguated to an ASN interpretation.

(15) a. They give **gifts** to *children.*
DC: **gift**-giving to children vs. **child*-giving of gifts

b. *Students* read **books**.
DC: **book**-reading by students vs. **student*-reading of books

In contrast to suffix-based deverbal nouns as in (15), she considers zero-derived nouns like *a sting* and *a bite* to always be RNs. She shows that the compounds these may head need not obey the hierarchy in (14) and allow agent non-heads. The grammatical compounds in (16) are RCs for Grimshaw.

(16) **bee** sting (vs. *bee-stinging), **dog** bite (vs. *dog-biting)

2.1.3 Deverbal compounds between ASNs and RNs: Borer (2013)

In spite of her extensive study on ASNs, Grimshaw does not go to great lengths to compare DCs with ASNs in terms of morphosyntactic properties such as those in Table 1. Di Sciullo (1992) investigates some of these tests in further support of the similarity between DC heads and ASNs. However, two decades later, Borer (2013) challenges Grimshaw's analysis of DCs by using some of these morphosyntactic tests. She argues that the behavior of DCs essentially differs from that of ASNs, and proposes that all DCs are headed by RNs.

We retain three of Borer's arguments. First, she argues that, unlike ASNs, DCs disallow aspectual *in*/*for*-adverbials and, second, that they also disallow argumental *by*-phrases. This contrast is illustrated in (17) (cf. (11) and (10)). In Borer's

system, the unavailability of aspectual modifiers indicates that event structure (with arguments) is entirely missing from DCs, so they cannot involve ASNs. Her conclusion is that DCs are headed by RNs and behave just like RCs.

(17) a. the demolition of the house **by the army** *in 2 hours* (ASN)
b. the stabbing of the emperor **by Brutus** *for 10 minutes* (ASN)
c. the house demolition (***by the army**) (**in 2 hours*) (DC)
d. the emperor stabbing (***by Brutus**) (**for 10 minutes*) (DC)

Third, Borer claims that the object reading of non-heads in DCs is just as available as a subject reading, depending on context. As evidence, she quotes DCs as in (18), parallel to those in (1b), whose non-heads may correspond to subjects.

(18) teacher recommendation, court investigation, government decision

Some criticism and re-interpretation of Borer's facts is found in Iordăchioaia et al. (2017) and Iordăchioaia (To appear). We briefly note here that aspectual adverbials are barely ever attested in corpora even with ASNs (Lieber 2016: 39–42), so an extensive empirical study is necessary to determine how much DCs differ from ASNs in this respect. Furthermore, *by*-phrases are broadly attested with DCs in corpora, as Grimshaw's (15b) also predicts, but they usually involve bare plurals and not definite noun phrases or proper names as in Borer's (17c-d). Given that DCs are often generic, this restriction is natural.

Having summarized these two theoretical approaches to DCs, we may add that we do not aim to argue for one or the other. Instead, we use morphosyntactic properties whose pertinence for ASN-hood is accepted by both to guide us in evaluating the impact of the head noun on the interpretation of the DC. Our hypothesis that a high level of ASN-hood in DC heads correlates with an object reading of the non-heads, however, follows Grimshaw's intuition that 'true' DCs involve ASN heads and are fully compositional. By contrast, Borer's claim is that DCs are always ambiguous like RCs and never as compositional as ASNs. Given that our results support the correlation between ASN-properties and an object reading in DCs, they also bring some evidence against Borer's analysis.

2.2 Computational approaches to compounds

Compounds have been the focus of quite a number of papers in the field of Computational Linguistics (CL) and NLP. In view of the topic of this paper there are two strands of research that are most relevant. The first focuses on determining the relation between the two components of a compound, the head and

the non-head. For our study this work is relevant to the extent that it discusses compounds whose head is a deverbal noun. The second strand of research is concerned with modeling the lexico-semantic transparency of noun-noun compounds. We will start by discussing the former and finish with an overview of the work that predicts the degree of transparency in compounds.

2.2.1 Predicting the interpretation of deverbal compounds

The goal of computational work on deverbal compounds (referred to as nominalizations) has been to predict the relation between the non-head and the deverbal head. The relation inventory has varied from two classes, OBJ and SUBJ, in Lapata (2002), to three classes, OBJ, SUBJ and prepositional complement in Nicholson & Baldwin (2006), and to 13 classes – OBJ, SUBJ and further specifications of the prepositional complement in Grover et al. (2005).

These works have mostly focused on encyclopedic, usage-based features such as the syntactic relations attested between the base verb of the head noun and the non-head in large corpora. The underlying assumption is that the frequency distribution of syntactic relations between a given noun and a verb, for example, between *taxi* and *drive*, is a good estimate for the distribution of the underlying relation between *taxi* and *driver*. Additional pragmatic knowledge is obtained from the direct context of the compound. In selecting these pragmatic features, these works are in line with lexicalist theoretical approaches that list several covert semantic relations typically available in compounds (cf. most notably, Levi (1978), see Fokkens (2007), for a critical overview). In addition to these pragmatic features, some straightforward morphological features are selected, such as the suffix of non-heads ending in *-ee* and *-er* (Lapata 2002).

Our study differs from these works in several ways. First, our aim is not to reach state-of-the-art performance in prediction, but to test linguistic hypotheses by measuring the predictive power of the various features discussed in theoretical linguistics, which are also indicative of the compositionality of the compound.

Second, and related to the previous point, our features are all head-specific. This is because, following Grimshaw's theory, the behavior of the derived nominal heads (as ASNs or RNs) should mirror the structural correlation between DCs and the compositional structure of the original verb. The presence (or absence) of such a correlation is expected to have a great impact on the relation between the head and the non-head. In order to measure the individual impact of these theoretically-defined features, we do not rely on pragmatic features that involve both the head and the non-head as in the studies above.

Lastly, because our goal is to uncover in how far the behavior of the derived nominals (as ASNs or RNs) can predict the relation between head and non-head, we carefully selected equal numbers of DCs with the suffixes *-al*, *-ance*, *-ing*, *-ion*, and *-ment*. These suffixes are all ambiguous in their formation of ASNs and RNs, so we eliminate any bias for particular readings (cf. *-ee* and *-er*, §3.2).

2.2.2 Predicting the degree of transparency in noun-noun compounds

For the transparency of compounds two types of CL work are relevant, which focus on different tasks, but share the same assumptions. One type aims to predict the meaning of compounds based on composite functions between the vector-based representations of their parts, e.g., Ó Séaghdha (2008) and Mitchell & Lapata (2010). These works compare different types of mathematical functions for the combination of the vectors for heads and non-heads to best represent the meaning of compounds. In the same spirit, but closer to our interest in the syntactic-semantic relationship between the parts, Marelli & Baroni (2015); Baroni & Zamparelli (2010) investigate linguistically-informed composite functions.

The other line of work aims to predict the degree of lexico-semantic transparency (i.e., what they call 'compositionality'; cf. §1.3) of compounds . For this, they compare the vector-based representations of the parts and composite functions to the vector-based representations of the compound as a whole, e.g., Schulte im Walde, Hätty & Bott (2016); Reddy et al. (2011).

This second line of work also draws upon psycholinguistic insights, such as Libben et al. (1997; 2003), which groups noun-noun compounds into four different categories, depending on the transparency of the head and the non-head. The four classes are: TT for compounds with both a transparent head and non-head, OO for compounds with opaque heads and non-heads, and OT and TO for compounds whose parts differ along the dimension of transparency. They found that both semantically opaque and semantically transparent compounds show morphological constituency. However, they found the semantic transparency of the head to play a significant role. This confirms previous results from the psycholinguistic literature (Zwitserlood (1994)).

In this literature, several datasets have been created, which collect human ratings on the degrees of lexico-semantic transparency of compounds with respect to their constituents: e.g., in English (Reddy et al. 2011; Juhasz et al. 2015) and in German (Schulte im Walde, Hätty, Bott & Khvtisavrishvili 2016). Schulte im Walde, Hätty, Bott & Khvtisavrishvili (2016) have enriched the semantic transparency ratings with several empirical features related to the constituents of

the compound in order to measure the influence of these features on the transparency of the compound. These features include:

- Corpus frequencies of the compounds and their parts;
- Productivity of the parts, as in the number of compound types the part (head/non-head) appears in;
- Number of senses for the parts as retrieved from GermaNet (Hamp & Feldweg 1997; Henrich & Hinrichs 2010) for the German dataset and WordNet (Fellbaum 1998) for the English dataset.

Schulte im Walde, Hätty & Bott (2016) use vector space models to model the meaning of the compounds and their parts. Subsequently, they model the transparency of the compound by measuring the distance between the composite vector of its parts and the vector for the actual compound. The assumption behind this work is that the vectors of transparent compounds should be closer to the composite function of their vectors than the vectors of opaque compounds.

The main question Schulte im Walde, Hätty & Bott (2016) try to answer is whether the above-mentioned properties (frequency of the compound and its parts, productivity, and ambiguity of its parts) play a major role in the quality of the predictions. They found that for the head all properties had a significant effect on the predictions, whereas for the modifier the effect was not consistent. This converges with our results in predicting the compositionality of DCs from the properties of the head.

Furthermore, they attribute the influence of these features to the underlying ambiguity that they seem to be correlated with: e.g., frequent heads that are highly productive are often highly ambiguous. We note, however, that these studies are not concerned with DCs, as ours, but especially with what we call RCs, some of which are lexico-semantically less transparent than our DCs (cf. *hogwash*).

3 Methodology

In this section we present the corpus and the tools for automatic pre-processing, the procedure in the DC extraction, as well as the annotation and post-processing of our collection of DCs.

3.1 Corpus and tools

For the selection of DCs and to gather corpus statistics on them, we exploited the Annotated Gigaword corpus (Napoles et al. 2012), one of the largest general-domain English corpora, which contains several layers of linguistic annotation. This corpus encompasses ten million documents from seven news sources and more than four billion words. We made use of the following available automatic preprocessing steps and annotations, which we accessed via the Java API provided along with the corpus: sentence segmentation (Gillick 2009), tokenization, lemmatization and POS tags (Stanford's CoreNLP toolkit[6]), and constituency parses (Huang et al. 2010) converted to syntactic dependency trees with Stanford's CoreNLP toolkit. The POS tags adhere to the Penn Treebank tagset (Santorini 1990); the dependency relations follow the Stanford typed dependencies (de Marneffe & Manning 2008). As news outlets often repeat news items in subsequent news streams, the corpus contains a considerable amount of duplication. To improve the reliability of our corpus counts, we removed exact duplicate sentences within each of the 1010 corpus files, reducing the corpus size by 16%.

3.2 Extraction of deverbal compounds

We created a balanced collection of DCs, which we extracted from the Gigaword corpus. We first gathered 25 nouns (over three frequency bands : high, medium, low) for each of the suffixes *-al*, *-ance*, *-ion*, *-ing*, and *-ment*. The highest frequency band ranges from 4.5 to 3.5 on the Zipf-scale (van Heuven et al. 2014), the medium frequency band ranges from 3 to 2.5, and the lowest one from 2 to 1.5. The suffixes may form both ASNs and RNs according to Grimshaw (1990).

We did not consider zero-derived nouns like *attack*, *abuse*, *bite*, because Grimshaw considers them RNs (see (16)). We also excluded deverbal nouns based on the suffixes *-er* and *-ee*, as they denote event participants corresponding to the subject and the object of the base verb, respectively, implicitly blocking this interpretation on the non-head (cf. $police_{subj}$ *train**ee*** – dog_{obj} *train**er***). In our attempt to capture the closeness of DCs to ASNs (and the base verbs), we considered only the suffixes that build eventive nominals, which could realize both a subject and an object argument. DCs headed by *-ee* and *-er* nouns would have been biased for one or the other. However, our selection of suffixes represents the large majority of deverbal nouns. They make up for 69.4% of the total number of deverbal nouns in the NOMLEX database (Macleod et al. 1998), which consists of 1025 lexicalized deverbal nouns.

[6]http://nlp.stanford.edu/software/corenlp.shtml

The nouns were selected such that their base verbs present transitive uses, making both subjects and objects available.[7] For illustration, Table 2 offers samples of deverbal nouns per each frequency range and suffix. For each such selected noun we then extracted the 25 most frequent compounds that they appeared as heads of, where available. A few deverbal nouns (in particular those with suffixes *-al* and *-ance*) were less productive in compounds and appeared with fewer than 25 different non-heads. Given these gaps and after removing a few repetitions due to capitalization, we obtained a collection of 3111 DCs.

Table 2: Samples of extracted deverbal nouns

Frequency	AL	ANCE	(AT)ION	ING	MENT
High	approval withdrawal rental	performance assistance surveillance	protection reduction consumption	building training trafficking	development movement punishment
Medium	renewal survival upheaval	assurance dominance tolerance	supervision cultivation instruction	killing counseling teaching	deployment placement adjustment
Low	retrieval disapproval dispersal	defiance endurance ignorance	demolition expulsion deportation	weighting chasing mongering	reinforcement empowerment abandonment

3.3 Annotation and post-processing of DCs

3.3.1 Interpretation of (non-heads in) DCs

All DCs were annotated by three trained American English speakers, who had a university level background in linguistics. They had to label the DCs as OBJ(ect), SUBJ(ect), OTHER, or ERROR, depending on the syntactic relationship that they considered the DC to establish between the base verb of the head noun and the non-head. For instance, DCs such as in (1) would be labeled as OBJ (1a), SUBJ (1b), and OTHER (1c). OTHER was an umbrella label for prepositional objects (e.g., *adoption counseling* 'somebody counsels somebody *on adoption*'), various adjuncts (e.g., *ultrasound examination* 'to examine somebody *with an ultrasound*', *sea burial* 'to bury somebody *by the sea*', *surprise arrival* 'somebody/something arrived *by surprise*'). ERROR was intended to identify errors of the POS tagger

[7] *Arrive* is the only intransitive unaccusative verb that we used, which realizes the object/internal argument as a subject.

(e.g., *face abandonment* originates in 'they face$_V$ abandonment'), but was also employed by the annotators when they considered the DC uninterpretable or ungrammatical. We allowed the annotators to use multiple labels and to indicate ambiguity (using '–') and the preferred order of the readings (using '>').

We used the original annotations to create a final list of compounds with the labels that all three annotators agreed on. For ambiguously labeled DCs we selected the one reading available for all three. If they all indicated the same ambiguity for a DC, we labeled the DC as ambiguous. The labels we used for the final dataset are OBJ, SUBJ, OTHER, DIS(agreement between annotators), AMBIG(uous), and ERROR. In spite of Borer's (2013) claims, we found only two cases of ambiguity which all three annotators agreed on – namely, *police killing* and *doctor referral*, which were both labeled SUBJ–OBJ. In the end we identified 772 DIS, 1377 OBJ, 404 OTHER, 286 SUBJ, and 273 ERROR cases of DCs. After removing the disagreements, the two ambiguous DCs and the errors, we obtained 2067 DCs. We based our study on the agreed-upon relations only. We note, however, that the simple inter-annotator agreement (IAA) among the three annotators, excluding the errors, was 72.8%. In a previous study with only two annotators (Iordăchioaia et al. 2016), the IAA was 81.5%.

We kept two versions of the data: one in which the classes OTHER and SUBJ are separate and one in which we conflated them to NOBJ (non-object). Given the purpose of this paper, i.e., verifying to what extent the OBJ reading of a DC correlates with particular morphosyntactic properties of the head noun, we focus here on the binary classification. The resulting data set is skewed with OBJ prevailing: 1377 OBJ and 690 NOBJ.

3.3.2 Process vs. result readings in DCs

An additional annotation task concerned feature 7. *process-vs-result* from Table 3 in §4.1 . This feature was designed to capture the three annotators' judgments with respect to how close the interpretation of the DC comes to the ASN and the verbal expression of a process/event in which the non-head is realized as SUBJ, OBJ, or OTHER. They had to rate DCs from 5 (very prominent process) to 1 (no process = result) (see Grimshaw (1990)).

We first explained to them the difference between an ASN and an RN as follows: "*The teacher's* ***assignment*** *of tasks* expresses a process in which the teacher assigns tasks. However, in *this long* ***assignment*** *took several hours to complete*, the noun *assignment* is interpreted as a result of the process of assigning something – namely, the task itself." We then instructed the annotators to check this contrast in DCs like *task assignment* and *Math assignment* and rate the ones that

relate to the process as closer to 5 and those that relate to the result as closer to 1. Another example was *apartment building*, which should be rated as closer to 5, if they interpret it as 'to build apartments', and closer to 1, if they interpret it as 'a building with apartments'. We fully encouraged the annotators to employ the scores 4, 3, 2 for unclear cases.

During this task, the annotators had access to their previous SUBJ/OBJ/OTHER annotation labels for each DC and could compare different DCs headed by the same head noun. In terms of the variation of ratings between DCs headed by the same noun, one annotator in particular assigned pretty similar scores, although the contrast was clear. This annotator also showed a tendency towards the extremes: either 5 or 1. In general, the task was perceived as difficult, especially by this annotator. We multiplied the scores from 5 to 1 by 20 to use them as percentages. For each DC we calculated the average between the three annotations obtaining values between 20 and 100.

4 The selection of morphosyntactic features

4.1 Theoretical considerations

To collect information on the morphosyntactic properties of the head nouns in DCs, we defined a total of nine morphosyntactic features, given in Table 3.

The first seven features are inspired by Grimshaw (1990), although only the first four directly correspond to the properties in §2.1.1. Two adjustments led us to four features instead of the six properties in Table 1: first, *in*/*for*-adverbials were discarded, because we found close to no relevant data; second, we counted agent-oriented and aspectual adjectives together, as they were also very few.[8] In line with our hypothesis, we expect all these seven features to have predictive power and to point to an OBJ interpretation of the DCs.

Feature 1. *of_outside_DC* encodes the first property in Table 1. Here we counted the percentage of occurrences of a (singular) head noun in which it also realizes an *of*-phrase. For feature 2. *by_outside_DC* (i.e., third property in Table 1), we collected the frequency of a *by*-phrase with a head noun. Feature 3. *sum_adjectives* collects all the (singular form) occurrences of the head nouns in a modifier relation with agent-oriented or aspectual adjectives (cf. second and fifth prop-

[8] We initially collected data on *in*- and *for*-adverbials, but only a few nouns had such occurrences. At closer inspection even these examples turned out not to illustrate *in*- and *for*-phrases that modify the telic/atelic aspect of the head noun, as Grimshaw and Borer used them. Instead, they mostly functioned as temporal modifiers, and we therefore discarded this feature.

Table 3: Indicative features for head nouns

Feature label	Description and illustration
1. *of_outside_DC* (Grimshaw 1990)	Percentage of the head's occurrences as singular outside compounds which realize a syntactic relation with an *of*-phrase. E.g., *assignment **of** problems*
2. *by_outside_DC* (Grimshaw 1990)	Percentage of the head's occurrences in the singular outside compounds which realize a syntactic relation with a *by*-phrase. E.g., *assignment (of problems) **by** teachers*
3. *sum_adjectives* (Grimshaw 1990)	Percentage of the head's occurrences in a modifier relation with one of the adjectives *frequent*, *constant*, *intentional*, *deliberate*, or *careful*.
4. *sg_outside_DC* (Grimshaw 1990)	Percentage of the head's occurrences as singular outside compounds.
5. *by_inside_DC* (≈ 2. *by_outside_DC*)	Percentage of the head's occurrences as singular inside compounds which realize a syntactic relation with a *by*-phrase. E.g., *task assignment **by** teachers*
6. *sg_inside_DC* (≈ 4. *sg_outside_DC*)	Percentage of the head's occurrences as singular inside compounds.
7. *process-vs-result* (≈ ASN vs. RN)	Native speaker annotation of each DC as a process (*car driving*) or result (*apartment building*) on a scale from 5 to 1.
8. *suffix* NEW	Suffix of the head noun: AL (rent**al**), ANCE (insur**ance**), ING (kill**ing**), ION (destruct**ion**), MENT (treat**ment**)
9. *head_in_DC* NEW	Percentage of the head's occurrences within a compound out of its total occurrences in the corpus.

erty in Table 1).[9] Feature 4. *sg_outside_DC* measures the percentage of singular occurrences of the head noun out of its total occurrences in the corpus (cf. last property in Table 1).

Grimshaw's properties in Table 1 characterize deverbal nouns as ASNs when they appear on their own, i.e., *outside* compounds. This is why features 1. to 4. are labeled correspondingly. Yet, if DCs are supposed to resemble ASNs, we considered that their head nouns should preserve these properties also within DCs, i.e., when the head noun is *inside* a DC.[10] For this reason, we also introduced the features 5. *sg_inside_DC* and 6. *by_inside_DC*. The former measures the percentage of singular DCs out of their total occurrences, and the latter the percentage of DCs that realize a *by*-phrase. We did not test *of*-phrases inside DCs, since DCs usually realize the object as a non-head (see our annotation results in §3.3.1) and collecting such occurrences would have mostly delivered noise. The adjectives modifying DCs were also left out, because their number was close to inexistent.

There are two caveats to these corpus-based features inspired by Grimshaw (1990). First, as we noted in §2.1.1, the individual ASN-properties are not fully reliable in determining ASN-hood: e.g., there is ambiguity in argument marking (i.e., *of*- and *by*-phrases), and deverbal nouns are easily coerced between the readings. For this reason, Grimshaw used several such properties together in her examples. However, we extracted these data from corpora, and most of the attestations were too few to allow any combined patterns beyond the one we ensured – that of a singular form of the head noun in each of the other properties. Second, and related to this, basing our study on a corpus comes with the risk that, no matter how large the corpus, it may not present enough relevant data. It was for these two reasons that we considered adding three more head-related features to our study. We first gathered native-speaker intuitions about the ASN vs. RN status of the head nouns in DCs (see feature 7. *process-vs-result*) and supplemented Grimshaw's tests with information about the suffix and the frequency of the head noun within compounds (features 8. *suffix* and 9. *head_in_DC*).

We designed feature 7. *process-vs-result* (P-R) in order to grasp Grimshaw's intuition about the contrast between ASNs and RNs by means of introspection. The process vs. result interpretation is the fundamental difference between ASNs and RNs in Grimshaw's understanding. It can be seen as the latent variable that her morphosyntactic properties are intended to identify: ASNs express processes or events like the corresponding verbs, while RNs depart from this meaning and

[9]Note that, given Grimshaw's assumption that ASNs do not appear in the plural, we counted all of these occurrences in the singular form of the head noun.

[10]Di Sciullo (1992) and Borer (2013) actually apply the same reasoning.

express results. Following this annotation (see §3.3.2), we gathered information on how salient the verbal process is in the meaning of a DC and, indirectly, how accessible the compositional structure of the base VP is within the DC.[11]

The last two features 8. *suffix* and 9. *head_in_DC* represent two further properties of the head nouns that we considered interesting for our study. The theoretical literature does not offer much on suffixes. *-Ing* has received most attention, to the extent that Grimshaw argued that it always forms ASNs, while Borer claims that it encodes what she calls an originator (i.e., subject argument), with the effect that in compounds, the SUBJ reading is blocked for non-heads and OBj is favored. Neither contention is true. First, *-ing* presents several examples of RNs (see *building(s), writing(s), reading(s)*). Second, we *do* find SUBJ-DCs headed by *ing*-nouns (see (1b)). In general, the information on the suffix is independent of ASN-hood, since all suffixes allow both ASN and RN readings, but we aimed to check whether some suffixes may be more informative than others.

Feature 9. *head_in_DC* delivers us the degree of compoundhood of a deverbal noun, i.e., how likely it is to appear within a compound. The expectation is that a noun that typically appears in compounds has undergone some meaning specialization, which requires another noun to be instantiated. One may rightly say that this makes the meaning of such head nouns less transparent than for those that freely appear both within and outside compounds. However, for deverbal nouns, to the extent that this slight meaning specialization requires a particular type of non-head, it can give us useful information about which (morpho)syntactic relationship between the base verb and one of its arguments is most likely to form a DC. If it is a non-OBJ relation, this shows that compositionality as in (6) is not a typical condition in the formation of DCs, weakening the relevance of our investigation. However, our results in Table 7 below indicate that high compoundhood correlates with an OBJ interpretation of the non-head, which supports the relevance of compositionality in the formation of DCs.

4.2 Technical support

To obtain statistics for the morphosyntactic features, we extracted counts for the selected DCs and their head nouns from the Gigaword corpus by matching patterns defined over word forms, lemmas, POS tags and dependency relations,

[11] The way we gathered estimates for our P-R feature comes close to the NLP studies which gather native speaker evaluations about the transparency of compounds. Namely, our three annotators had to evaluate how close the morphosyntactic (and semantic) relationship between the head noun and the non-head comes to the fully compositional relationship between the corresponding verb and its argument or adjunct.

as provided by the automatic corpus annotations. The specific patterns used for each feature are detailed in the following.

For the *inside_DC* features we extracted DCs from the Gigaword corpus by locating two adjacent nouns according to the POS tags NN for singular nouns and NNS for plural nouns, and excluding noun pairs directly preceded or succeeded by other nouns or proper nouns (POS tags NNP and NNPS). DCs were matched with the word form of the non-head and the lemma of the head, thereby extracting singular and plural occurrences. We determined the grammatical number of a noun or compound by its POS tag or the POS tag of its head, respectively. For example, we matched *security training(s)*, but not *airport security training* and *security training instructor*, to make sure that we do not extract parts of larger compounds. Conversely, the *outside_DC* features apply to head nouns (matched by their lemma and POS tag NN or NNS) without any noun or proper noun next to them.

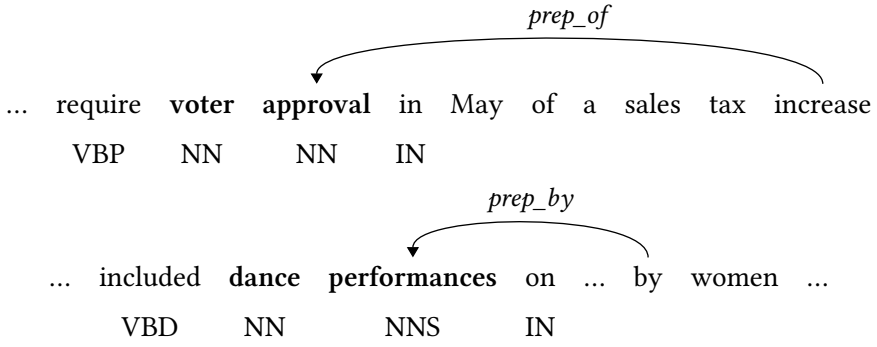


Figure 1: Illustration of morphosyntactic patterns to extract DCs heading *of-phrases* (top) and *by-phrases* (bottom)

We counted a DC (or its head noun) as being in a syntactic relation with an *of*-phrase or *by*-phrase, if it (or its head) governed a collapsed dependency labeled 'prep_of'/'prep_by'[12], as in Figure 1. Since we were interested in prepositional phrases that realize internal or external arguments, but not in temporal phrases (e.g., 'by Monday') or fixed expressions (e.g., 'of age', 'by chance'), we excluded phrases headed by words that typically appear in these undesired constructions. We semi-automatically compiled these lists based on a multiword

[12]By conflating dependencies involving prepositions or conjuncts, collapsed dependencies directly link content words. This simplifies the extraction patterns, as we can obtain the complement of the prepositional phrase depending on the noun or the DC, by following a single dependency arc.

expression lexicon[13] and manually added entries. To compute the feature *sum_adjectives* we counted how often each noun outside a DC governs a dependency relation labeled 'amod', where the dependent is an adjective (POS tag JJ) out of the lemmas *intentional, deliberate, careful, constant*, and *frequent*.

4.2.1 Reliability of the extracted features

Our extracted features rely on the automatic corpus annotations, the manually defined extraction patterns, and, in the case of the *of*-phrases and *by*-phrases, on heuristics, to exclude undesired matches of temporal phrases or fixed expressions. The constituency parser, which was used to obtain the syntactic analyses then converted to dependency trees, obtained an average F1-score of 91.4% on a standard test set, Section 22 of the Wall Street Journal corpus (Huang et al. 2010).

To measure the reliability of the extracted features, more in particular the most error-prone features based on heuristics, we exemplarily conducted a manual analysis of the counts of head nouns that appear in conjunction with *of*-phrases and *by*-phrases. For this, we implemented the following pattern to extract all candidate sentences in the corpus for this feature. We selected all sentences in which one of the target head nouns outside a compound was followed by a token with lemma *of* or *by* and POS tag IN, not separated by a punctuation mark.[14] On the one hand, this was driven by the motivation to keep the amount of sentences on a manageable level and focused on the feature of interest. On the other hand, we designed the pattern to maximize recall so as not to miss out on any true positives. We then randomly selected 2000 of these sentences for each preposition for a manual annotation of the target features by a single human annotator. A comparison of the annotated instances with the automatically extracted instances revealed a precision of 91.0% and recall of 90.1% for *of*-phrases, while the results for *by*-phrases were lower (85.0% precision, 73.8% recall).

5 Data exploration with machine learning techniques

Our goal is to test the features listed in Table 3 for their predictive power in determining the relation between the head and the non-head. These features are composed of numerical (1 to 7, and 9) and categorical features (8). The dependent variable is a binary feature that varies between one of the two annotation labels,

[13] http://www.cs.cmu.edu/~ark/LexSem/

[14] We used the following list of punctuation characters: '.', '?', '!', ';', ':', ','.

OBJ and NOBJ. We trained a Logistic Regression classifier to model the effect of these features.[15]

We divided the data described in §3.3.1 into a test and a training set. Because the features are all head-specific, as can be seen in Table 3, the model was tested on a test set for which we ensured that neither compounds, nor heads were seen in the training data. Therefore, we randomly selected two mid-frequency heads for each suffix and removed these from the training data to be put in the test data. We expect mid-frequency heads to lead to most reliable results, because high-frequency heads may show higher levels of idiosyncrasy and low-frequency heads may suffer from data sparseness.[16] This resulted in a division of roughly 90% training and 10% testing data.[17] The data set resulting from the annotation effort is skewed with OBJ being the majority class. Our selection of test instances introduces further differences in proportions of OBJ and NOBJ in the test and training set. Therefore, we balanced both the training and test set by randomly removing instances with the OBJ relation (the largest class) until both classes were equal in size.[18] The balanced training set consisted of 1248 examples, and the test set of 132 examples.

We compare our models with the random baseline, and two additional baselines to make sure that the features we are proposing are not just be a by-product of the impact of simpler variables. We computed the relative[19] frequency of the head and the relative family size, i.e., how many compound types we find with a given head.[20]

We ran ablation experiments to determine the individual contribution of each feature in addition to the other features. However, because features might be interdependent and one feature could overshadow another, we first looked at the performance of each feature individually. This way, we could measure the

[15] We used version 3.8 for Linux of the Weka toolkit (Hall et al. 2009) and experimented with several other classifiers that have interpretable models (decision trees, but also support vector machines and naive Bayes classifiers). All of these underperformed on our test set.

[16] We remind the reader that our goal is not to determine the realistic performance of our model, but to measure the contribution of the features. Therefore we believe that the bias introduced by selecting mid-frequency items for the test set is acceptable.

[17] Multiple divisions of training and test data would lead to more reliable results, but we have to leave this for future work.

[18] We also ran experiments with non-balanced data, because we reasoned that more data might result in higher performance, but the performance proved to be comparable. A balanced dataset facilitates comparisons to the random baseline of 50%.

[19] By providing relative counts, we make sure these features are on the same scale as our other features.

[20] These additional baselines were computed on a slightly different test and training set, due to the random process in balancing the data.

exact predictive power of each individual feature in comparison to the baselines. Lastly, we combined the top-*n* features from ablation experiments and individual feature experiments to see the overall predictive potential of the model.

The first row in Table 4 shows that, when using all features, the classifier significantly outperforms[21] the baselines with a large margin (78.8%). This proves that the combination of features driven by linguistic theory has strong predictive power.

Table 4: Percent accuracy for individual features. "†" indicates a statistically significant difference from the performance of all features. All results are statistically significant in comparison to the baselines.

Features	Accuracy (%)
All features	**78.8**
process-vs-result	76.5
suffix	72.0†
sg_outside_DC	68.9†
sg_inside_DC	68.9†
head_in_DC	66.7†
sum_adjectives	61.4†
of_outside_DC	59.8†
by_outside_DC	56.0†
by_inside_DC	54.5†
process-vs-result and *suffix* combined	78.0
Random baseline	50.0
Head frequency baseline	50.0
Head family size baseline	46.8

With respect to the upper bound, we cannot directly compare the numbers in Table 4 with the IAA reported in §3.3.1, because the data we use for testing and training includes only examples on which all annotators agree; neither can we use the 100% IAA on this selected test set as an upper bound. We expect the IAA for this high-agreement test set to lie between 100% and the 81.5% reported in §3.3.1 for the complete dataset and two annotators. The 78.8% we attain is not

[21]Significance numbers for these experiments, in which training and test data are fixed, were computed with a McNemar test with $p < .05$, as it makes relatively few type I errors (Dietterich 1998).

too far from the upper bound we can estimate from these IAA values.[22]

Furthermore, the results for the individual features in Table 4 show that each feature outperforms the baselines significantly. This means that each feature contributes significantly to the prediction of the relation. The 78.0% performance of the model that combines the top-2 features is comparable to the 78.8% of the model that includes all features. This means that although all features contribute to the quality of the prediction of the model individually, the best features overshadow the effect of the less well-performing features.

Table 5 shows the results from the ablation experiments. Only the removal of features *suffix, of_outside_DC*, and *P-R* result in a significant drop in performance, which means that their contribution in addition to the other features is particularly important. Their performance together is not significantly higher than that of all features (cf. 80.3% vs. 78.8%).

Table 5: Percent accuracy in ablation experiments. "†" indicates a statistically significant difference from the performance of all features.

Features	Accuracy (%)
All features	78.8
All features, except *sg_inside_DC*	**80.3**
All features, except *head_in_DC*	79.5
All features, except *sg_outside_DC*	78.8
All features, except *by_inside_DC*	78.8
All features, except *sum_adjectives*	78.8
All features, except *by_outside_DC*	75.0
All features, except *suffix*	73.5†
All features, except *of_outside_DC*	72.0†
All features, except *P-R*	72.0†
P-R, *of_outside_DC*, *suffix*, *by_outside_DC* combined	**80.3**

For the sake of comparison, Table 6 shows the results of a model using corpus-based features only, i.e., the data does not include the *P-R* feature that is based on human judgments. Like in Table 5, we see that the features *of_outside_DC* and *suffix* are particularly important also in this model, since their absence triggers a significant drop in performance. In this model, however, the contribution of

[22] A realistic upper bound for the test set could be determined by getting an independent annotator to annotate the items in the test set and measuring the agreement with the previous annotations. We leave this for future work.

the feature *by_outside_DC* also becomes significant, by contrast to the model in Table 5, which included the *P-R* feature.

Table 6: Ablation experiment with corpus-based features only (no P-R). "†" indicates a statistically significant difference from the performance of all corpus-based features.

Features	Accuracy (%)
All corpus-based features	72.0
All corpus-based features, except *sg_outside_DC*	72.0
All corpus-based features, except *sum_adjectives*	72.0
All corpus-based features, except *sg_inside_DC*	72.0
All corpus-based features, except *by_inside_DC*	72.0
All corpus-based features, except *head_in_DC*	68.2
All corpus-based features, except *suffix*	66.7†
All corpus-based features, except *by_outside_DC*	59.1†
All corpus-based features, except *of_outside_DC*	54.5†
of_outside_DC, *by_outside_DC*, and *suffix* combined	**72.7**

Table 7 shows the direction of the prediction of the features in all three models (Tables 4 to 6). In other words, it shows whether higher values of a given feature are indicating higher chances of an OBJ or NOBJ relation. We gathered these directions by inspecting the coefficients of the logistic regression model.[23]

6 Discussion

In what follows we offer a detailed discussion of our results and interpret them in view of our initial hypothesis (§6.1). We then show their implications for compositionality and for our starting hypothesis (§6.2). In the end we present the main comparison points with respect to previous NLP literature (§6.3).

[23]We inspected the weights in the models as well, but they are not very informative, because there is a high level of collinearity in the features and the weights are calculated based on all other features staying equal. For this reason we report results on single feature models and ablation tests instead.

Table 7: Direction of prediction per feature in different models. Consistent values across studies in bold

Feature	Table 4	Table 5	Table 6
P-R	**OBJ**	**OBJ**	N/A
suffix=ment	OBJ	OBJ	OBJ
suffix=ance	OBJ	NOBJ	NOBJ
suffix=ion	NOBJ	OBJ	OBJ
suffix=al	OBJ	NOBJ	OBJ
suffix=ing	NOBJ	OBJ	NOBJ
sg_inside_DC	NOBJ	OBJ	NOBJ
by_inside_DC	OBJ	NOBJ	NOBJ
sg_outside_DC	**OBJ**	**OBJ**	**OBJ**
head_in_DC	**OBJ**	**OBJ**	**OBJ**
sum-adjectives	NOBJ	OBJ	OBJ
of_outside_DC	**OBJ**	**OBJ**	**OBJ**
by_outside_DC	NOBJ	NOBJ	NOBJ

6.1 Interpretation of results

6.1.1 *Process-vs-result* (*P-R*)

According to Table 4, the best individual feature is the *process vs. result* reading of the DC with 76.5% accuracy. The accuracy resulting from the combined model with all features (78.8%) is not significantly higher (McNemar two-tailed p-value of 0.2482), showing that this single feature is indeed very strong, and stronger than any of the corpus-based features on their own or in combination (cf. Table 6). This is not surprising, given that this feature encodes direct estimates for the ASN-hood of the head based on introspection.[24] In the ablation experiment in Table 5, *P-R* also proves to be very strong, since its removal yields a significantly lower result (72.0% vs. 78.8%), the lowest in this experiment. Still, the ablation study shows that removing *of-outside* is as detrimental to the model as removing *P-R*. This indicates that these two features capture characteristics that complement the rest of the corpus-based features to a similar extent.

Importantly, in line with our hypothesis, an increase in the *P-R* value correlates with an OBJ interpretation of the compounds in both experiments (see Table 7).

[24]It is interesting to see though that manual annotation was better at predicting ASN-hood than any of the corpus-based features, in spite of the huge corpus we used. This suggests that we need even larger corpora to make up for the performance of (expensive) manual annotation.

To be precise, the *P-R* feature is so designed that a high value indicates that the DC is headed by an ASN, which parallels the verbal construction in (6). Given that such a compositional structure requires the object to be realized first, the fact that a high *P-R* value correlates with an OBJ reading of the DC in our models confirms our hypothesis that compositional DCs involve object non-heads.

The two columns in Table 8 illustrate pairs of DCs which, despite having the same head, reveal contrasting *P-R* values. In these examples, one can see that whenever the DC pair differs between an OBJ and a NOBJ reading, the OBJ reading receives the higher *P-R* value. This is predicted by our hypothesis and also supported by the results in Table 7. However, we also find examples with two considerably different *P-R* values under the same OBJ (or NOBJ) interpretation, which shows that there is no one-to-one correspondence between a (high) process reading and an OBJ interpretation of the DC.[25]

Table 8: DC pairs with contrasting *P-R* values

High *P-R* > 60%			Low *P-R* < 60%		
DC	P-R (%)	Reading	DC	P-R (%)	Reading
home building	100	OBJ	police building	20	NOBJ
book reading	100	OBJ	temperature reading	40	OBJ
ship breaking	93.3	OBJ	record breaking	40	OBJ
science teaching	93.3	OBJ	church teaching	46.7	NOBJ
career counseling	93.3	NOBJ	telephone counseling	53.3	NOBJ
slum clearance	80	OBJ	safety clearance	20	NOBJ
body movement	80	OBJ	student movement	33.3	NOBJ
nicotine withdrawal	80	NOBJ	summer withdrawal	33.3	NOBJ
refuse disposal	80	OBJ	garbage disposal	46.7	OBJ
temperature tolerance	73.3	OBJ	alcohol tolerance	20	OBJ
cancer treatment	73.3	OBJ	spa treatment	46.7	NOBJ

The confusion matrix for the feature *P-R* in Table 9 confirms that the machine learning algorithm was not able to find a clear cut-off value for this feature above which we find only OBJ readings. The *P-R* feature misclassifies 18 OBJ-DCs as NOBJ, and 13 NOBJ-DCs as OBJ. Examples of the former case are the OBJ-DCs in the second column of Table 8, which have a low *P-R* value, because they involve RN heads (see *temperature reading*, *alcohol tolerance*). In the latter case, the errors concern the NOBJ-DCs from the first column of Table 8, which have a high *P-R* value (see *career counseling* and *nicotine withdrawal*).

[25]NOBJ-DCs with a high *P-R* value are usually headed by simple event nominals like the nouns

Table 9: Confusion matrix for *P-R*

		Classified as		
		OBJ	NOBJ	Totals
Gold	OBJ	48	18	66
	NOBJ	13	53	66
	Totals	61	71	132

In our study, the *P-R* annotation feature comes closest to the transparency rating of compounds carried out in some NLP studies (cf. §2.2). The difference is that we correlated the rating with the semantics of the base verb in combination with its argument or adjunct, following Grimshaw's (1990) insight. At the same time, our design primarily targeted compositionality.

6.1.2 *Of_outside_DC*

The next most important feature in our endeavor to capture compositionality in DCs is the realization of an *of*-phrase by the deverbal noun. This feature is intended to measure how often the deverbal noun realizes an *of*-phrase introducing the object argument, when appearing outside DCs. As explained in §2.1.1, this is the necessary and sufficient condition to diagnose an ASN in Grimshaw's theory. If the head noun of a DC shows a high tendency to realize *of*-phrases introducing objects, we expect it to also require object non-heads in DCs.

Although on its own the feature *of_outside_DC* yields a value of only 59.8% (see Table 4, insignificantly lower than the next higher value of 61.4%), the ablation study in Table 5 shows that its removal is just as detrimental for the system as the removal of the *P-R* feature: The accuracy drops from 78.8% to 72.0%. Similarly, in the model with corpus-based features in Table 6, its removal triggers the largest drop, showing that in combination with the other features, the contribution of *of_outside_DC* is very important. This confirms Grimshaw's claim that the realization of the object argument is essential in identifying ASNs. Even more important for our hypothesis is the fact that *of_outside_DC* systematically correlates with an OBJ-DC in all our models (see Table 7). That is, to the extent that this feature identifies DCs with ASN heads, a high value indicates an object reading for the DC, as expected under our hypothesis.

The question is why the *of_outside_DC* feature does not score better than 59.8% on its own. First, as shown in §2.1.1, the presence of an *of*-phrase per se, as extracted from the corpus, is no guarantee for ASN-hood, since *of*-phrases may

in (11c, d).

introduce possessive modifiers of RNs, besides the object arguments of ASNs. Second, even in their ASN reading, deverbal nouns attested in corpora do not always realize their object arguments (cf. Grimm & McNally (2013)).

Table 10: Head nouns with (in)frequent *of*-phrases. Outliers in bold

Head noun	*Of_outside_DC* (%)	OBJ-reading (%)
creation	80.5	72.7
avoidance	70.4	100
obstruction	65.3	90.5
assassination	52.3	**11.8**
preservation	52.1	100
proposal	1	**76.2**
counseling	0.5	10
mongering	0	**100**

The samples in Table 10 show various mismatches between the realization of *of*-phrases and the formation of OBJ-DCs. For instance, *avoidance* and *preservation*, which build only OBJ-DCs in our database, have fewer occurrences with an *of*-phrase than *creation*, which forms only 72.7% OBJ-DCs. Moreover, *proposal*, which forms a high proportion of OBJ-DCs, realizes *of*-phrases in only 1.0% of its occurrences. In spite of the many OBJ-DCs like *book/contract/marriage/investment proposal*, the verbal relation is lost in this noun. It mostly functions as an RN, i.e., it refers to the proposal made, and not to the process/event of proposing. In confirmation of this, these DCs received a *P-R* rating as low as 20% to 26.7%. This is an example of how our individual features complement each other.

The confusion matrix for the feature *of_outside_DC* in Table 11 shows indeed that the model based on this feature makes many false predictions, notably, it attributes 38 OBJ readings to DCs that in fact have a NOBJ reading. This means that the prediction power of *of_outside_DC* is misled by the presence of *of*-phrases with head nouns that form NOBJ-DCs (see Table 10). These DCs involve RN heads, which realize *of*-phrases as modifiers and not object arguments. The head noun *assassination* in Table 10 is one example. That this noun behaves like a RN is confirmed by the *P-R* value of the DCs it forms, which is below the average of 60%. A similar problem is posed by the DCs headed by, e.g., *creation*, which also allows RN readings and forms NOBJ-DCs, in spite of the high frequency with *of*-phrases (Table 10). In these critical cases, the results in Tables 5 and 6 show

that the other corpus-based features compensate for the errors made by the *of_outside_DC* feature, helping the model.

Table 11: Confusion matrix for *of_outside_DC*

		Classified as		
		OBJ	NOBJ	Totals
Gold	OBJ	51	15	66
	NOBJ	38	28	66
	Totals	89	43	132

All in all, when comparing *of_outside_DC* with *P-R* in the ablation study, their contribution in combination with the other corpus features is similar. The difference is that the other features negatively affect the 76.5% individual contribution of *P-R* (cf. 72%), while they substantially improve the 59.8% contribution of *of_outside_DC* (cf. Table 4). Thus, the contribution *of_outside_DC* greatly relies on the other ASN-features in the ablation models in Tables 5 and 6. This is not surprising, given the ambiguity of *of*-phrases, a reason for which Grimshaw (1990) used this test in combination with others (see §2.1.1). The contrast between *P-R* and *of_outside_DC* is also expected, since *P-R* is manually annotated and targets the underlying ASN-hood of the deverbal noun; the corpus features can only capture some aspects of it.

6.1.3 *Suffix*

Suffix is an important feature in all our models (see Tables 4, 5, and 6). It is the strongest corpus-based feature, as we can see from the performance of the individual features in Table 4, and has additional predictive power compared to the combination of all features (see Tables 5 and 6). However, Table 7 demonstrates a high variance in the direction of prediction of each suffix. Except for *-ment*, which correlates with OBJ readings, none of them is constant across models.

As noted in §4.1, the theoretical literature does not offer much on the role of suffixes in the ASN vs. RN disambiguation of deverbal nouns. Grimshaw (1990) and Borer (2013) suggest that *-ing* should form ASNs, which is disconfirmed by some data and by our models, where *-ing* oscillates between OBJ- and NOBJ-DCs. It is difficult to draw any conclusions on the role of the *suffix* feature for our compositionality hypothesis for two reasons. First, more theoretical research must be pursued to draw some definite conclusions on possible correlations between suffixes and ASN-hood, since the one suffix that was expected to show a preference did not. Second, we must also consider that the dataset of DCs for each suffix was five times smaller than for the other features in our study: i.e., the feature *suffix*

subsumes five different suffix features. The small dataset may also be a reason for the inconclusiveness of the results in Table 7.[26]

The high variation between OBJ and NOBJ readings in Table 7 indicates that the valuable contribution of the *suffix* feature in the prediction task (72.0% in Table 4) comes from the complementarity between the individual suffixes. Similarly, in the ablation models in Tables 5 and 6, the contribution of the suffixes – which, recall, is independent of Grimshaw's tests – is complementary to the features that diagnose ASN-hood. Thus, the *suffix* feature is not informative about the relation between compositionality and interpretation in DCs, but improves the predictive power of the models.

6.1.4 *Sg_outside_DC* and *sg_inside_DC*

The frequency of the noun head in a singular form whether outside or inside a DC yields similar accuracy levels (68.9% in Table 4, 78.8% and 80.3% without a significant difference in Table 5, and 72% in Table 6). This similarity supports our assumption that within DCs the head nouns should preserve the properties from outside DCs (see §4.1). However, an interesting difference appears with respect to the direction of prediction, since only *sg_outside_DC* constantly predicts OBJ-DCs across all the models in Table 7, while *sg_outside_DC* is less reliable. This suggests that Grimshaw's morphosyntactic ASN-properties may be more reliable when the deverbal noun appears outside a DC than inside DCs.[27]

6.1.5 *Head_in_DC* (Compoundhood)

As an individual feature, the accuracy of *head_in_DC* is just above average among the other features in the present study (see Table 4). Its removal in our ablation experiments yields slight and non-significant drops in accuracy. In §4.1, we conjectured that an OBJ reading of DCs whose head nouns present high compoundhood would show us that a compositional construction with an object non-head is very likely to form DCs. The direction of prediction in Table 7 indicates that high values of this feature consistently correlate with OBJ-DCs, supporting this assumption. However, why does this feature not perform better? Our full database shows that its values are not informative enough: there are a few head

[26] To check correlations between individual suffixes and ASN-hood, one could measure how the *suffix* feature fares with respect to the *P-R* value and not the OBJ-NOBJ readings of DCs. This, however, would digress from the focus of this paper and we leave it for future research.

[27] The *inside* features do not damage our model, since removing *sg_inside_DC* and *by_inside_DC* from the ablation model yielded 77.3% accuracy – lower than 78.8% for all features together, though not significantly so.

nouns which display high compoundhood and frequently form OBJ-DCs, but the majority of DCs have very low such values. Only 5.1% of our DCs have a *head_in_DC* value above 50% and as many as 70.3% of them have one under 20%.

Table 12: Head nouns with high compoundhood

Head noun	*Head_in_DC* (%)	OBJ-reading (%)
laundering	94.8	95.5
mongering	91.8	100
growing	68.7	95.2
trafficking	62	100
enforcement	53.7	66.6

Table 12 illustrates the few head nouns that most often appear in DCs and the frequency of an OBJ reading among the DCs they appear as heads of. As visible there, a high frequency of a deverbal noun in DCs correlates with a high value for an OBJ reading of the compound's non-head, as predicted (cf. §4.1).

6.1.6 *Sum-adjectives* and *by*-phrases

The last three features we employed in our study are *sum-adjectives*, *by_outside_DC* and *by_inside_DC*. On their own, they have some predictive power (Table 4), but their removal in Table 5 has no significant impact on the results, showing that *P-R* compensates for their absence. Interestingly, in the corpus-based model in Table 6, the removal of *by_outside_DC* triggers a significant drop, indicating that in the absence of *P-R*, this feature becomes important. Yet, in spite of Grimshaw's expectation for this feature to identify OBJ-DCs, its direction of prediction is NOBJ in all models (see Table 7). As we saw in Section 2.1.1, *by*-phrases are ambiguous and their presence indicates ASN-hood only when the object argument is also realized (see (10)). We considered using the frequency of *by*-phrases co-occurring with *of*-phrases, but the numbers were extremely low. Thus, the unexpected direction of prediction of *by*-phrases might be due to their ambiguity. The other two features do not preserve the direction of prediction (Table 7).

The inconclusiveness of these three features most likely resides in data sparsity. Namely, for the feature *by_outside_DC* the range of frequency in our full database is 0% – 6.22% with 60% of the deverbal head nouns realizing a *by*-phrase in fewer than 1% of their occurrences outside DCs. For *by_inside_DC* the range is between 0% and 4.36%, with 74% of the DCs displaying a *by*-phrase in fewer

than 1% of the cases. For *sum-adjectives* the value is even lower: the frequency ranges between 0% and 1.8%, with 99% of the cases having a value under 1%.

6.1.7 Summary

In summary, *P-R*, the feature based on introspection, is the strongest. It provides a high performance individually and its removal from the model considerably hurts the results. *Suffix* is the strongest corpus-based feature. It brings additional value over the combination of all features including *P-R*, but it does not reach the performance of *P-R* on its own. *Of-outside* is the next valuable feature. On its own, it is not very strong, but it is a very important addition to the other features. Its removal from the combined models hurts the performance considerably. The feature *by-outside* is valuable when only corpus-based features are considered. If *P-R* is present in the model, *by-outside* is unimportant. This indicates that this feature has a considerable overlap with *P-R*. The other features all have predictive power, but their additional predictive power is not very important. They capture the same signal in a less reliable way.

The latent variable that we are trying to capture with the features presented in this study, the ASN-hood of the head, is best represented by the introspection-based feature *P-R*. The corpus-based features *suffix* and *of-outside* have additional value in the combined model, that includes P-R, as the ablation studies show. They seem to help the strong feature P-R to move the model in the right direction. However, although the combination of P-R and the best corpus-based features leads to an improvement (80.3% vs 78.8%), we could not prove that their addition to *P-R* as a single feature model improves the results significantly.

6.2 Implications for our hypothesis

We have identified four features which are important for the interpretation of DCs: *P-R*, *of_outside_DCs*, *by_outside_DC*, and *suffix*. The first three were inspired by Grimshaw (1990), the fourth was introduced by us. As mentioned in §6.1.3, the suffix does not tell us anything about ASN-hood or the compositionality of the DC. It is a corpus-based feature, which scores well on its own and better than most ASN-features from Grimshaw (1990); yet, in ablation studies, it is weaker than *of_outside_DCs*, which is Grimshaw's most important ASN-feature.

The other three features all give us input on ASN-hood, but in different ways. An unexpected result comes from *by_outside_DC*, whose direction of prediction is for NOBJ-DCs, instead of OBJ-DCs. In §6.1.6, we reasoned that this is due to the ambiguity of *by*-phrases, which we could not eliminate by measuring their co-occurrence with *of*-phrases, given data sparsity. The only way we can interpret

this result is that, in combination with other ASN-features which usually point to OBJ-DCs, the input from the ambiguous *by*-phrases was used by the model for the other direction, of NOBJ-DCs.

The features *P-R* and *of_outside_DCs* are the most important for the ASN-hood of head nouns and the implicit compositional interpretation of DCs. They both behave as predicted by our hypothesis. *P-R* represents human intuitions with respect to the ASN-hood of the head noun and scores best in our models. In addition, in line with Grimshaw's claims and our hypothesis, its direction of prediction consistently points to OBJ-DCs. *Of_outside_DCs* is not very strong on its own, but extremely important in combination with the other ASN-features. This is in fact what Grimshaw's combined use of two or three of these morphosyntactic tests (in order to circumvent ambiguity) leads us to expect (see §2.1.1).

These results immediately confirm two things: 1) the validity of these features as identifying ASN-hood and correlated OBJ readings in DCs (i.e., Grimshaw's theory, which is also part of our hypothesis in §1.5); 2) DCs are compositional and easily interpretable to the extent that their head nouns exhibit ASN-properties (i.e., the starting point of our hypothesis in §1). A further implication of these observations is that, indeed, the (deverbal) head noun plays a crucial role in the compositionality and overall transparency of DCs, a conclusion that was reached by other computational studies as well (see §2.2.2).

For the DCs whose heads fail to exhibit ASN-properties and behave like RNs, our features cannot get very far. These DCs behave like RCs, and the relation between their two parts may even be unrelated to the base verb and its modifiers. For these DCs, the addition of other features, especially some designed for non-heads, should improve the results. In this case, it would be worth including features from previous NLP work, which deals with noun-noun compounds in general, especially that reported in §2.2.2. We leave such a study for future research, since it departs from our focus here.

6.3 Comparison to other NLP approaches

We mentioned in §2.2.1 that the aims of previous work on predicting the relation between heads and non-heads in DCs are different from ours. Whereas this work focuses on building classifiers that reach state-of-the-art performance on the task of predicting the relation between the head and the non-head of deverbal compounds, our interest lies in uncovering in how far the behavior of the derived nominals (as ASNs or RNs) can help in predicting the (compositional) relation between head and non-head. As a result, the datasets are very different.

However, we present here some meaningful comparisons with this work. In the two-class prediction task, Lapata (2002) reaches an accuracy of 86.1% compared to a baseline of 61.5%, i.e, 24.6% above the baseline. The accuracy we achieve is 80.3%, i.e., 30.3% above the 50% baseline of our balanced test set. Relative improvements are comparable. Note that the data set of Lapata (2002) included DCs ending in suffixes such as *-er* and *-ee* that are biased in the relation they select. Including them in our dataset could have resulted in better accuracy overall and a stronger predictive power for the *suffix* feature.

Apart from the differences in the data set, we also see large differences in the type of features selected. In this paper we exclusively tested the predictive power of morphosyntactic features of the deverbal noun for determining the covert relation. In the future, it would be interesting to compare these to the encyclopedic/pragmatic features prevalent in the CL literature, by incorporating the latter into our models.

Schulte im Walde, Hätty & Bott (2016) evaluate the influence of several properties of the constituents (frequency, productivity and ambiguity) on the performance of the model in its predictions on transparency. Just as they attribute the influence of these properties to the underlying property of ambiguity, so do we attribute the non-compositionality in the relation between head and compounds (in RNs) to the greater underspecification of RNs in comparison to ASNs. Although we do not have access to transparency ratings for our DCs, we have gathered annotations on their process vs. result interpretation (see §3.3.2). This information can be seen as a proxy for the transparency of the head, because by default the more result-like the DC is, the less transparent it will be.

Furthermore, Schulte im Walde, Hätty & Bott (2016) emphasize the importance of properties of the head and the compound, and to a lesser extent of the modifier (i.e., non-head) for the prediction of the transparency of the compound. The authors stress the need to carefully balance datasets according to the empirical and semantic properties of the compounds, as well as of their heads. We have balanced our data set for corpus frequency of the head and measured the family size of the heads. We have not measured other properties that they have used, but will consider these in future work.

7 Conclusions

In this paper we have presented a study on the (syntactic) compositionality of DCs, as predictable from the morphosyntactic properties of their head nouns. We have employed theoretical insights on the behavior of deverbal nominals, on

the basis of which we collected corpus data, as well as manual annotations. We used this data collection in the form of indicative features in a logistic regression model, by means of which we evaluated the prediction power of each feature for the OBJ (vs. NOBJ) interpretation of the compounds.

Our approach to compositionality comes from the theoretical linguistic perspective, according to which the compositionality of a complex expression (here, the DC) depends on the meanings of its parts, as well as the syntactic relationship between them. To the extent that DCs are headed by deverbal nouns, the fully compositional ones encode the syntactic-semantic relationship between the base verb and its object, while the less compositional ones are underspecified/ambiguous. This difference is traced back to the ambiguity of deverbal nouns between ASN and RN uses from Grimshaw (1990). ASNs preserve the compositional requirements of the base verb, while RNs do not.

Our results confirm our hypothesis that DCs with ASN-heads are compositional and receive an OBJ reading. This study, however, raises a few questions for future research. It especially highlights the need for more study on the role of individual suffixes in the interpretation of the deverbal noun, since previous claims on *-ing* as primarily building OBJ-DCs have not been confirmed. In addition, some tests which are popular in the theoretical literature (e.g., *in/for*-adverbials, agentive and aspectual adjectives, as well as *by*-phrases) could not be used or were not reliable enough as features, probably due to data sparsity. On the one hand, their low attestation in corpora throws doubts on their authenticity, requiring further empirical study. On the other hand, this is also an alarm signal for the need of even larger corpora in order to reliably test theoretical insights, which human intuitions are considerably better at, as proven by our *P-R* feature.

By comparison to the previous NLP work on the transparency of (root) compounds, we did not consider both constituents to evaluate the mapping with the compound; we focused on the head noun, which has a crucial influence on the relationship that it establishes with the non-head in DCs. In future work, we will consider including some predictive features of the non-head. We expect that the encyclopedic features exploited in the NLP literature such as in Nicholson & Baldwin (2006), Lapata (2002), and Grover et al. (2005) will benefit the disambiguation of RNs and the DCs headed by these.

Abbreviations

ASN	Argument Structure Nominal	*P-R*	*process-vs-result* feature
CL	Computational Linguistics	PoC	Principle of Compositionality
DC	Deverbal Compound	POS	Part of Speech
DS	Distributional Semantics	RC	Root Compound
IAA	inter-annotator agreement	RN	Result Nominal
NLP	Natural Language Processing	TL	Theoretical Linguistics

Acknowledgements

We are grateful to Katherine Fraser, Bethany Lochbihler and Whitney Frazier Peterson for annotating our database, to Kerstin Eckart and the INF project in the SFB 732 for important technical support, and to Alla Abrosimova for help with further technical details. This research has been funded by the German Research Foundation (DFG) via a grant offered to the Projects B1 *The Form and Interpretation of Derived Nominals* and D11 *A Crosslingual Approach to the Analysis of Compound Nouns*, as part of the SFB 732 *Incremental Specification in Context* at the University of Stuttgart.

References

Ackema, Peter & Ad Neeleman. 2004. *Beyond morphology*. Oxford: Oxford University Press.

Alexiadou, Artemis & Jane Grimshaw. 2008. Verbs, nouns, and affixation. In Florian Schäfer (ed.), *Working papers of the sfb 732 incremental specification in context*, vol. 1, 1–16. Universität Stuttgart.

Alexiadou, Artemis, Gianina Iordăchioaia & Elena Soare. 2010. Number/aspect interactions in the syntax of nominalizations. *Journal of Linguistics* 46:3. 537–574.

Baroni, Marco & Roberto Zamparelli. 2010. Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. In *Proceedings of the 2010 conference on empirical methods in natural language processing*. Boston.

Borer, Hagit. 2013. *Taking form*. Oxford: Oxford University Press.

Chomsky, Noam. 1970. Remarks on nominalization. In Roderick A. Jacobs & Peter S. Rosenbaum (eds.), *Readings in English transformational grammar*, 184–221. Waltham, MA: Ginn.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: MIT Press.

de Marneffe, Marie-Catherine & Christopher D. Manning. 2008. *Stanford typed dependencies manual.*

Di Sciullo, Anna Maria. 1992. Deverbal compounds and the external argument. In Iggy M. Roca (ed.), *Thematic structure: its role in grammar*, 65–78. Berlin: Foris.

Dietterich, Thomas G. 1998. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation* 10(7). 1895–1923.

Dowty, David. 2007. Compositionality as an empirical problem. In Chris Barker & Pauline I. Jacobson (eds.), *Direct compositionality*, 14–23. Oxford University Press.

Fellbaum, Christiane (ed.). 1998. *WordNet: An electronic lexical database.* MIT Press.

Fokkens, Antske Sibelle. 2007. *A Hybrid Approach to Compound Noun Disambiguation. MA-Thesis. Universität des Saarladens, Saarbrücken.*

Gillick, Dan. 2009. Sentence boundary detection and the problem with the u.s. In *Proceedings of human language technologies: the 2009 annual conference of the north american chapter of the association for computational linguistics, companion volume: short papers*, 241–244. Boulder, Colorado.

Grimm, Scott & Louise McNally. 2013. No ordered arguments needed for nouns. In Maria Aloni, Michael Franke & Floris Roelofsen (eds.), *Proceedings of the 19th amsterdam colloquium*, 123–130. Institute for Logic, Language & Computation, University of Amsterdam.

Grimshaw, Jane. 1990. *Argument structure.* Cambridge, MA: MIT Press.

Grover, Claire, Mirella Lapata & Alex Lascarides. 2005. A comparison of parsing technologies for the biomedical domain. *Journal of Natural Language Engineering* 11:01. 27–65.

Hall, Mark, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann & Ian H. Witten. 2009. The WEKA data mining software: An update. *SIGKDD Explorations* 11(1). 10–18.

Hamp, Birgit & Helmut Feldweg. 1997. GermaNet - a Lexical-Semantic Net for German. In *Proceedings of acl workshop automatic information extraction and building of lexical semantic resources for nlp applications*, 9–15.

Henrich, Verena & Erhard Hinrichs. 2010. GernEdiT - The GermaNet Editing Tool. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner & Daniel Tapias (eds.), *Proceedings of the 7th international conference on language resources and evaluation.* Valletta, Malta.

Huang, Zhongqiang, Mary Harper & Slav Petrov. 2010. Self-training with products of latent variable grammars. In *Proceedings of the 2010 conference on empirical methods in natural language processing*, 12–22. Cambridge, Massachusetts.

Iordăchioaia, Gianina. To appear. English deverbal compounds with and without arguments. In *Proceedings of the 54th annual meeting of the chicago linguistic society*. Chicago.

Iordăchioaia, Gianina, Artemis Alexiadou & Andreas Pairamidis. 2017. Morphosyntactic sources for nominal synthetic compounds in English and Greek. *Zeitschrift für Wortbildung/Journal of Word-formation* 1. 47–72.

Iordăchioaia, Gianina, Lonneke van der Plas & Glorianna Jagfeld. 2016. The grammar of English deverbal compounds and their meaning. In Eva Hajicova & Igor Boguslavsky (eds.), *Proceedings of the workshop on grammar and lexicon: interactions and interfaces*, 81–91. Osaka, Japan.

Juhasz, Barbara J., Yun-Hsuan Lai & Michelle L. Woodcock. 2015. A database of 629 English compound words: Ratings of familiarity, lexeme meaning dominance, semantic transparency, age of acquisition, imageability, and sensory experience. *Behavior Research Methods* 47. 1004–1019.

Kratzer, Angelika. 1996. Severing the external argument from its verb. In Johan Rooryck & Laurie Zaring (eds.), *Phrase structure and the lexicon*, 109–137. Kluwer Academic Publishers.

Lapata, Mirella. 2002. The disambiguation of nominalizations. *Journal of Computational Linguistics* 28:3. 357–388.

Larson, Richard K. 1988. On the double object construction. *Linguistic Inquiry* 19:3. 335–391.

Levi, Judith N. 1978. *The syntax and semantics of complex nominals*. New York: Academic Press.

Libben, Gary, Martha Gibson, Yeo Bom Yoon & Dominiek Sandra. 1997. *Semantic transparency and compound fracture*. Tech. rep. 9. CLASNET Working Papers.

Libben, Gary, Martha Gibson, Yeo Bom Yoon & Dominiek Sandra. 2003. Compound fracture: The role of semantic transparency and morphological headedness. *Brain and Language* 84. 50–64.

Lieber, Rochelle. 2004. *Morphology and lexical semantics*. Cambridge: Cambridge University Press.

Lieber, Rochelle. 2016. *English nouns. The ecology of nominalizations*. Cambridge: Cambridge University Press.

Macleod, Catherine, Ralph Grishman, Adam Meyers, Leslie Barrett & Ruth Reeves. 1998. NOMLEX: A lexicon of nominalizations. In *Proceedings of EURALEX*.

Marelli, Marco & Marco Baroni. 2015. Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review* 122:3. 485–515.

Mitchell, Jeff & Mirella Lapata. 2010. Composition in distributional models of semantics. *Cognitive Science* 34. 1388–1429.

Montague, Richard. 1970. Universal grammar. *Theoria* 36. 373–398.

Napoles, Courtney, Matthew Gormley & Benjamin Van Durme. 2012. Annotated gigaword. In *Proceedings of the joint workshop on automatic knowledge base construction and web-scale knowledge extraction*, 95–100. Montreal, Canada.

Nicholson, Jeremy & Timothy Baldwin. 2006. Interpretation of compound nominalisations using corpus and web statistics. In *Proceedings of the workshop on multiword expressions: identifying and exploiting underlying properties*, 54–61. Montreal, Canada.

Ó Séaghdha, Diarmuid. 2008. *Learning Compound Noun Semantics*. Tech. rep. UCAM-CL-TR-735. University of Cambridge, Computer Laboratory.

Partee, Barbara H. 1984. Compositionality. In Fred Landman & Frank Veltman (eds.), *Varieties of formal semantics: proceedings of the 4th amsterdam colloquium*, 281–311. Dordrecht: Foris Publications.

Reddy, Siva, Diana McCarthy & Suresh Manandhar. 2011. An empirical study on compositionality in compound nouns. In *Proceedings of the 5th international joint conference on natural language processing*.

Roeper, Thomas & Muffy Siegel. 1978. A lexical transformation for verbal compounds. *Linguistic Inquiry* 9. 199–260.

Santorini, Beatrice. 1990. *Part-Of-Speech tagging guidelines for the Penn Treebank project (3rd revision, 2nd printing)*. Tech. rep. Philadelphia, PA, USA: Department of Linguistics, University of Pennsylvania.

Schulte im Walde, Sabine, Anna Hätty & Stefan Bott. 2016. The role of modifier and head properties in predicting the compositionality of English and German noun-noun compounds: A vector-space perspective. In *Proceedings of the 5th joint conference on lexical and computational semantics*, 148–158. Berlin, Germany.

Schulte im Walde, Sabine, Anna Hätty, Stefan Bott & Nana Khvtisavrishvili. 2016. G$_h$ost-NN: A Representative Gold Standard of German Noun-Noun Compounds. In *Proceedings of the 10th international conference on language resources and evaluation*, 2285–2292. Portoroz, Slovenia.

Selkirk, Elisabeth O. 1982. *The syntax of words*. Cambridge, MA: MIT Press.

van Heuven, Walter J. B., Pawel Mandera, Emmanuel Keuleers & Marc Brysbaert. 2014. SUBTLEX-UK: A new and improved word frequency database for British English. *Journal of Experimental Psychology* 67(6). 1176–1190.

Zwitserlood, Pienie. 1994. The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes* 9. 341–368.

Chapter 4

Internal constituent variability and semantic transparency in N Prep N constructions in Romance languages

Inga Hennecke
University of Tübingen

Constructions of the type N Prep N represent one of the most controversial issues in Romance word formation. In particular, their lexical status and their degree of productivity are still crucial points of discussion. Hence, it remains unclear whether these constructions fall within the category of morphological word formation or of syntax. Furthermore, the possibilites for internal prepositional variation remain uncertain. This article takes a constructionist approach within the framework of construction morphology in order to describe the internal constituent variability and transparency of the prepositional element in N Prep N constructions in Spanish, Portuguese, and French, as in Sp. *juego de niños*, *juego para niños* ('kid's game') or in Sp. *cabaña de árbol* and *cabaña en árbol* ('tree house'). A qualitative analysis of large-scale corpus data from the TenTen corpus family indicates that Romance N Prep N constructions may undergo internal prepositional variation. The analysis focuses on the semantic relations of the internal nominal constituents and the semantic transparency of the constructions in the three Romance languages under investigation. The results indicate that semantic relations and semantic transparency play a role in the internal constituent variability of the prepositional element.

1 Introduction

Compounds of the type N Prep N, such as Sp. *bicicleta de montaña* 'mountain bike', Fr. *salle de bain* ('bath room'), or Pt. *história em quadrinhos* ('comic strip'), are generally considered to be the most problematic aspect of research on compounding and word formation in Romance languages. This is because these constructions represent nominal lexical units that clearly approach free syntactic

Inga Hennecke. 2019. Internal constituent variability and semantic transparency in N Prep N constructions in Romance languages. In Sabine Schulte im Walde & Eva Smolka (eds.), *The role of constituents in multiword expressions: An interdisciplinary, cross-lingual perspective*, 99–122. Berlin: Language Science Press. DOI:??

structures (de Bustos Gisbert 1986). Compounds of the type N Prep N have been treated very differently in research on compounding and have also been labeled with many different terms, such as syntagmatic compounds (Buenafuentes de la Mata 2010), syntactic compounds (Rio-Torto & Ribeiro 2009), improper compounds (Kornfeld 2009), phrasal lexemes (Masini & Thornton 2007), frozen multiword units (Guevara 2012), lexicalized syntactic constructions (Villoing 2012), lexicalized phrases (Fradin 2009), and syntactic words (Di Sciullo & Williams 1987). Gen(Guevara 2012)erally, compounding is a mechanism whereby two lexical units are combined. Compounds of the type N Prep N are characterized as lexical units that consist of (at least) two lexical elements that are not orthographically combined. As a result, compounds of the type N Prep N, such as Sp. *traje de baño* ('bathing suit'), do not differ on a formal level from syntactic phrases of the type N Prep N, such as Sp. *libro para niños* ('book for children') (de Bustos Gisbert 1986: 69).

The most problematic issue in current research on compounding of the type N Prep N is the question of the delimitation of syntactic and lexical structures in Romance languages. As the treatment of these constructions is based largely on the theoretical background of the individual author, there is no general agreement on whether or not N Prep N constructions should be included in the class of compounds. Related to this issue is the question of whether these constructions emerge by means of productive word formation processes or are merely 'fossilized' or lexicalized syntactic structures. These two crucial issues will be discussed and analyzed in what follows, with a focus on one particular case of internal constituent variability, the alternation of the internal preposition in N Prep N constructions. A large-scale corpus analysis of this alternation in French, Spanish, and Portuguese supports the adoption of a constructionist approach within a framework of construction morphology. Such an approach allows the internal constituent variation of N Prep N constructions to be represented without recourse to traditional notions of lexicon and syntax.

2 Definition and classification of syntagmatic compounds

As mentioned above, constructions of the type N Prep N are often excluded from descriptions of Romance compounding. Typically, they are classed together with other compound-like constructions lacking an orthographical union, as in the examples from (Masini 2009: 257):

According to Masini, these examples are separated orthographically, show no strong degree of idiomaticity, and appear quite frequently in each of the four

Table 1: Phrasal lexemes in Romance languages

Language	Types	Phrasal lexems	Lit.	Glosses
French	[ADJ N]$_N$	*première violon*	first violin	'first violin'
Italian	[N *da* N]$_N$	*camera da letto*	room from bed	'bedroom'
Portuguese	[N *de* N]$_N$	*cadeira de rodas*	chair of wheels	'wheelchair'
Spanish	[N ADJ]$_N$	*luna nueva*	moon new	'new moon'

languages. The question nevertheless remains as to whether these constructions form part of the class of compounds.

According to Guevara (2012), Spanish syntagmatic compounds, such as *fin de semana* 'weekend' or *sabelotodo* 'know-it-all', should be excluded from the class of Spanish compounds, as these units are clearly syntactic units that contain "certain effects of lexicalization and atomicity in their distribution" (Guevara 2012: 180). In the same way, Villoing (2012) excludes French constructions such as *fil de fer* 'iron wire' and *brosse à dents* 'tooth brush' from her description of French compounds, as they are "lexicalized syntactic constructions that behave like lexical units" (Villoing 2012: 35). The approach taken by Guevara and Villoing indicates, on the one hand, that constructions of the type N Prep N are often considered as syntactic units that lie outside of the core of word formation processes. For this reason, they are regularly neglected in research papers on Romance word formation. On the other hand, this approach shows that N Prep N constructions are frequently interpreted as lexicalized syntactic constructions and, more precisely, as syntactic constructions that have somehow attained a high degree of fixedness. If this is the case, they should also be excluded as belonging to the class of Romance-language compounds, as lexicalization cannot be considered a morphological word formation process.

There is an opposing perspective according to which the constructions mentioned in Table 1 constitute a productive type of word formation and clearly follow productive morphosyntactic rules. According to Rainer, constructions of the type N Prep N are "very productive lexical patterns, which normally continue to obey the rules of [...] syntax (for example, agreement rules), but may occasionally also deviate from them" (Rainer 2016: 2724). This perspective is not new and was already adopted by Benveniste (1974) in his work on French compounds of the type *robe de chambre* ('robe') and *plat à barbe* ('shaving bowl'), for which he claims indefinite productivity (Benveniste 1974: 172). In the course of the present paper, I will provide new empirical evidence in favor of this perspective, using

large-scale corpus data. The analysis will show that N Prep N constructions in Romance languages are highly frequent and productive and that their internal variability follows clear morphological rules that can be mapped using construction morphology.

In order to distinguish N Prep N constructions from other phrase-like constructions, their characteristics must be clearly delineated. According to Buenafuentes de la Mata (2010), a syntagmatic compound may be defined as a lexical element that has been created by the fixation of a syntagm, that keeps its sentential structure, and that therefore shows neither orthographic nor accentual union (Buenafuentes de la Mata 2010: 21ff.). De Bustos Gisbert (1986) states that Spanish N Prep N compounds differ from syntactic units on the syntactic level in two respects. First, they have a fixed word order, for example *ojo de buey* ('porthole') cannot be reordered as * *buey ojo de*. Second, there is generally no unproblematic substitution of their constituents; for example **ojo de vaca* (Val Àlvaro 1999: 4825). On a morphological level, he adds that N Prep N constructions show the same characteristics as other compounds in terms of gender and number agreement, the presence of composition markers, and the ability to undergo further derivation and to form collocations (de Bustos Gisbert 1986: 77). According to Masini, N Prep N constructions are of major interest, as they follow the syntactic rules of head modification of a nominal phrase by a prepositional phrase. That means that in Romance languages, N Prep N constructions are generally left-headed, and that inflectional processes are performed on the head of the construction (Masini 2009: 257). Val Àlvaro (1999) adds a fundamental characteristic on the semantic level: the absence of compositional meaning that may lead to syntactic reinterpretation of the complex nouns (Val Àlvaro 1999: 4827). This means, that syntagmatic compounds, in contrast to syntactic units, represent one single naming unit at the semantic level; that is, they refer to one specific conceptual representation, as in Fr. *sac à main* ('purse').

In this paper, I will focus on the syntactic criteria given by de Bustos Gisbert, specifically, on the impossibility of constituent substitution. This criterion does not appear to be suitable for purposes of differentiating syntactic and lexical elements, as the delimitation between syntactic and lexical N Prep N constructions remains a matter of controversy. Here, I will show that variation of the internal preposition can best be explained within a constructional framework. I will then argue with regard to the internal preposition not only that the substitution of internal constituents is possible in N Prep N constructions, but also that the substitution is a rule-governed process and depends largely on semantic factors, particularly the semantic relation of the nominal constituents.

When investigating the semantic relations of constituents of N Prep N constructions, it is crucial to consider the notions of semantic transparency and semantic opacity. In current research, the term semantic transparency denotes the way in which the meaning of a complex construction can be derived from the meaning of its constituents (Zwitserlood 1994). In that sense, the French N Prep N construction *salle de bains* 'bathroom' is considered semantically transparent, whereas the Spanish construction *ojo de buey* ('porthole', lit. 'bull's eye') is considered semantically opaque. Bell and Schäfer view semantic transparency and semantic opacity as scalar notions, lying at either end of a continuum (see Bell & Schäfer (2016) for a detailed discussion on semantic transparency). Later in the present study, I will discuss whether the semantic transparency of an N Prep N construction determines the possibility of internal constituent variation.

3 Internal constituent variation in N Prep N constructions – the role of the preposition

Characteristic of N Prep N constructions, and a crucial factor in their delimitation, is their resistance to paradigmatic variation. In the context of the delimitation of nominal compounds and noun phrases of the type N Prep N in Portuguese, Rio-Torto and Ribeiro state, that the "(im)possibility of lexical insertion" is one of the most important tests of compoundhood (Rio-Torto & Ribeiro 2012: 9). They go further, claiming that if internal changing is allowed, "we are no longer dealing with compounds ($[N[PrepN]]_N$) but with noun phrases ($[N[PrepN]]_{NP}$)" (ibid.). When speaking of internal change, Rio-Torto and Ribeiro refer principally to changes in determination, as in Pt. *fim de semana* ('weekend') and Pt. *fim da semana* ('end of this week'), and to changes effected through insertion of lexical material, as in *fim de última semana* ('end of last week'). As these examples suggest, internal constituent variation is generally seen as a crucial test of delimitation between compounds and syntactic structures. Similarly, Masini argues that, for lexical elements, "paradigmatic variation is blocked, since the words in the construction cannot be substituted by a near-synonym, which should not be a problem for normal phrases" (Masini 2009: 259). Masini defines paradigmatic blocking as the inability to replace a constituent of the construction by another paradigmatically fitting constituent. She also refers to cases of paradigmatic blocking of a nominal unit of a N Prep N construction, as the Italian examples *casa di cura* ('nursing home') and **abitazione di cura* ('*nursing domicile'). In this case, *casa di cura* is a fixed naming unit that loses its semantic meaning when there is paradigmatic variation of a nominal element. The following analysis will show

that paradigmatic blocking holds particularly true for N Prep N constructions with a stronger degree of semantic opacity and idiomaticity. More transparent N Prep N constructions allow productive and rule-governed internal constituent alternation, as the analysis will show by means of the prepositional constituent.

In the literature, all references to a delimitation test of constituent variability neglect the prepositional constituent in N Prep N constructions. The prepositional element is fundamental in N Prep N constructions, but its status is far from clear. In all the Romance languages under investigation here, the preposition *de* is the most frequently used prepositional constituent in N Prep N constructions. In the case of Spanish, Buenafuentes de la Mata (2010) cites various examples of N Prep N constructions with prepositions other than *de*, such as *leche en polvo* ('milk powder'), *cita a ciegas* ('blind date'), *caridad con uñas* ('self-serving favor'), *pozo sin fondo* ('bottomless pit'), and *caballo con arcos* ('pommel horse'). She adduces the appearance of prepositions other than *de* as evidence for the structural complexity of N Prep N constructions in Spanish. The same case can be made for the other languages under investigation in this paper (i.e. French and Portuguese), which show the same ability to form N Prep N constructions with other prepositions.

This paper concentrates on a specific set of (partially) synonymous prepositions in French, Spanish, and Portuguese, which are Fr. *de* ('of'), *à* ('to'), *en* ('in'), and *pour* 'for'; Sp.*de* ('of'), *a* ('to'), *en* ('in'), and *para* ('for') as well as Pt. *de* ('of'), *a* ('to'), *em* ('in') and *para* ('for'). These prepositions may all appear in N Prep N constructions and they may all undergo internal alternation and variation in the three languages under investigation. Consider examples (1) to (3) from the TenTen corpora:

(1) a. Sp. *fuente de horno – fuente para horno* 'casserole'
b. Pt. *água de lavagem – água para lavagem* 'wash water'
c. Fr. *livre d'enfant – livre pour enfants* 'children's book'

(2) a. Sp. *motores de gasolina – motores a gasolina* 'gas engine'
b. Fr. *jauge d'essence – jauge à essence* 'fuel gauge'
c. Pt. *fogão de lenha – fogão a lenha* 'wood stove'

(3) a. Fr. *chemise de coton – chemise en coton* 'cotton shirt'
b. Pt. *bracelete de aço – bracelete em aço* 'steel bracelet'
c. Sp. *ciclismo de pisto – ciclismo en pisto* 'track cycling'

Example (1) shows internal variation of the prepositional elements *de* and *pour/para*. While the constructions containing *de* are considered to have a lexical status, the constructions with *pour/para* are generally considered to be syntactic constructions, as they pass certain of the classification tests mentioned above. In contrast to the construction with *de*, they allow substitution and insertion, as the two tests of compoundhood demonstrate: *fuentes de vidrio para horno* ('glass casserole for the oven'), *fuentes profundas para horno* ('deep casserole for the oven'), but **fuentes de vidrio de horno* and **fuentes profundas de horno*. Example (3b) demonstrates the internal alternation of the prepositions *de* and *en/em*. Here, alternation is possible without changing the semantic context of the whole construction or its degree of semantic transparency. Example (2), the prepositions *de* and *a/à* alternate in N Prep N constructions without changing the lexical status of the respective constructions. Nonetheless, these constructions differ in their frequency of usage, productivity, and fixedness, as well as in their degree of lexicalization and of idiomaticity. Especially in French, alternation of *de* and *à* may indicate a change in meaning, as in *verre de vin* ('glass of wine') and *verre à vin* ('wine glass'). In this case, the interpretation of both constructions as two distinct products of word formation is reasonable (this specific case will be discussed in detail in the course of the corpus analysis). In other cases, such as Pt. *fogão de lenha – fogão a lenha* ('wood stove'), no clear semantic difference is visible, as attested by native speakers of Brazilian and European Portuguese: internal variation is possible inside one construction (a more detailed discussion of the examples will follow in the upcoming section). As mentioned above, authors including Rio-Torto & Ribeiro (2009) interpret constructions as of the type in Examples (1–3) as syntactic units, on the grounds that they do not pass all the delimitation tests for compoundhood. The following theoretical discussion and empirical analysis will show that it is neither necessary nor possible to draw a clear distinction between syntactic constructions and lexical constructions; the possibility of alternating prepositional elements in N Prep N constructions depends largely on the semantic function of the N2, and the fixedness, semantic transparency, the idiomaticity of the whole construction.

Another problem in analyzing alternation of prepositional elements concerns the role of the prepositions. In Romance languages, the prepositions *de* and *à/a* in particular have often been considered as semantically 'empty' units that do not contain meaning. This perspective has often been applied to the prepositional element in N Prep N constructions, for example, Bartning states that prepositions in French N Prep N constructions do not code any specific meaning and that they function only as linking elements (Bartning 1993: 164). Similarly, Cadiot describes these prepositional elements as 'colorless prepositions' (Bartning 1993:

164). For the French prepositions *de* and *à*, Bosredon and Tamba use the term of "opérateur de couplage" ('linking operator') (Bosredon & Tamba 1991: 44). Cadiot (1997) sees prepositions in N Prep N constructions as elements that express the operation of a construction or the denomination of a subclass of N1; he associates the prepositional element with a 'referential calibration' of N1. In contrast, Laumann (1998) notes the importance of distinguishing between different types of meaning, when investigating the function and meaning of prepositions in nominal compounds. Laumann differentiates between system meaning (*Systembedeutung* - the sum of the meaning patterns of the constituents in the N Prep N construction), word meaning (*Wortbedeutung* - the meaning of the construction on the level of word formation), and lexicon meaning (*Wortschatzbedeutung* - the meaning of the construction as a naming unit in the lexicon) (Laumann 1998: 32).

Other authors interpret the possibility of elision of the prepositional element, as Sp. *ducha de teléfono* > *ducha teléfono* ('detachable shower head') or Sp. *crédito de vivienda* > *crédito vivienda* ('residence loan'), as evidence of the semantic emptiness of the preposition. However, the elision of the prepositional element is only possible in certain strongly lexicalized constructions. Therefore, a counterargument may be based on the same evidence, given the elision of the prepositional element is not possible in most cases. In the present paper, I argue that the elision of prepositional elements is not proof of a lack of semantic content. The elision can be explained in terms of common processes of language change that may or may not take place in certain lexicalization processes within complex units. The alternation between prepositional elements, exemplified above, is a productive word-formation process that differs clearly from mere lexicalization processes. Therefore, the following qualitative corpus analysis considers the internal constituent alternation of the prepositional element from a comparative perspective and does not focus the elision of this element. The analysis adopts a constructionist approach based on Goldberg (1995; 2006), with a special focus on construction morphology as introduced by Booij (2010; 2015).

4 N Prep N constructions in construction grammar and morphology

Since Goldberg's seminal work *Constructions* in 1995, the constructional approach has had a strong impact on linguistic research. Constructions are considered as conventionalized form-meaning pairs that can be found at all levels of abstraction in language, are dynamically formed, and may be changed continuously.

They are acquired via general processes of abstraction, generalization, and categorization. Goldberg (2006) considers any linguistic unit to be a construction, if "some aspect of its form or function is not strictly predictable from its component parts" (Goldberg 2006: 5). Furthermore, units are considered to be stored as constructions if they can be fully predicted and if they are sufficiently frequent (ibid.).

In his theory of construction morphology, Booij (2015) applied the general notions and concepts of construction grammar to morphological units that have traditionally been regarded as morphological. The underlying assumption of the theory of construction morphology is that a construction may have characteristics that cannot be derived from their constituents (Booij 2015: 3). Booij cites the example of the reduplication of nouns in Spanish in order to express the notion 'real', as in *un café café* ('a real coffee'). He establishes the notion of conceptual schemas and subschemas, defined as schematic representations of morphological constructions. These schemas represent a correlation between form and meaning:

(4) (Booij 2015: 2)
<[[x]Vi er]Nj ↔ [Agent of SEMi]j>

This example indicates that a word with base x, in this case an English infinitive verb form, can transform into a noun with the meaning 'agent of the base word (SEM)' by adding the suffix *-er* (Booij 2015: 2). The variable x denotes the phonological content of the base word, i denotes the meaning of the base word, and j shows that the meaning of the complete construction depends on the form of the complete construction (ibid.). Masini (2009) applies the theory of construction morphology to constructions of the type N Prep N in Italian, taking them to represent an abstract template that is stored in the mental lexicon (Masini 2009: 261). Masini further notes that this abstract template features a certain degree of productivity and is associated with a concrete naming function (ibid.). By means of a specific inheritance mechanism, based on instance inheritance links (Goldberg 1995), constructions that are more and more specific can be derived from the abstract template. This may be done by categorical specification (filling an unspecified slot with a specific category), as in N Prep N or N Prep V, by lexical specification (filling a slot with specific lexical material), as in N *de* N or by a completely lexical construction, such as *casa di cura* (Masini 2009: 261). Figure 1 demonstrates the application of this theory to French constructions of the type N Prep N.

This figure shows the inheritance hierarchy from the abstract template [N1 *de* N2]$_N$, which here is an intermediate construction of the abstract template [N1

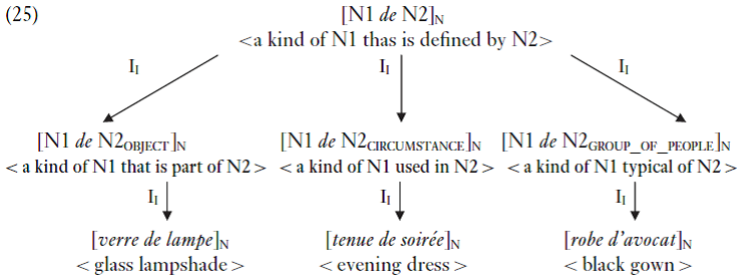


Figure 1: Inheritance hierarchy for N Prep N templates in French (Masini 2009: 263)

Prep Y] and [N1 Prep N2]. From the level [N1 *de* N2]$_N$, it is possible to proceed to a second intermediate lexical level, which indicates the semantic function of N2, and to conclude at a completely lexical level, which shows the lexical result with a concrete naming function. According to Masini, this model can also clarify and describe new occurrences of the N1 *de* N2 construction (Masini 2009: 263).

The following qualitative corpus analysis aims to apply the concept of construction morphology presented by Booij (2010; 2015) and exemplified by Masini (2009) to a cross-linguistic comparative analysis of large-scale corpus data for Spanish, French, and Portuguese N Prep N constructions. The focus of the analysis is on constituent variation of the internal prepositional element in N Prep N constructions in these three languages, and I will apply Masini's inheritance hierarchy template (Figure 1) to the internal variability of prepositional constituents. It is useful to include a further intermediate level prior to the first and second levels of the hierarchy for N Prep N templates mentioned above. This additional level contains the abstract template with the semantic function of N2, which in the following corpus analysis is shown to be a crucial factor in determining the possibility of internal prepositional variation. For the purposes of the present analysis, the inheritance hierarchy for N Prep N templates may be visualized as in Figure 2.

This figure shows the inheritance hierarchy adapted from Fig. 1 by means of Example (3a). As mentioned above, the added abstract intermediate levels are intended to reflect the possibility of that prepositional variability for certain N Prep N constructions and the dependence of this variability on the semantic function of the nominal constituents of the construction. The objectives of the following qualitative corpus analysis are to apply the inheritance hierarchy in Fig. 2 to a large-scale corpus of natural speech data for French, Spanish and Portuguese and to compare the internal prepositional variability of N Prep N constructions

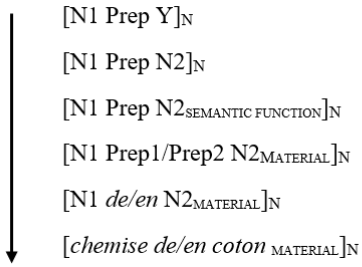


Figure 2: Inheritance hierarchy for internal variation in N Prep N templates in French (adapted from Masini (2009))

in these three languages.

5 Qualitative corpus analysis

As mentioned above, the present corpus analysis is intended to investigate internal constituent alternation of the prepositional element in N Prep N constructions in Spanish, French and Portuguese. The focus is on the alternation between *de* and *à/a*, *de* and *en/em*, and *de* and *pour/para*. This study builds on a quantitative corpus survey on the internal alternation of the prepositional element in N Prep N constructions in Spanish, French, and Portuguese by means of large-scale corpus data (Hennecke & Baayen 2017). Hennecke and Baayen showed that internal prepositional variation in the three languages under investigation is possible, but that these languages show different characteristics in terms of frequency and productivity of such alternation (Hennecke & Baayen 2017: 144). The quantitative analysis of the three languages focused on frequency of types and token, productivity (i.e. probability of previously unobserved types), and population size (i.e. potential number of formations) (Hennecke & Baayen 2017: 139). The results show that Portuguese and, to a lesser extent, French, allows productive internal constituent variation of the prepositional element. In contrast, Spanish does not show productivity in internal variation, which is demonstrated by the absence of hapax legomena (ibid.). At the same time, Spanish has the greatest tendency to employ the preposition *de* in N Prep N constructions. In French, the prepositions *à* and *pour* are slightly more productive than in the other two languages, French tends to avoid constructions using *avec*, whereas constructions with *com* are productive in Portuguese. The latter tendency may be explained by the fact that French prefers NA-constructions over constructions of the type N *avec* N.

The aim of the present corpus analysis is to investigate the results from the above-mentioned study from a qualitative perspective. In this qualitative survey, the internal prepositional variability will be investigated from a mostly semantic perspective, combined with a constructionist approach. Here, the focus will be on which nominal semantic functions allow prepositional variability and whether the variability depends on the semantic transparency of the construction. To that end, this corpus analysis is based on the same dataset as in Hennecke & Baayen (2017), namely three web corpora from the TenTen corpus family from Sketchengine, the French corpus frTenTen12, the Spanish corpus esTenTen11 and the Portuguese corpus ptTenTen11. The TenTen corpora are large-scale web corpora with the counts displayed in Table 2.

Table 2: Corpus Information of the TenTen corpora for French, Spanish and Portuguese (https://the.sketchengine.co.uk)

	frTenTen12	esTenTen11	ptTenTen11
Tokens	11,444,973,582	10,994,616,207	4,626,584,246
Words	9,889,689,889	9,497,402,122	3,900,501,097
Sentences	456,065,104	407,205,587	190,221,913
Paragraphs	188,079,362	213,364,685	91,248,976
Documents	20,400,411	22,287,566	10,216,060

In order to perform a qualitative analysis of the data, all N Prep N constructions were extracted automatically from the corpora, keeping only those that appear with more than one internal prepositional element. The present analysis focuses exclusively on N Prep N constructions and therefore excludes constructions of the type N Prep Det N. The data were manually inspected by excluding grammaticalized constructions (for example Fr. *face à N*, Sp. *gracias a N*), binominal pairs (i.e. Fr. *temps en temps*, Sp. *dia a dia*), and antonyms (Fr. *chien avec/sans laisse* 'dog with/without leash'). Table 3 demonstrates the underlying dataset for the qualitative analysis.

This dataset shows important differences in type-token frequency between the three languages (Hennecke & Baayen 2017). Portuguese presents by far the greatest number of different types and tokens of N Prep N constructions with more than one internal preposition. In contrast, Spanish has very few different types but a considerable number of tokens. This can be interpreted as a small number of different N Prep N constructions, but these few types appear quite often in the corpus data. The French data show a significantly higher number of different

Table 3: Type and token counts for the underlying dataset with all pairs of nouns that are attested with at least two different internal prepositions

	Types	Tokens
French	1062	6991
Spanish	547	10219
Portuguese	6795	58932

types than the Spanish data, but a lower number of different tokens. Here, more different types occur less often in the corpus data (for a detailed quantitative analysis of the data see Hennecke & Baayen 2017). In what follows, a qualitative analysis of selected pairs of internal prepositions is presented in order to investigate whether these differences also appear at a qualitative level, with a special focus on the semantic functions of the nominal constituents and the semantic transparency of the constructions. The specific semantic relations were established with regard to the current literature on the semantic relations of nominal constituents in nominal compounds (Gagné & Shoben 1997; Gagné & Spalding 2009; Girju et al. 2005). They were subsequently modified and adapted to the specific case of N Prep N constructions in the corpus data under investigation. It is not possible to list and discuss all occurrences of all types in the present paper; only selected examples will therefore be discussed and analyzed. Where necessary, references will be made to frequency of occurrence.

5.1 The preposition *de* in N Prep N constructions

In all three languages under investigation, the preposition *de* is most often used to combine two nominal expressions, as in Fr. *salle de bain* ('bathroom'), Sp. *botas de agua* ('rubber boots'), or Pt. *moinho de vento* ('wind mill'). Therefore, the preposition *de* appears in all pairs of internal prepositional variation analyzed in the following section. The three data sets also show internal variation for prepositions other than *de*, but these pairs are not the subject of the present analysis. As mentioned above, the preposition *de* has been much discussed; it has often been considered an 'empty' or 'colorless' preposition that lacks any kind of semantic content and that merely fulfills a linking functions. This completely functional approach is not adopted in the present paper, for reasons given above. In the present account, I follow a constructionist approach (see Masini 2009), in which the prepositional constituent in N Prep N constructions is an element of semantic

consequence to the whole construction (Masini 2009: 262). Masini states the example of Italian N1 *di* N2 intermediate lexical constructions, which clearly differ semantically from N1 *a* N2 intermediate lexical constructions. She also emphasizes that, according to Johnston and Busa (1996), "the prepositions *da*, *di* and *a* in Italian N+PREP+N expressions, under certain conditions and in combination with certain classes of nouns, are specialized for different kinds of modification" (Masini 2009: 262). In the further analysis, the statement from Masini will be refined, since in the present data, the intermediate lexical constructions N1 *de* N2 and other intermediate lexical constructions (e.g. *N1 para/pour* N2) may overlap semantically under certain conditions. These cases will be exemplified below in a cross-linguistic comparative analysis. In the present analysis, the preposition is seen not as a semantically opaque constituent but as a constituent with a specific semantic value determined by the semantic functions of the nominal constituents.

The preposition *de* in French, Spanish, and Portuguese has been described as expressing various relations (Bartning 1993: 187). In binominal constructions, it expresses, for instance, a relation of possession (Sp. *el ordenador de Luis* 'Luis' computer', Fr. *la voiture de Jean* 'John's car'), characterization (Fr. *statut de valeur* 'status'), instrument (Fr. *coup de baton* 'blow'), material (Fr. *papier de soie* 'silk paper'), a part-whole relation (Sp. *puerta de casa* 'front door', Pt. *ponta do dedo* 'fingertip'), an affiliation (Fr. *fils de roi* 'king's son'), a content (Fr. *tasse de café* 'cup of coffee'), a defining characteristic (Sp. *hotel de lujo* 'luxury hotel') or a purpose (Pt. *vestido de novia* 'wedding dress') (for more examples in French see (Lang 1991: 291 ff.).

5.2 Internal variation between *de* and *a/à*

Internal variation between the prepositional constituents *de* and *à* has been the subject of several articles and books on French prepositions and nominal syntagms (e.g. Anscombre 1990; Lang 1991; Bosredon & Tamba 1991; Cadiot 1997). However, it is interesting that this discussion has no equivalent in the literature on Spanish and Portuguese prepositions. This is because such internal variation does not take place in Spanish and only to a small extent in Portuguese. The sole example of internal variation of *de* and *a* in the Spanish corpus data is the following:

(5) [N1 *de/a* N2 $_{\text{TYPE/SPECIFICATION}}$]$_{\text{N}}$> *freno de/a disco*, 'disk brake'

Here, the construction containing *de* is far more frequent and the lexicalized form can be found in dictionaries. Still, the construction *freno a disco* also occurs

regularly in the corpus data of the esTenTen corpus, with a frequency of 0.10 occurrences per million. However, the corpus data shows that the internal variation of *de* and *a* is neither frequent nor productive in Spanish, as only one example of one type can be found in this large-scale internet corpus. In Portuguese, the ptTenTen data shows at least two Intermediate lexical constructions with variation of *de* and *a* that present a certain productivity:

(6) [N1 *de/a* $N2_{PURPOSE}$]$_N$
forno de/a microondas, forna de/a lenhas
'microwave oven', 'wood stove'

[N1 *de/a* $N2_{TYPE/SPECIFICATION}$]$_N$
lampião de/a gás, pilhas de/a combustível
'gas lantern', 'fuel cell'

The template [N1 *de/a* $N2_{TYPE/SPECIFICATION}$]$_N$, in particular, is frequently present in the corpus data and is expressed via different types, as in *motor de/a combustão* ('combustion motor') or *bomba de/a vácuo* ('vacuum pump'). It is striking that many of these types are technical terms. It is possible to perceive a semantic difference in both intermediate constructions, where the type N1 *a* N2 more clearly indicates the material part of the N2 constituent, the type N1 *de* N2 focuses semantically on complementing N1 and creating a construction that is a subtype of N1. However, the first sample surveys and questionnaires revealed that native speakers of European and Brazilian Portuguese do not perceive a difference in the semantic meaning patterns or, more precisely, in the semantics of the whole construction.

For the French data, a very different pattern appears in the analysis of internal variability of *de* and *à*:

(7) [N1 *de/à* $N2_{PURPOSE}$]$_N$
fil de/à pêche
'fishing rod'

[N1 *de/à* $N2_{TYPE/SPECIFICATION}$]$_N$
course à/d'obstacles
'obstacle course'

(8) [N1 *de/à* $N2_{INGREDIENT}$]$_N$
crème au/de citron
'lemon creme'

[N1 *de/à* $N2_{CONTAINER}$]$_N$
conteneur de/à déchets
'waste bin/bin with waste'

(9) [N1 *de/à* $N2_{TRANSPORT}$]$_N$
course de/à vélo
'biking trip'

The existing literature on *de-à* alternation in French emphasizes that there is a semantic difference between binominal constructions containing *de* and *à* and that this semantic difference affects not only the prepositional element itself but also the whole naming unit. This becomes very clear in more detailed analysis of the examples from the template [N1 *de/à* $N2_{CONTAINER}]_N$. In all these examples, the intermediate lexical construction N1 *à* N2 designates the container itself, as in *flûte à champagne* ('champagne glass') or *corbeille à fruit* ('fruit bowl'). In contrast, the intermediate lexical construction N1 *de* N2 denotes the content of the container, as in *flûte de champagne* ('a glass of champagne') or *corbeille de fruit* ('a bowl of fruits'). In these cases, according to Cadiot (1997), *de* turns the interpretation of the construction toward the N2 and constructs a quantified image of the referent, whereas *à* turns the interpretation toward the N1 and permits a qualified image of the reference (Cadiot 1997: 44). That is, *de* carries an effect of quantification whereas *à* carries a semantic notion of qualification. For cases of the intermediate lexical construction [N1 *de/à* $N2_{INGREDIENT}]_N$, such as *salade d'écrevisses* and *salade aux écrevisses* ('crawfish salad'), Lang (1991) states that the preposition *à* connects N1 and N2, whereas the preposition *de* derives N1 from N2. That is to say that *à* describes an ingredient, whereas *de* describes a substance (Lang 1991: 283). In the same way, in the examples of [N1 *de/à* $N2_{TYPE/SPECIFICATION}]_N$ and [N1 *de/à* $N2_{MEANS\ OF\ TRANSPORT}]_N$, it can be seen that *à* points to the material object *vélo* or *obstacles*, whereas *de* more likely complements the N1, and hence the whole construction describes a subtype of N1. According to Cadiot, the semantic differences that occur through the variation of the prepositions *de* and *à* can be accounted for in terms of the more abstract, categorization that is the opposition of intension and extension (Cadiot 1997: 43). On this view, *de* constructs an extensional reference directly, whereas *à* creates an extensional reference indirectly by passing over an intentional reference (Cadiot 1997: 62).

From a constructionist perspective, it can be stated that only in the template [N1 *de/à* $N2_{CONTAINER}]_N$ does the semantic value of the whole construction change, as in *conteneur de déchets* ('bin containing waste') and *conteneur à déchets* ('waste bin'). In this case only, we have to deal with two different naming units when *de* and *à* alternate. Therefore, only here it is appropriate to refer to two different constructions, [N1 *de* $N2_{CONTAINER}]_N$ and [N1 *à* $N2_{CONTAINER}]_N$, which lead to two different naming units at the lexical level. In all the other cases mentioned above, the variability of *de* and *à* does not lead to different semantic interpretations of the lexical outcome, but only to a difference in the semantic weight of certain meaning patterns in the interpretation. Therefore, in all other cases, the inheritance hierarchy from the previous section of this paper can be applied in order to capture the internal constituent variation.

To conclude this analysis, it can be stated that all constructions that allow internal constituent variation of the prepositional element are semantically transparent. The analysis shows that alternation of the internal prepositional constituent does not go along with the semantically more opaque constructions in the languages under investigation, since normally in these cases, the semantic functions of the nominal constituents cannot always be clearly determined. In Spanish and Portuguese, the internal variation is only possible in very specific cases of semantic function of the nominal constituents. French, however, allows more internal variation of *de* and *à*, and this appears to be governed by the semantic functions of the nominal constituents.

5.3 Internal variation between *de* and *em/en*

The variation between *de* and *en/em* in N Prep N constructions has received little attention in the literature. On a general level, Lang states that in French, *en* between two nouns indicate the location of N1, as in *arc-en-ciel* ('rainbow') and *une ville en Italie* ('a city in Italy'), the characterization of N1, as in *ange en stuc* ('stucco angel'), a way of preparation of N1, as in *une salade en vinaigrette* ('a salad with dressing'), the material of N1, as in *robe en soie* ('silk dress'), the form in which N1 appears, as in *fleurs en bouquet* ('bouquet of flowers'), the condition in which N1 stands, as in *arbre en fleur* ('blooming tree'), or a field in which N1 operates, as in *expert en assurances* ('insurance expert') (Lang 1991: 411). According to Laumann, French *de* and *en* are not always interchangeable in the case where N2 refers to the material of N1 (Laumann 1998: 55). On the basis of an analysis of French grammar and dictionary entries, Laumann states that *en* appears more regularly with a predicative supplement than *de*, gives more concrete information about the material, and is less strongly linked to the N1. However, the most important difference seems to be that *en* cannot appear in more opaque constructions with a (partially) idiomatic reading. Laumann cites the examples of *homme de fer* ('iron man') and *yeux d'acier* ('steely eyes'), where it is not possible to substitute *en* for *de* (Laumann 1998: 55). In the French data, most of these relations can also be seen in variations of *de* and *en*, as in the following examples:

(10) [N1 *de/en* $N2_{MATERIAL}$]$_N$
chemise de/en coton
'cotton shirt'

[N1 *de/en* $N2_{FIELD}$]$_N$
étudiant de/en Sciences Po
'student of politics'

(11) [N1 *de/en* $N2_{LOCATION}$]$_N$
course de/en montagne
'mountain race'

[N1 *de/en* $N2_{CONDITION}$]$_N$
maison de/en vente
'house for sale'

(12) [N1 *de/en* $N2_{GROUP}$]$_N$
dîner de/en famille
'family dinner'

[$N1_{ACTION}$ *de/en* $N2_{MATERIAL}$]$_N$
dépenses d'/en énergie
'energy expenditures'

In each of these cases, the variation between *de* and *en* does not trigger any strong meaning difference between the two construction types: it is possible to talk about internal variability rather than about two different types of constructions referring to different naming units. Nonetheless, certain differences are visible, as Laumann (1998) pointed out. For instance, *en* is generally less closely linked to N1 and more often introduces a complement. Constructions with *en* also appear to have a lesser degree of fixedness and put the focus on the N2. The present analysis confirms Laumann's observation that the alternation of *de* and *en* is only possible in semantically transparent constructions that do not include any idiomatic meaning.

(13) [N1 *de/em* $N2_{MATERIAL}$]$_N$
bracelete de/em aço
'steel bracelet'

A very similar picture emerges from the analysis of the Portuguese data, as shown in the following examples:

(14) [N1 *de/em* $N2_{MATERIAL}$]$_N$
bracelete de/em aço
'steel bracelet'

[N1 *de/em* $N2_{FIELD}$]$_N$
profissional de/em artes
'art professional'

(15) [N1 *de/em* $N2_{LOCATION}$]$_N$
surf de/em ondas
'surf (on waves)'

[N1 *de/em* $N2_{CONDITION}$]$_N$
crianças de/em risco
'children at risk'

(16) [N1 *de/em* $N2_{GROUP}$]$_N$
almoço de/em família
'family lunch'

[N1 *de/em* $N2_{MEDIUM}$]$_N$
comentário de/em áudio
'audio commentary''

The Portuguese data show almost the same intermediate lexical constructions that function with a variation between *de* and *em*. The only difference is in the intermediate construction [N1 *de/em* $N2_{MEDIUM}$]$_N$, where N2 designates the medium via which N1 is transferred. In contrast, the French data offer the intermediate construction [$N1_{ACTION}$ *de/en* $N2_{MATERIAL}$]$_N$, which indicates a concrete action referring to a specific (raw) material. Nevertheless, from a quantitative perspective, the internal variation between *de* and *en/em* is by far more frequent in the Portuguese data.

From a quantitative perspective, the variation between *de* and *en* is quite rare in the Spanish data, but the qualitative analysis shows a more variate picture:

(17) [N1 *de/en* $N2_{MATERIAL}$]$_N$
construcción de/en madera
'wood construction'

[N1 *de/en* $N2_{FIELD}$]$_N$
grado de/en ingeniería
'engineering degree'

(18) [N1 *de/em* $N2_{LOCATION}$]$_N$
ciclismo de/en pista
'track cycling'

[N1 *de/en* $N2_{CONDITION}$]$_N$
obras de/en construcción
'construction site'

(19) [N1 *de/en* $N2_{MEDIUM}$]$_N$
entrevista de/en radio
'radio interview'

The examples show that Spanish allows the same internal variation as Portuguese, except that the template [N1 *de/en* $N2_{GROUP}$]$_N$ was not present in the data. In Portuguese and French, there are no strong meaning differences between the two templates, and therefore they can be counted as variants rather than as two distinct forms. In Spanish, as in French and Portuguese, the same subtle differences in degree of fixedness and focus of the constituents can be observed.

Overall, it is possible to state that variation between *de* and *en/em* is possible in all three languages under investigation. The differences appear to exist at the quantitative level rather than on the specific semantic meaning patterns. In all three languages, different templates can demonstrate and explain the possible alternation between *de* and *en/em*. For most cases, these templates overlap in the three languages. Therefore, it is possible to apply the inheritance hierarchy mentioned in section 4 to all of the examples.

5.4 Internal variation between *de* and *pour/para*

For French binominal compounds of the type N Prep N, Laumann (1998) states that the preposition *pour* occurs quite rarely. This may be explained by the fact that *pour* is less abstract than other prepositions, such as *de* or *à*: *pour* indicates a very concrete meaning of purpose or determination, whereas *de* shows a less definite meaning pattern. Therefore, *de*, as a semantically more opaque constituent, offers a wider scope for application than *pour*, but in some cases both prepositions are interchangeable, as in the following examples:

(20) [N1 *de/pour* $N2_{USER}$]$_N$
collier de/pour chien
'dog collar'

[N1 *de/pour* $N2_{PURPOSE}$]$_N$
décoration de/pour mariage/table
'marriage/table decoration'

(21) [N1 *de/pour* $N2_{USER(OBJECT)}$]$_N$
musique de/pour piano
'piano music/music for piano'

The French data show that variation between *de* and *pour* only is possible in cases where N2 designates a user (or a benificiary), or where N2 specifies the purpose of N1. In all three templates given above, N2 serves to form a subtype of N1. However, the templates containing the preposition *de* point more clearly to the N1 and focus on the interpretation of the whole template as a subtype of N1. In templates containing the preposition *pour*, the preposition is clearly attached to the N2, and the semantic emphasis is on N2. Furthermore, the preposition *pour* clearly carries the interpretation 'for', whereas the constructions containing *de* leave room for ambiguous interpretation. While *musique pour piano* clearly designates music (a piece of music or composition) for piano, *musique de piano* may also refer to music played by a piano (and not necessarily composed for playing on a piano). In this sense, *pour* helps to resolve ambiguity and allows only the interpretation 'designed for'. For the Spanish data, the pattern is quite similar to the French data, as in the following examples:

(22) [N1 *de/para* $N2_{USER}$]$_N$
club/ropa de/para niños
'children's club/clothes'

[N1 *de/para* $N2_{PURPOSE}$]$_N$
alimentos de/para consumo
'consumer goods'

(23) [N1 *de/para* $N2_{USER(OBJECT)}$]$_N$
juego de/para pc
'PC game'

These cases show that variation between *de* and *para* is possible only in contexts in which N2 semantically represents a user of (a person or an object) or a specific purpose for N1. These are the same templates that were found for the French data above. This result contradicts the findings from López (1970), who indicates that variation of *de* and *para* is also possible in contexts in which N1 designates a container, as in *cesto de/para basura* 'waste bin/bin for waste'. In her corpus data of Argentinian Spanish from Buenos Aires, Pacagnini (2003) also finds constructions of the type *loción de/para limpieza* ('cleaning lotion') or *crema*

de/para hidración ('hydration crème'), in which the preposition expresses the utility of an object (Pacagnini 2003: 164). Furthermore, she describes examples of the type *lápiz de/para labios* 'lipstick' and *esmalte de/para uñas* 'nail polish', in which N1 represents an instrument. From this, Pacagnini deduces a schema in which, on a continuum between morphology and syntax, *de* lies closer to the morphological pole, whereas *para* is closer to the syntactic pole. In this paper, I can confirm Pacagnini's hypothesis that N Prep N constructions in Spanish show a certain internal variation in respect of the prepositions *de* and *para*, which might therefore be considered as lying at different points of a continuum between the morphological and the syntactic pole. In this case, it is evident that constructions with *para* are located closer to the syntactic pole than constructions with *de*. Pacagnini observes that 75 percent of the participants in her data used a determiner or a qualifying adjective with the preposition *para* in cases where N1 denotes an instrument, as in *loción de/para la limpieza* ('lotion for cleaning') or *esmalte para uñas sensibles* ('polish for sensitive nails') (Pacagnini 2003: 166). In the esTenTen corpus data, this type of variation between *de* and *para* does not occur at all. However, a closer look at the Portuguese data offers interesting findings:

(24) [N1 *de/para* $N2_{USER}$]$_N$
brinquedos de/para crianças
'children's toys'

[N1 *de/para* $N2_{PURPOSE}$]$_N$
acessórios de/para decoração
'accessories for decoration'

(25) [N1 *de/para* $N2_{USER\ (OBJECT)}$]$_N$
concerto de/para piano
'piano concert'

[N1 *de/para* $N2_{REASON}$]$_N$
cirurgia de/para correção
'reconstructive surgery'

(26) [N1 *de/para* $N2_{PERIOD}$]$_N$
aluguel de/para férias
'vacation rental'

[$N1_{INSTRUMENT}$ *de/para* N2]$_N$
produto de/para limpeza
'cleaning product'

(27) [N1 *de/para* $N2_{DETERMINATION}$]$_N$
animais de/para abate
'animals for slaughter'

The Portuguese data illustrate that variation of *de* and *para* is possible in a larger number of nominal semantic relations in Portuguese than in Spanish or French. On the one hand, Portuguese offers the same templates as French and

Spanish: N2 as a user (object or person) and N2 as a specific purpose of N1. On the other hand, Portuguese provides additional templates, including N2 as a specific time or period of time, and N2 designating a specific determination for N1 (which in most cases is a living being). One additional template, N1 is an instrument for N2, is of particular interest. Here, we find the Portuguese example *produto de/para limpeza* 'cleaning product', which Pacagnini cited for Argentinian Spanish. This template appears to be productive in Portuguese, as shown in the additional examples *creme de/para mãos* ('hand cream') and *máscara de/para cílios* ('mascara for eyelashes'). Although our Spanish data contradict Pacagnini's findings for Spanish, the same template can be found in Portuguese. Further investigation of this phenomenon is necessary, particularly in light of the possibility that Spanish used in Buenos Aires, where Pacagnini collected her data, may be influenced by Portuguese from Brazil. Initial informal speaker assessments of native Spanish speakers in Spain reveal that the template [$N1_{INSTRUMENT}$ *de/para* $N2]_N$ is not productive in Spain and that the template $[N1_{INSTRUMENT}$ *para* $N2]_N$ is considered incorrect.

The analysis of the variation between *de* and *para*, and *de* and *pour*, in Spanish, French, and Portuguese reveals that Portuguese has the largest number of templates at an intermediate lexical level for the variation of *de* and *pour/para*. The Spanish and French data overlap in their templates for the variation of *de* and *pour/para*, while the Spanish data from Buenos Aires (Pacagnini 2003) offers a slightly different picture. The analysis here supports the findings from the previous subsections on the semantic transparency of the constructions under investigation. The present analysis does not feature any (partially) opaque or (partially) idiomatic constructions. I mentioned at the beginning of this subsection that the prepositions *de* and *pour/para* vary in their semantic transparency; nevertheless, they undergo internal constituent variation in all three languages under investigation. While traditional accounts generally mention the different syntactic status of constructions containing *de* and *pour/para*, the constructionist approach introduced in section 4, makes possible an unproblematic mapping of this internal constituent variation.

6 Conclusion

The present study of internal constituent variation in N Prep N constructions allows numerous conclusions to be drawn as to their nature in Romance languages as well as on the role and variability of the prepositional element. The discussion and analysis here have shown that it is not always possible or expedient to differentiate clearly between lexical and syntactic N Prep N constructions. In many

cases, not even the numerous delimitation tests may lead to a clear distinction. Therefore, the present account has abandoned this strict, dichotomous distinction in favor of a more holistic approach. When considering internal constituent variability, the determining factor is not the lexical or syntactic status of the elements, instead, it is the nominal semantic relation expressed via the preposition. Here, it is not crucial to differentiate between the lexical status, as in *libro de niños* 'children's book' and the syntactic status, as in *libro para niños* 'children's book'. In order to conduct a fruitful qualitative comparative analysis of N Prep N constructions in Romance languages, it is necessary to adopt a theoretical accounts that does not focus on the lexicon-syntax distinction. In the present paper, construction morphology, a constructionist approach that expands the notion of construction to the word level, offers the appropriate tools for analysis. Following Masini (2009), N Prep N constructions are analyzed as abstract templates, which are, to some degree, productive and associated with a naming function (Masini 2009: 261). For the present analysis, a constructionist inheritance hierarchy has been adapted to internal constituent variation in one construction (see section 4). The latter analysis focused on the intermediate lexical level, that is, the alternation between [N1 Prep1 N2] and [N1 Prep2 N2], at which Prep1 and Prep2 designate alternative prepositions. This constructionist approach revealed the possible templates for prepositional variation in three different languages: Spanish, French, and Portuguese.

The analysis of three alternating pairs, specifically *de* and *à/a*, *de* and *en/em*, and *de* and *pour/para*, demonstrates important differences and common features between the languages. The quantitative aspect, which was not the primary focus of this paper, demonstrates the strong frequency and productivity of the different templates in Portuguese. This holds to a lesser extent in French and is even less in Spanish. This result is in line with the results from Hennecke & Baayen (2017). The qualitative analysis demonstrates that, in the underlying datasets, Portuguese offers the greatest number of different templates for internal prepositional variation, followed by French, and then Spanish. In this connection, it should be mentioned that Portuguese also offers the largest number of constructions (or types) for each template. This result confirms the impression from the quantitative study that Portuguese N Prep N templates are frequent in speech and very productive. From a qualitative perspective, it is striking that most templates of internal prepositional variation exist across languages. In the case of the pair *de* and *en/em*, the templates that allow internal prepositional variation vary only slightly between the languages. For variation between *de* and *à/a*, the French data show the greatest tendency to internal variation. This is mainly because the preposition *à* is relatively productive and frequent in French, which

is not the case for Spanish and Portuguese. In cases where French relies on the preposition *à*, Spanish and Portuguese mostly employ the preposition *de*, as in Fr. *verre à vin*, Sp. *copa de vino* and Pt. *copo de vinho* ('wine glass', in each case). Spanish does not offer any internal variation of *de* and *a*, whereas Portuguese shows certain tendencies in this direction. For the variation between *de* and *pour/para*, the French and Spanish data do not show any qualitative differences; they overlap exactly in terms of which templates allow internal prepositional variation. Studies based on data from Argentinian Spanish indicated the existence of further templates; these were not found in the present data in Spanish, but many of them were present in the Portuguese data.

A very important finding from the qualitative analysis is that internal prepositional variation in the three languages is possible only for semantically transparent constructions. This can be explained by the fact that in opaque N Prep N constructions, the semantic relation between the nominal constituents often cannot be determined explicitly.

In conclusion, a constructionist approach to N Prep N constructions may solve certain problems in defining and delimitating these constructions in Romance languages. Furthermore, a constructionist approach allows an accurate investigation of the differences and common features of templates for internal prepositional variation in the three languages under investigation here. Future studies should investigate these templates in more detail, extending the approach to other types of internal variation.

References

Anscombre, Jean-Claude. 1990. Pourquoi un moulin à vent n'est pas un ventilateur. *Langue française* 86. 103–125.

Bartning, Inge. 1993. La préposition 'de' et les interprétations possibles des syntagmes nominaux complexes: essai d'approche cognitive. In Anne-Marie Berthonneau & Pierre Cadiot (eds.), *Les prépositions: méthodes d'analyse*, vol. 11 (Lexique), 163–191. Lille: Presse Universitaires de Lille.

Bell, Melanie & Martin Schäfer. 2016. Modelling semantic transparency. *Morphology* (26). 157–199.

Benveniste, Émile. 1974. Formes nouvelles de la composition nominale. In Émile Benveniste (ed.), *Problèmes de linguistique générale*, vol. 2 (Bibliothèque des sciences humaines), 163–176. Paris: Gallimard.

Booij, Geert. 2010. Compound construction: schemas or analogy? a construction morpgology perspective. In Sergio Scalise & Irene Vogel (eds.), *Cross-disciplinary issues in compounding*, vol. 311 (Current Issues in Linguistic Theory), 93–109.

Booij, Geert. 2015. Word-formation in construction grammar. In O. Peter Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word formation*, vol. 1, 188–202. Berlin & Boston: Mouton de Gruyter.

Bosredon, Bernard & Irène Tamba. 1991. Verre à pied, moule à gaufres: préposition et noms composés de sous-classe. *Langue française* (91). 40–55.

Buenafuentes de la Mata, Cristina. 2010. *La composición sintagmática en español*. Vol. 7 (Instituto histororia de la lengua. Monografías). España: Cilengua.

Cadiot, Pierre. 1997. *Les prépositions abstraites en français* (Linguistique). Paris: Armand Colin/ Masson.

de Bustos Gisbert, Eugenio. 1986. *La composición nominal en español*. Salamanca: Universisad de Salamanca.

Di Sciullo, Anne-Marie & Edwin Williams. 1987. *On the definition of word*. Vol. 14 (Linguistic inquiry. Monographs). Cambridge, MA.

Fradin, Bernhard. 2009. Romance: French. In Rochelle Lieber & Pavol Štekauer (eds.), *The oxford handbook of compounding*, 417–435. Oxford University Press.

Gagné, Christina & Edward Shoben. 1997. The influence of thematic relations on the comprehension of modifier-noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition* (23). 71–87.

Gagné, Christina & Thomas Spalding. 2009. Constituent integration during the processing of compound words: does it involve the use of relational structures? *Journal of Memory and Language* (60). 20–35.

Girju, Roxana, Dan Moldovan, Marta Tatu & Daniel Antohe. 2005. On the semantics of noun compounds. *Journal of Computer Speech and Language – Special Issue on Multiword Expressions,* 19(4). 479–496.

Goldberg, Adele. 1995. *Constructions: 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: Oxford University Press.

Guevara, Emiliano. 2012. Spanish compounds. *Probus. International Journal of Latin and Romance Linguistics* 24(1). 175–195.

Hennecke, Inga & R. Harald Baayen. 2017. A quantitative survey of N Prep N constructions in romance languages and prepositional variability. *Quaderns de Filologia: Estudis Lingüístics* 22. 129–146.

Kornfeld, Laura Malena. 2009. Romance: Spanish. In Rochelle Lieber & Pavol Štekauer (eds.), *The oxford handbook of compounding*, 436–453. Oxford University Press.

Lang, Jürgen. 1991. *Die französische Präposition: Funktion und Bedeutung*. Vol. 95 (Beiträge zur Literatur-, Sprach- und Medienwissenschaft). Heidelberg: Carl Winter Universitätsverlag.

Laumann, Ferdinand. 1998. *Französische Nominalkomposita mit Präposition: eine Untersuchung aufgrund der naturwissenschaftlich-technischen Fachsprache*. Vol. 5 (Studien zur Linguisitk). Münster: Lit.

López, Maria Luisa. 1970. *Problemas y métodos en el análisis de preposiciones*. Madrid: Gredos.

Masini, Francesca. 2009. Phrasal lexemes, compounds and phrases: a construcionist perspective. *Word Structure* 2(2). 254–271.

Masini, Francesca & Anna Thornton. 2007. Italian VeV lexical constructions. In Geert Booij, Angela Ralli & Sergio Scalise (eds.), *Morphology and dialectology*, 148–189.

Pacagnini, Ana Maria Judith. 2003. Compuestos sintagmáticos y alternancia preposicional. *Moenia* 9. 159–172.

Rainer, Franz. 2016. Italian. In O. Peter Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word formation*, vol. 4, 2712–2731. Berlin & Boston: Mouton de Gruyter.

Rio-Torto, Graça & Sílvia Ribeiro. 2009. Compounds in Portuguese. *Lingue e Linguaggio* 8(2). 271–291.

Rio-Torto, Graça & Sílvia Ribeiro. 2012. Portuguese compounds. *Probus. International Journal of Latin and Romance Linguistics* 24(1). 119–145.

Val Àlvaro, José Francisco. 1999. La composición. In Violeta Demonte & Ignacio Bosque (eds.), *Gramática descriptiva de la lengua espanola*, vol. 3, 4747–4842. Barcelona: Espasa Calpe.

Villoing, Florence. 2012. French compounds. *Probus. International Journal of Latin and Romance Linguistics* 24(1). 29–60.

Zwitserlood, Pienie. 1994. The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes* 9. 341–368.

Chapter 5

Production of multiword referential phrases: Inclusion of over-specifying information and a preference for modifier-noun phrases

Christina L. Gagné
Department of Psychology, University of Alberta

Thomas L. Spalding
Department of Psychology, University of Alberta

J. Claire Burry
Department of Psychology, University of Alberta

Jessica Tellis Adams
KidsAbility Centre for Child Development, Ontario

We examined the underlying psycholinguistic and cognitive factors that give rise to the production of multiword expressions. For example, if a story describes a woman buying a dog with blue fur, will people include the color of the dog when referring to the animal and, if so, in what syntactic form? In the experiment, participants read short stories that contained a concept that was presented as either a modifier-noun phrase (e.g., the blue dog) or full phrase (e.g., the dog that was blue). We also varied whether the property being highlighted was normal (e.g., brown) or distinctive (e.g., blue) for the head noun concept (e.g., dog). We found that participants are more likely to include distinctive properties than normal properties when referring to the concept. Although the selection of a syntactic form was partially influenced by the form of the information in the story, there was a strong overall bias toward using a modifier-noun phrase structure.

Christina L. Gagné, Thomas L. Spalding, J. Claire Burry & Jessica Tellis Adams. 2019. Production of multiword referential phrases: Inclusion of over-specifying information and a preference for modifier-noun phrases. In Sabine Schulte im Walde & Eva Smolka (eds.), *The role of constituents in multiword expressions: An interdisciplinary, cross-lingual perspective*, 123–141. Berlin: Language Science Press. DOI:??

Christina L. Gagné, Thomas L. Spalding, J. Claire Burry & Jessica Tellis Adams

1 Introduction

1.1 Aim and background

The aim of this chapter is to explore when and how multiword expressions are used within a referential context. In particular, we focus on production of referential expressions and examine what drives the inclusion of modifying information and the syntactic form of the expression. When referring to an object, person, or event, a speaker/writer is faced with the challenge of assigning linguistic labels to conceptual entities; often, several linguistic expressions can be used. For example, the same object can be referred to as "cup", "ceramic cup", or "cup that is made of ceramic". What influences this decision? Two aspects of forming a referential expression are particularly relevant and will be the focus of our investigation. First, the speaker/writer might or might not include modifying information in the referential expression. Second, if modifying information is included, the expression might be a compound (e.g., ceramic cup) or a full noun phrase (e.g., cup that is made of ceramic). Although it is tempting to think of these as two separate ordered decisions (first decide whether or not to modify, then decide the form of the modification) we should note that these two aspects are not necessarily deliberate, conscious choices, nor need they be, strictly speaking, independent or sequential. Rather, the ultimate form of the expression may reflect underlying cognitive processes carried out within the language system that, working together, give rise to the form of the expression, and hence to both the syntax and the presence (or not) of modifying information.

Much of the existing work on compounds and modifier-noun phrases has focused on compound access and interpretation. The current study takes a different approach to this problem. Rather than focusing on the interpretation per se, we examine production to identify some of the expectations and biases that human users have about the use of modifying information during referential communication. When using referential expressions, speakers/writers attempt to establish both semantic co-ordination and lexical co-ordination with the addressee (e.g., Clark & Wilkes-Gibbs 1986; Garrod & Anderson 1987; Clark & Schaefer 1989). An attempt is made to synchronize the underlying mental model of the current situation as well as the specific expressions that are applied to particular entities within that model. In doing so, the speaker/writer draws on many different types of knowledge, including world knowledge, knowledge about information expressed in the conversation/discourse, and knowledge about linguistic conventions. Identifying the expectations that people have about the use of multiword expressions provides insight into how people are conceptualizing both the entities denoted by these constructions and the scenarios in which the constructions

are or should be used. Consequently, this area of research has implications for a variety of areas within the psycholinguistic and linguistic literature. In particular, the current project contributes to research that examines the contribution of the individual constituents to the understanding of the meaning of the whole expression, and the appropriateness of the use of the whole construction in a given situation.

The semantic transparency of the constituents of a compound has been a widely studied aspect of compound processing (Libben 1998; Jarema et al. 1999; Gagné & Spalding 2016; Smolka & Libben 2017). In general, compounds with opaque constituents (e.g., *humbug*) are more difficult to process than compounds with transparent constituents (e.g., *schoolyard*). Of course, in creating a multiword referential phrase that is new (as opposed to a known compound word, for example), the constituents will need to be relatively transparent in order to provide the information that would allow the communicative task to be successfully completed. However, at any given level of transparency there are other aspects that will influence whether a head noun is modified. In the current chapter, we will consider one of these factors, namely the distinctiveness of the property denoted by the modifier, which, like semantic transparency, is a semantic factor. Both *blue dog* and *brown dog* are semantically transparent expressions in that the meaning of the constituents contribute to the meaning of the whole. However, *blue dogs* are more distinctive compared to the concept *dog* than are *brown dogs*. We explore whether people are sensitive to the distinctiveness of a property during the formation of multiword expressions.

1.2 Overview of chapter

In this chapter, we begin by providing an overview of the theoretical issues concerning the inclusion of modifying information and the use of either full phrases or modifier-noun phrases. Next, we present an experiment in which we manipulated two factors that might influence the production of referential expressions. In particular, we examined whether the distinctiveness of the modifying information influences whether that information is used when referring to the antecedent. In addition, we examined whether the syntactic form in which the modifying information is presented influences the form in which modifying information is conveyed. Finally, we discuss the relevance of the empirical data within a psycholinguistic context and highlight the implications of the data for multiword expressions and for modifier-noun phrases in particular.

1.3 What motivates the inclusion of modifying information?

The expressions used to denote referents reflect how the speaker/writer is conceptualizing the object and, in particular, how he/she chooses to distinguish it from other items (Brown 1958; Olson 1970). Indeed, speakers are sensitive to both nonlinguistic- and linguistic-ambiguity during referential communication and attempt to avoid producing ambiguous expressions (Ferreira et al. 2005). A key issue for the current research concerns the factors that lead people to include modifying information rather than using an unmodified noun when producing a referential expression. The inclusion of modifying information serves several linguistic and psychological functions. Most often, modifying information is used to distinguish among potential referents (Downing 1977; Brekle 1986). There are often situations in which using the category label alone would not be sufficient. Consider a situation in which there are several cups on a table. To refer to a particular cup, for example, a speaker might specify its material and use either a full noun phrase (e.g., May I have the cup that is ceramic?) or a compound (e.g., May I have the ceramic cup?). Both utterances involve combining information about the head noun concept (e.g., cup) with information about a modifying concept (e.g., ceramic). This combination of information, in turn, allows the unambiguous identification of the referent within the available set of potential referents.

Several experiments on referential communication that used a visual display of objects (Tanenhaus et al. 1995; see also Frank & Goodman 2012) have found that speakers use a pre-nominal adjective (e.g., tall glass) in a context in which there are contrasting members (e.g., a short glass), which is consistent with the hypothesis that speakers try to make their utterances as informative yet as economical as possible (Grice 1975). The pre-nominal adjective is used to uniquely identify one object among several objects. However, the motivation for using modifying information appears to go beyond merely disambiguating among multiple possible referents because it is often included even when there is no need to provide additional information. This phenomenon of providing modifying information even in cases where such information is not needed to identify the referent is known as over-specification. Indeed, there are a number of studies showing that participants include adjectives during referential communication even though this additional specification is not required to identify the referent (e.g. Pechmann 1989; Sedivy 2003; Maes et al. 2004; Koolen et al. 2013).

Over-specification performs various functions in addition to identifying referents. For example, modifying information (e.g., the cup that is on the shelf near the plate) is used to shift the addressee's focus of attention (Ariel 1990; Prince 1992; Gundel et al. 1993; Chafe 1994). Another reason for using modifying infor-

mation is to conform to pre-established conversational pacts (Brennan & Clark 1996; Ibarra & Tanenhaus 2016). Conversational partners often converge on an expression and will persist in using that expression even when there is no longer a need to include the additional information. To use an example from Brennan & Clark (1996), the term *pennyloafer* was initially used to denote a particular shoe among other possible shoes. However, the speaker continued to use this term rather than switching to using the simpler term *shoe* even when no other shoes were present in the display.

From a cognitive processing perspective, over-specification appears beneficial to both the speaker and the listener. For example, it aids in the identification of objects in a visual array and, consequently, speakers are more likely to produce over-specified expressions when they were asked to imagine that the task was very important (i.e., when told to imagine that the control panel is being used for long-distance surgery) than when they were not given such a scenario (Arts et al. 2011). Over-specification also benefits production (Pechmann 1989). Consistent with this idea, redundant information is more likely to be included when the speaker is under time pressure. Koolen et al. (2016) conducted a study in which participants referred to target objects in a visual array of objects. Participants were more likely to use over-specifying information when they were under a time constraint (e.g., they had to respond within 1000 ms) than when they had an unlimited amount of time to refer to the target object. Koolen et al. (2016) concluded that when individuals are under pressure, they are more likely to use quick heuristics and therefore select properties of an object based on their perceptual salience rather than discriminatory power.

Overall, there appear to be many reasons for why speakers might choose to include modifying information in referential expressions. In the current experiment, we focus on additional usage of modifying information that has not been fully explored in the literature. In particular, we propose that modifying information might be used to mark a conceptual distinction among category members and, in particular, to make explicit note of particularly distinctive information.

Studies on referential expressions within a visual context (i.e., situations in which objects are presented visually) indicate that the distinctiveness of visual properties within the display influences referential expressions. Participants were more likely to provide modifying information (i.e., to produce over-specified expressions) when the property of an object is atypical (e.g., Westerbeek et al. 2015). For instance, Rubio-Fernández (2016) used a referential communication task in which participants asked the researcher to click on objects that were presented in an array on the computer screen. In the first experiment, participants saw pictures of paper dolls and a display of paper clothes that were either all the same

color (e.g., brown purse, brown shirt, brown dress, and brown shoes) or different colors (e.g., yellow purse, pink shoes, blue dress, and red pants). In the second experiment, participants saw arrays of animals, fruits, vegetables, and artifacts that either had typical colors (e.g., brown camel) or atypical colors (e.g., blue camel). Participants tended to use a redundant color adjective in instances where such modifying information would be unnecessary (e.g., *"the blue dress"*, where only one dress could be a possible referent) more often when the object was an atypical color than when the object was a stereotypical color. These results suggest that modifying information is used when the concept has been modified with a distinctive property. Furthermore, participants provided modifying information more often when the color was a central property of the object category (e.g., referents such as clothing yielded a higher usage of redundant color adjectives than did geometrical figures). Taken together, these results suggest that a key characteristic in terms of determining whether modifying information is provided is conceptual distinctiveness rather than perceptual/visual distinctiveness. That is, the distinctiveness of the information relative to the category itself, rather than just within the visual display.

The aim of the current study is to explore the role of conceptual distinctiveness by examining whether the tendency to mention distinctive properties extends to situations in which the objects are not physically present. In particular, we will focus on a situation in which the contrast with other category members is implied or based on conceptual knowledge within a story context, rather than presenting the objects in a visual display. For example, mentioning that flowers are either fresh or wilted implicitly contrasts the flowers with ones that are not fresh or not wilted. Moreover, in the context of buying flowers as a gift, it is more typical to buy ones that are fresh than ones that are wilted. Thus, from a conceptual perspective, the property *wilted* is more distinctive for flowers than is *fresh*.

Conceptual distinctiveness is related to the issue of contrast. The notion of contrast between categories and subcategories has long played an important role in linguistic and psycholinguistic theories. Indeed, the principles of contrast and mutual exclusivity (Clark 1983; Carstairs-McCarthy 2010) are well-known constraints on word learning. In terms of multiword expressions, previous research on conceptual combination suggests that the notion of contrast influences how people use noun phrases. For example, Gagné & Murphy (1996) found that when verifying whether a property is true of a modifier-phrase (e.g., *submarine door*), people took less time to verify a property that was true of the phrase but not generally true of the head noun (e.g., *made of metal*) than to verify a property that was true of both the phrase and the head noun (e.g., *solid*). This finding suggests that people are sensitive to the extent to which the modified concept (e.g.,

submarine door) is semantically/conceptually distinctive from other members of the head noun concept (e.g., *door*).

In terms of judgments about whether a concept has a particular property, several studies (Connolly et al. 2007; Gagné & Spalding 2011; Hampton et al. 2011; Jönsson & Hampton 2012; Gagné & Spalding 2014b) have shown that properties that are true of the head noun (e.g., *kites have strings*) are viewed as being less true of the modified head (e.g., *silk kites have strings*). This effect (known as the modification effect) appears to be driven by the expected level of contrast between the combined concept (e.g., *silk kites*) and the head concept (e.g., *kites*); when making judgments about the likelihood that a property is true, participants are influenced by the meta-knowledge that modified concepts are used to signal that the subcategory is similar to the category (e.g., silk kites have many properties in common with kites) but also that the subcategory is somehow different than the category (Gagné & Spalding 2011; 2014b; Spalding & Gagné 2015). These two expectations account for why properties that are true of the head noun are judged as being less true of the modified concept, and that properties that are false of the head noun (e.g., *candles have teeth*) are judged to be more true (but still unlikely) of the modified concept (e.g., *purple candles have teeth*). Indeed, the effects of the expected contrast is so strong that the same effects are seen even when the modifier is a non-word (e.g., Gagné & Spalding 2015).

Thus, we conclude that conceptual contrast or conceptual distinctiveness is a critical factor in the use and understanding of multiword phrases and compound words in general and is therefore likely to contribute to the production of such phrases.

1.4 When modifying information is included, how is it expressed?

If modifying information is included, the syntactic form which expresses this information can still vary. In English, modifying information can be expressed as a full noun phrase (e.g., a dog that is blue) or as a modifier-noun phrase (e.g., a blue dog). Do people have a priori biases toward using one linguistic expression over another? The answer is not immediately obvious because intuitions based on ease of processing do not correspond with the tendency for expressions to become shortened over time.

In terms of ease of processing, there is an advantage to using a full phrase because noun compounds are particularly challenging to interpret (Lapata 2002; Copestake & Briscoe 2005; Libben 2014). Much of the difficulty lies in recovery of an implicit underlying relation between the modifier and head noun concept. A modifier-noun phrase is more ambiguous than a full phrase, in that the

full noun phrase explicitly describes the exact nature of the modification that is being performed (e.g., *oil for babies*) whereas, for modifier-noun phrases (e.g., *baby oil*) the nature of the modification is implicit and must be reconstructed by the listener/reader (see Levi 1978; Gagné & Shoben 1997). The term "modifier-noun phrase" most often refers to constructions that are novel (e.g., *apple juice seat*; *mountain magazine*), but, can also refer to lexicalized open (unspaced) compounds (e.g., *hunting dog*; *paper bag*). Indeed there seems to be commonalities in the processing of novel noun phrases and lexicalized compounds (Gagné & Spalding 2006). Psycholinguistic research has shown that human language users actively make use of relations during the processing of both novel and established/lexicalized compounds (Gagné & Shoben 1997; Gagné 2002; Gagné & Spalding 2009; 2014a). This research indicates that, during the comprehension of noun compounds, the more available the required relation is, the easier it is to select the relation and, consequently, the less time it takes to interpret the compound. In other words, the more difficult it is to recover the implicit underlying relation, the more difficult it is to interpret a compound (see, for example, Gagné & Shoben 1997; Spalding & Gagné 2014; Schmidtke et al. 2018).

Given the difficulty inherent in recovering implicit semantic relations, one would presume that it would be advantageous to overtly express the relation and, consequently, to avoid the use of compounds. Yet, this is not what happens within the human language system. Over time, lexicalized phrases are often truncated and become compounds (e.g., *our lady's bug* became *ladybug*). Similarly, compounds can become non-compounds (e.g., *electronic mail* became *e-mail* and, more recently, *email*); the words *lord* and *lady* are derived from Old English compounds *half-weard* "bread-keeper" and *halfdige* "bread-kneader". This truncation that occurs on a global (and more long-term) level within a language also occurs during local interactions. During referential communication, for example, linguistic expressions are often shortened (Garrod & Anderson 1987; Brennan & Clark 1996). For example, in one experiment, a geometric figure that was initially described as looking "…like a person who's ice skating, except they're sticking two arms out in front" became "the ice skater" (Clark & Wilkes-Gibbs 1986). Similarly, an object that was initially referred to as "the car that has like …blueprints painted on the side of it sorta" was later referred to as "the blueprint car" (Metzing & Brennan 2003). In sum, there appears to be a preference toward using syntactically simpler expressions such as compounds, even though such expressions are inherently more ambiguous than full expressions which specify the relation overtly.

On the basis of these findings, one would expect an overall bias towards using a truncated expression (e.g., using *wilted flowers* or *even flowers*, rather than

flowers that are wilted). However, this bias must also be considered in light of another bias reported in the literature – namely, the tendency for people to re-use recently encountered syntactic structures. For example, Bock (1986) demonstrated that speakers tend to re-use a syntactic structure from the priming sentence when describing a scene. This effect has been examined in a variety of context including examinations of whether it can be driven by a single word as in the case of featural accounts of syntactic priming. For example, Melinger & Dobel (2005) found that production preferences for dative alternation can be biased by prior exposure to a single verb. However, most relevant for the current project concerns studies that focus on the creation of referential expressions. Syntactic convergence occurs during referential communication. For example, participants were more likely to describe a picture of a red sheep as "The sheep that's red" when the confederate recently described a picture of a red door as "The door that's red" than when it was described as "a red door" (Cleland & Pickering 2003). This result suggests that participants tend to re-use syntactic structures, especially when the prime and target sentences share lexical items such as *red* (see also Chang et al. 2003). Similarly, Tarenskeen et al. (2015) found that when participants use modifying information to describe a target item from a visual array of six drawings of clothing, there is a tendency to continue to re-use the same syntactic structures.

These studies all demonstrate that participants have a tendency to re-apply the same syntactic structure that was used with one object/entity (e.g., *sheep*) when subsequently referring to a separate object/entity (e.g., *door*). However, an unresolved question concerns whether syntactic priming will occur in a task in which participants are introduced to a concept (e.g., *apples that are rotten*) and then are asked a question requiring them to refer to that same concept. This situation directly pits the bias towards truncation against the bias towards re-using syntactic expressions. The current experiment will investigate this issue.

2 Experiment

We examine the types of referring expressions that people produce when referring to a concept that has been encountered in a short description of a scenario. The experiment was designed to address two key issues: 1) whether the distinctiveness of the modifying information being conveyed about a target entity in a story influences whether that information is included when the participant is asked to refer back to the entity and 2) whether the syntactic form in which the modifying information is presented influences the form in which modifying in-

formation is conveyed. Participants read short stories and then answered a comprehension question that would require them to refer to something in the story. For example, one story described a woman buying a pet. The target antecedent was the dog that she purchased. We varied the type of modifying information that was presented with the target antecedent. The information was either normal or typical for the object or was distinctive. To illustrate, all participants read a version of the story in which the color of the dog was mentioned. For half of the participants, the dog was described as having brown fur (a normal feature for dogs), and for the other half, the dog was described as having blue fur (a distinctive feature for dogs). We were interested in what the participants would produce when they were asked *What kind of pet did Sally buy?*

We predict that distinctiveness will influence whether participants choose to include modifying information in their linguistic expression. Properties that are unusual or distinctive for the head noun will be seen as especially relevant and, consequently, will be more likely to be included in the description provided by the participants. However, properties that are not unusual will be deemed less relevant (because the majority of members of the head noun category have the same property) and therefore less likely to be included. Thus, when referring to a dog that was previously mentioned in a short story, participants will be more likely to include modifying information when the dog was described as having an atypical color such as *blue* relative to when the dog was described as having a typical color such as *brown*, because the resulting subcategory is more distinctive and therefore will tend to more readily identify the appropriate referent. In short, there are lots of brown dogs, but relatively few blue dogs in the world, and, consequently, it should be more informative to refer to the subcategory of *blue dogs* than to the subcategory of *brown dogs*. Note, however, that in no case is the modifying information required to uniquely identify the referent.

In terms of the syntactic form that is used to convey the modifying information, the existing literature points to two conflicting predictions. On one hand, people might show a tendency toward using a modifier-noun phrase even when the information is presented as a full noun phrase. Two considerations arise here. First, the modifier-noun phrase is shorter and syntactically simpler and, thus, might generally be preferred. Second, a modifier-noun phrase is more ambiguous than a full phrase, in that the full noun phrase explicitly describes the exact nature of the modification that is being performed (e.g., a dog that is blue) whereas, for modifier-noun phrases the nature of the modification is implicit and must be reconstructed by the listener/reader (Downing 1977; Levi 1978). Having the relation directly specified (e.g., *crayon that is made of plastic*, or *sunshine in the morning*) removes uncertainty about relation selection (Gagné & Spalding 2014a;

2015). Thus, there could be some trade-off, in which speakers or writers generally prefer to use the shorter modifier-noun phrase, as long as they have reason to believe that the recipient will understand the implied connection between the modifier and the head noun concepts. Gagné & Spalding (2004) found that the presence of a referent in a discourse made modifier-noun phrases easier to comprehend, even though the phrase itself had not been presented. In the present study, all of the stories include information (either in the form of the full noun phrase or the modifier-noun phrase) that should make it easy for a recipient to understand the modifier-noun phrase. Therefore, the participants, in responding to the question about the target antecedent, might show a general preference for the modifier-noun phrase.

On the other hand, the form in which the information was initially presented in the preceding discourse might influence the manner in which the information is later conveyed due to syntactic priming. That is, when information is presented as a modifier-noun phrase, then people should be more likely to produce a modifier-noun phrase than when the information is presented as a full noun phrase. This prediction is derived from research on the activation of syntactic structure during speech production that demonstrates that speakers tended to reuse a syntactic structure from the priming sentence when describing a scene (Bock 1986; Bock & Loebell 1990).

2.1 Method

2.1.1 Participants

Fifty-four introductory psychology students participated for partial course credit. All participants were native speakers of English. The data from two participants were not used because they did not follow instructions. Thus, data from 52 participants were included in the analyses.

2.1.2 Materials and procedure

Twenty-eight short stories were constructed. Each story was under 65 words long and contained a target antecedent (i.e., the antecedent that we will be eliciting) for which we provided modifying information. We varied whether the modifying information was distinctive (e.g., *blue fur*) or usual (e.g., *brown fur*) for the head noun (e.g., *dog*) in the context of the story. In addition, we varied the syntactic form in which the modifying information was presented: the information was presented as a modifier-noun phrase (e.g., *brown dog; blue dog*) or full noun phrase (e.g., *a dog that is brown; a dog that is blue*). These two variables were

crossed which yielded four experimental conditions. For example, one story was: "Sally loves animals. She decided to get a pet. So she went to the pet store to see what was there. Sally immediately set her eyes on a [*blue dog/brown dog/dog that was blue/dog that was brown*]. She picked him up and knew instantly that he was going to be a great companion for her." Only one of the expressions within the square brackets was presented to a particular participant. The items were counter-balanced such that each participant saw an equal number of stories in each of the four conditions and each item was seen only once by each participant. Order of presentation was randomized for each participant. The full list of target items (i.e., the unusual, normal, and head noun) is listed in the Appendix.

Participants viewed the stories one at a time on a computer screen. They were instructed to read each passage carefully and were allowed as much time as necessary to complete the task. After each story, participants answered two questions about the story. The first question required people to recall the referent of the target noun phrase from the story. It specifically required the participant to respond by describing the target concept. For example, a question might ask *What kind of pet did Sally get?* The participant typed in their answer. The second question was also associated with the passage, and asked about another aspect of the story.

2.2 Results

Two of the authors classified the responses into four categories based on how the participants referred to the antecedent: modifier-noun phrase (e.g., *blue dog* or *brown dog*), full phrase (*dog that is blue* or *dog that is brown*), and head noun only (*dog*). In addition, a fourth category was used for "other" responses. Three main types of responses fell under this category. The first were responses that did not provide a specific answer (e.g., "I don't know", "it doesn't say"). The second were responses that did not address the question (e.g., "What does Nathan cut quickly" was intended to elicit either green or yellow grass, but the participant responded "because his parents are coming home"). The third type of response did not directly refer to the target reference (e.g., "What does Katie wear to keep her feet warm" was intended to elicit snake slippers or soft slippers but the participant responded "fuzzy slippers").

Inter-rater agreement was 100%. Table 1 displays the number of responses (for each condition) in each category. Overall, participants generally did include modifying information; modifying information was provided in 962 out of 1456 responses, and the vast majority (84%) of these responses were in the form of

a modifier-noun phrase. The responses that were coded as Other were not included in further analyses and, thus, the percentage with which a category was used within each of the four experimental conditions was calculated based only on responses in the form of a modifier-noun phrase, full phrase, and head noun only.

Table 1: Number of responses and row percentages (in parentheses) for each condition that were modifier-noun phrase, full phrase, head noun only, or other. Each row sums to 364.

Experimental Condition		Response Type			
Property	Form	modifier NP	full phrase	noun	other
non-distintive	modifier-noun phrase	220 (60.44)	5 (1.37)	91 (25.00)	48 (13.19)
non-distintive	full phrase	111 (30.49)	48 (13.19)	157 (43.13)	48 (13.19)
distintive	modifier-noun phrase	303 (83.24)	2 (0.55)	37 (10.16)	22 (6.04)
distintive	full phrase	177 (48.63)	96 (26.37)	59 (16.21)	32 (8.79)
	Total	811 (55.70)	151 (10.37)	344 (23.63)	150 (10.30)

We conducted two separate analyses. The first analysis focused on whether Form and Distinctiveness affected the likelihood of including modifying information. The second analysis examined whether Form and Distinctiveness influenced the form (e.g., full phrase vs. modifier- noun phrase) in which the modifying information was conveyed. In both analyses, the dependent variable was binary (i.e., is modified vs. not modified for the first analysis, and compound vs. phrase for the second analysis) and, consequently, we used the *melogit* function in Stata 15 to fit a mixed-effects model for binary responses. The experimental variables, Form and Distinctiveness, were included as fixed effects, and subjects and items were included as crossed random effects. The estimates of the fixed effects are reported as log odds.

To examine whether the syntactic form (e.g., *wilted flowers* vs. *flowers that are wilted*) in which the information had been presented in the story and the distinctiveness of the property influenced the likelihood of including modifying information when referring to the antecedent, we fit a model in which the dependent variable was whether the participant's response included modifying information; modifier-noun phrase and full phrase responses were coded as 1 and the head noun only responses were coded as 0. Both the distinctiveness of the property and the form in which the information was presented in the story influenced whether modifying information was included in the response. Participants were more likely to provide modifying information when the property presented in the story

was distinctive (e.g., *wilted* as a property of *flowers*) rather than usual (e.g., *fresh* as a property of *flowers*), 86% vs. 61%, $b = 1.48, \text{SE} = 0.24, z = 6.22, p < 0.0001$, and when the property had been presented as a modifier-noun phrase rather than a full-phrase (81% vs. 67%), $b = -1.15, \text{SE} = 0.19, z = -5.96, p < 0.0001$. The two predictor variables (Form and Distinctiveness) did not interact with each other, $b = 0.46, \text{SE} = 0.31, z = 1.48, p = 0.14$.

The second analysis was conducted using only the responses that included modifying information (i.e., only the full phrase and modifier-noun responses) so that we could test whether the form in which the modifying information was presented in the story and the distinctiveness of the property influenced the way in which participants conveyed the modifying information in their response. The dependent variable corresponded to whether the response was a modifier-noun phrase (1 = modifier-noun phrase and 0 = full phrase). Participants were more likely to provide a modifier-noun response when the story used a modifier-noun form (predicted $M = 0.99, \text{SE} = 0.009s$) than when the story used a full phrase form ($M = 0.64, \text{SE} = 0.05$), $\chi^2(1) = 31.47, p < 0.0001$. Note that because there are only two levels of the variable, the reverse is also true: namely, that participants are more likely to provide a full-phrase response when the story used a full-phrase form than when the story used a modifier-noun form. The type of property used in the story did not strongly influence whether participants used a modifier-noun form, $\chi^2(1) = 3.05, p < 0.08$.

Distinctiveness and Form interacted, $b = -1.99, \text{SE} = 0.58, z = -3.41, p = 0.001$, and, therefore, we examined the simple effects at each level of form. Distinctiveness of the property had no effect on whether the response was a full phrase or modifier-noun phrase when the modifying information was presented as a full phrase, $\chi^2(1) = 2.43, p < 0.12$. However, when the modifying information was presented as a modifier-noun phrase, the response was more likely to be a modifier-noun phrase when the property was unusual/distinctive than when the property was normal, $\chi^2(1) = 8.91, p < 0.003$.

3 Discussion

We explored two aspects of the production of multiword referential expressions: inclusion of modifying information and syntactic form, with a particular focus on modifier-noun phrases (e.g., *blue dog* and *brown dog*) and full noun phrases (e.g., *dog that is blue* and *dog that is brown*). The experiment directly pitted the bias towards truncation against the bias towards re-using syntactic expressions. The findings make three primary contributions to the literature on multiword expressions. First, we demonstrate the influence of semantic/conceptual knowledge on

the inclusion of modifying information. In particular, the degree of conceptual contrast seems to be critical in determining whether modifying information is included when the referential expression is produced. Second, our results reveal the primacy of modifier-noun phrase constructions (over full phrase constructions) as a means of conveying that information. Third, while it is possible that there are small effects of syntactic repetition, or a general bias to use shorter syntactic forms for a reference to an already identified object from the story, the bias towards the modifier-noun phrase appears to be the main driver of the syntactic form of the referential expression, at least in this particular communicative task.

3.1 Including modifying information

Previous research using visual displays of objects found that over-specification was more likely when a property was visually distinctive or salient such as when one object was a different color than other objects in the display (e.g., in a visual display in which one dog is blue and the others are orange). The current results extend this finding to a situation where the objects are not physically present and the distinctiveness of a property is based on conceptual knowledge about the modifier and head noun concepts. For example, *blue* is distinctive for *dogs* but not for *skies*. The knowledge needed to determine distinctiveness comes from past history and knowledge of the concepts involved rather than from visual information that is presented in the experiment. Therefore, our finding suggests that people are sensitive to conceptual distinctiveness in addition to (as shown in previous research) referential distinctiveness. To illustrate, in general language usage, a category name (e.g., dog) typically refers to a generic type (i.e., to the category of dogs). However, in our study, the referent was always a particular category member, not a generic category. Whether participants used a generic label or modified construction depended on the distinctiveness of the property (relative to the head noun category) used in the story. In this respect, our data highlights the role of a particular type of implicit information, namely knowledge about the nature of the category-subcategory similarity. In particular, the category label (i.e., dog) was used when the particular referent in the story was not unusual; that is, when the entity being described was similar to the generic representative of the category. Note that the modifying information was not required to uniquely identify the referent (i.e., there was only one dog in the story), yet participants often opted to include this information, especially when it was distinctive. Thus, the inclusion of modifying information corresponded to a conceptual distinction rather than a purely referential one in that participants were

sensitive to semantic and conceptual knowledge about the category to which the referent belonged.

There are several possibilities for why participants tended to provide over-specified expressions especially when the referent had a distinctive property than when it had a normal property. One possibility is that the distinctive properties are just much more salient. For example, work on memory has suggested that features that violate expectations are often noticed and remembered particularly well (e.g., a skull in an office setting, see Brewer & Treyens 1981). In general, people make note of properties that are not similar to those they have seen before and, when communicating, they might prefer to explain these differences to others in the simplest way possible (Garrod & Anderson 1987; Markman et al. 1997). In the current experiment, the distinctive properties might have been more noticeable than normal properties, and this difference might have prompted participants to include them in their response. Another possible explanation is that the distinctive features are more likely to be incorporated into the representation of the target referent because they tend not to be true of the head noun. This explanation is consistent with previous research on novel combined concepts that suggests that features that are true of the entire phrase but not of the head noun in general (e.g., *white* for *peeled apples*) are more available than features that are true of the head noun (e.g., *round*) (Springer & Murphy 1992; Gagné & Murphy 1996) and also with evidence suggesting that people strongly expect property differences between things named with modified and unmodified nouns (Gagné & Spalding 2011; 2014b; Spalding & Gagné 2015). In our experiment, the normal properties were ones that tended to be true of the head noun concept, whereas the distinctive properties were not generally true of the head noun concept. Thus, it is possible that the distinctive property was more likely to be integrated into the representation of the target referent than was the normal property. If so, then distinctive properties would be more likely to be included in the participant's response than would normal properties.

3.2 Selection of syntactic form

There is some tendency to reproduce the syntactic form in which the information was first presented; Responses using a modifier-noun phrase are common, but are even more used when the story also uses a modifier-noun phrase then when the story uses a full-phrase. Furthermore, although responses using a full response were relatively rare, the vast majority of responses that used a full phrase (n=144) were produced when the story also used a full phrase whereas only 7 responses using a full phrase were produced when the story did not use a full phrase. This

finding is consistent with previous research on syntactic priming (Bock 1986; Bock & Loebell 1990) that found that people are more likely to produce passive constructions when describing a scene when previous sentences contained passive constructions than when previous sentences did not contain passive constructions. The current experiment examined part of a sentence, namely, the structure of a noun phrase, and also found support for syntactic priming.

However, the selection of syntactic form was not completely determined by the form presented in the story. Instead, there was a strong preference toward using a modifier-noun phrase (e.g., wilted flowers) rather than a full noun phrase (e.g., flowers that are wilted). Previous work on referential communication has indeed shown an overall trend towards the use of shortened expressions (Brennan & Clark 1996; Markman et al. 1997) and analyses of text corpora also show evidence of text compression (Marsh 1984). Thus, the preference for a modifier-noun phrase might reflect a tendency to select a syntactically simpler construction. Modifier-noun phrases are syntactically simpler than full noun phrases and yet still provide information that allows the reader/listener to identify a subcategory of head noun (e.g., *ceramic cup* refers to a particular subcategory of the category *cup*). Thus, modifier-noun phrase constructions offer a balance between syntactic simplicity and informativeness. At the same time, there was little evidence to suggest that participants selected a head noun only structure over a modifier-noun structure, even though head noun only structures are syntactically simpler than modifier-noun phrase structures. That is, rather than exhibiting an overall bias towards shortening, per se, our data indicate a bias towards modifier-noun phrase use, which suggests that modifier-noun phrase might have a special status in the language. Although full phrases (e.g., *flowers that are fresh*) were almost always shortened (to either a modifier-noun phrase or noun, e.g., *fresh flowers* or *flowers*), modifier-noun phrases were rarely shortened to noun-only. Thus, the use of a modifier-noun phrase rather than a full phrase might reflect something about the special status of modifier-noun phrases rather than a general bias toward syntactically simple constructions, per se. That is, it seems likely that modifier-noun phrases are particularly useful for conveying subcategory information. People are sensitive to overt cues that indicate the existence of a contrast set, such as the presence of the word *only*, and the inclusion of this cue affects the relative ease of resolving main clause/reduced relative clause ambiguities (Sedivy 2002). Perhaps the inclusion of modifying information in the context implied the existence of a contrast set. This might have encouraged people to use a modifier-noun phrase when referring to the target referent because this construction indicates a contrast set (Markman 1991).

In sum, we see some evidence for syntactic priming in that the form of the presentation in the story could reduce the bias to producing modifier-noun phrases, but the influence of the prior form was relatively weak in that it was not able to overturn the strong preference for modifier-noun phrases constructions. Similarly, although we see some degree of shortening of the referring phrase, there still seems to be a preference for maintaining at least a modifier-noun construction, rather than just a generic noun, even though no modifying information was required in order to identify the referent in the story. This was particularly true when the modifying information was atypical.

4 Conclusion

Our data reveal that the context in which the linguistic expressions are used provides useful cues as to the form that the linguistic expression will take and provide insight into the expectations/biases that languages users use during referential communication. During conversation and referential communication, modifier-noun phrases (e.g., *rotten apple*) are produced for several reasons including distinguishing among potential referents and maintaining conversational pacts. The current experiment demonstrates that modifier-noun phrases also are produced in order to highlight conceptually distinctive properties. The finding that distinctiveness influenced the use of modifying information provides insight into how people use multiword expressions to convey information about how they are conceptualizing the various entities about which they are communicating. In particular, the form of the linguistic construction (e.g., noun versus modifier-noun phrase) provides useful cues as to the intended meaning. Furthermore, although the participants were somewhat sensitive to the syntactic form with which the target was presented, there was a strong bias for the modifier-noun phrase form. In sum, it appears that modifier-noun phrases have a privileged status among multiword expressions and provide a good compromise between competing principles of conveying sufficient information and using simple syntactic structures.

References

Ariel, Mira. 1990. *Accessing noun-phrase antecedents*. New York: Routledge.

Arts, Anja, Alfons Maes, Leo G M Noordman & Carel Jansen. 2011. Overspecification in written instruction. 49(3). 555–574.

Bock, Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18. 355–387.

Bock, Kathryn & Helga Loebell. 1990. Framing sentences. *Cognition* 35(1). 1–39.

Brekle, Herbert E. 1986. The production and interpretation of ad hoc nominal compounds in German: A realistic approach. *Acta Linguistica Academiae Scientiarum Hungaricae* 36(1-4). 39–52.

Brennan, Susan E & Herbert H Clark. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22(6). 1482–1493.

Brewer, William F & James C Treyens. 1981. Role of schemata in memory for places. *Cognitive Psychology* 13(2). 207–230.

Brown, Roger. 1958. How shall a thing be called? *Psychological Review* 65(1). 14–21.

Carstairs-McCarthy, Andrew. 2010. *The evolution of morphology*. Oxford: Oxford University Press.

Chafe, Wallace. 1994. *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.

Chang, Franklin, Kathryn Bock & Adele E Goldberg. 2003. Can thematic roles leave traces of their places? *Cognition* 90(1). 29–49.

Clark, Eve V. 1983. *The lexicon in acquisition*. Cambridge: Cambridge University Press.

Clark, Herbert H & Edward F Schaefer. 1989. Contributing to discourse. *Cognitive Science* 13(2). 259–294.

Clark, Herbert H & Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition* 22. 1–39.

Cleland, Alexandra A & Martin J Pickering. 2003. The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language* 49(2). 214–230.

Connolly, Andrew C, Jerry A Fodor, Lila R Gleitman & Henry Gleitman. 2007. Why stereotypes don't even make good defaults. *Cognition* 103(1). 1–22.

Copestake, Ann & EJ Briscoe. 2005. Noun compounds revisited. In *Charting a new course: Natural language processing and information retrieval*, vol. 16, 129–154. Netherlands: Springer.

Downing, Pamela. 1977. On the creation and use of English compound nouns. *Language* 53. 810–842.

Ferreira, Victor S, L Robert Slevc & Erin S Rogers. 2005. How do speakers avoid ambiguous linguistic expressions? *Cognition* 96(3). 263–284.

Frank, Michael C & Noah D Goodman. 2012. Predicting pragmatic reasoning in language games. *Science* 336(6084). 998.

Gagné, Christina L. 2002. Lexical and relational influences on the processing of novel compounds. *Brain and Language* 81(1-3). 723–735.

Gagné, Christina L & Gregory L Murphy. 1996. Influence of discourse context on feature availability in conceptual combination. *Discourse Processes* 22(1). 79–101.

Gagné, Christina L & Edward J Shoben. 1997. Influence of thematic relations on the comprehension of modifier–noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23(1). 71–87.

Gagné, Christina L & Thomas L Spalding. 2004. Effect of discourse context and modifier relation frequency on conceptual combination. *Journal of Memory and Language* 50(4). 444–455.

Gagné, Christina L & Thomas L Spalding. 2006. Using conceptual combination research to better understand novel compound words. *SKASE Journal of Theoretical Linguistics* 3(2). 9–16.

Gagné, Christina L & Thomas L Spalding. 2009. Constituent integration during the processing of compound words: Does it involve the use of relational structures? *Journal of Memory and Language* 60(1). 20–35.

Gagné, Christina L & Thomas L Spalding. 2011. Inferential processing and metaknowledge as the bases for property inclusion in combined concepts. *Journal of Memory and Language* 65(2). 176–192.

Gagné, Christina L & Thomas L Spalding. 2014a. Conceptual composition: The role of relational competition in the comprehension of modifier-noun phrases and noun–noun compounds. In Brian Ross (ed.), *Psychology of Learning and Motivation*, vol. 59, 97–130. New York: Elsevier.

Gagné, Christina L & Thomas L Spalding. 2014b. Subcategorisation, not uncertainty, drives the modification effect. *Language, Cognition and Neuroscience* 29(10). 1283–1294.

Gagné, Christina L & Thomas L Spalding. 2015. Semantics, concepts, and metacognition: Attributing properties and meanings to complex concepts. In Laurie Bauer, Lívia Körtévlyessy & Pavol Štekauer (eds.), *Semantics of complex words*, 9–25. New York: Springer.

Gagné, Christina L & Thomas L Spalding. 2016. Effects of morphology and semantic transparency on typing latencies in English compound and pseudo-compound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42(9). 1489–1495.

Garrod, Simon & Anthony Anderson. 1987. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition* 27(2). 181–218.

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.

Gundel, Jeanette K, Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307.

Hampton, James A, Alessia Passanisi & Martin L Jönsson. 2011. The modifier effect and property mutability. *Journal of Memory and Language* 64(3). 233–248.

Ibarra, Alyssa & Michael K Tanenhaus. 2016. The flexibility of conceptual pacts: Referring expressions dynamically shift to accommodate new conceptualizations. *Frontiers in Psychology* 7. 561.

Jarema, Gonia, Céline Busson, Rossitza Nikolova, Kyrana Tsapkini & Gary Libben. 1999. Processing compounds: A cross-linguistic study. *Brain and Language* 68(1-2). 362–369.

Jönsson, Martin L & James A Hampton. 2012. The modifier effect in within-category induction: Default inheritance in complex noun phrases. *Language and Cognitive Processes* 27(1). 90–116.

Koolen, Ruud, Martijn Goudbeek & Emiel Krahmer. 2013. The effect of scene variation on the redundant use of color in definite reference. *Cognitive Science* 37(2). 395–411.

Koolen, Ruud, Emiel Krahmer & Marc Swerts. 2016. How distractor objects trigger referential overspecification: Testing the effects of visual clutter and distractor distance. *Cognitive Science* 40(7). 1617–1647.

Lapata, Maria. 2002. The disambiguation of nominalizations. *Computational Linguistics* 28(3). 357–388.

Levi, Judith N. 1978. *The syntax and semantics of complex nominals.* Academic Press.

Libben, Gary. 1998. Semantic transparency in the processing of compounds: Consequences for representation, processing, and impairment. *Brain and Language* 61(1). 30–44.

Libben, Gary. 2014. The nature of compounds: a psychocentric perspective. *Cognitive Neuropsychology* 31(1-2). 8–25.

Maes, Alfons, Anja Arts & Leo Noordman. 2004. Reference management in instructive discourse. *Discourse Processes* 37(2). 117–144.

Markman, Arthur B, Takashi Yamauchi & Valerie S Makin. 1997. The creation of new concepts: A multifaceted approach to category learning. In TB Ward, SM Smith & J Vaid (eds.), *Conceptual structures and processes: emergence, discovery, and change*, 174–208. American Psychological Association.

Markman, Ellen M. 1991. *Categorization and naming in children: Problems of induction.* MIT Press.

Marsh, Elaine. 1984. A computational analysis of complex noun phrases in Navy messages. In *Proceedings of the 10th international conference on computational linguistics and 22nd annual meeting on association for computational linguistics*, 505–508.

Melinger, Alissa & Christian Dobel. 2005. Lexically-driven syntactic priming. *Cognition* 98(1). B11–B20.

Metzing, Charles & Susan E Brennan. 2003. When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language* 49(2). 201–213.

Olson, David R. 1970. Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review* 77(4). 257–258.

Pechmann, Thomas. 1989. Incremental speech production and referential overspecification. *Linguistics* 27(1). 89–110.

Prince, Ellen F. 1992. The ZPG letter: Subjects, definiteness, and information-status. In William C Mann & Sandra A Thompson (eds.), *Discourse description: diverse linguistic analyses of a fund-raising text*, 295–325. Amsterdam: John Benjamins.

Rubio-Fernández, Paula. 2016. How redundant are redundant color adjectives? An efficiency-based analysis of color overspecification. *Frontiers in Psychology* 7. 153.

Schmidtke, Daniel, Christina L Gagné, Victor Kuperman, Thomas L Spalding & Benjamin V Tucker. 2018. Conceptual relations compete during auditory and visual compound word recognition. *Language, Cognition and Neuroscience* 33(7). 923–943.

Sedivy, Julie C. 2002. Invoking discourse-based contrast sets and resolving syntactic ambiguities. *Journal of Memory and Language* 46(2). 341–370.

Sedivy, Julie C. 2003. Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research* 32(1). 3–23.

Smolka, Eva & Gary Libben. 2017. 'Can you wash off the hogwash?' – Semantic transparency of first and second constituents in the processing of German compounds. *Language, Cognition and Neuroscience* 32(4). 514–531.

Spalding, Thomas L & Christina L Gagné. 2014. Relational diversity affects ease of processing even for opaque English compounds. *The Mental Lexicon* 9(1). 48–66.

Spalding, Thomas L & Christina L Gagné. 2015. Property attribution in combined concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 41(3). 693–707.

Springer, Ken & Gregory L Murphy. 1992. Feature availability in conceptual combination. *Psychological Science* 3(2). 111–117.

Tanenhaus, Michael K, Michael J Spivey-Knowlton, Kathleen M Eberhard & Julie C Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268(5217). 1632–1634.

Tarenskeen, Sammie, Mirjam Broersma & Bart Geurts. 2015. Overspecification of color, pattern, and size: Salience, absoluteness, and consistency. *Frontiers in Psychology* 6. 1703.

Westerbeek, Hans, Ruud Koolen & Alfons Maes. 2015. Stored object knowledge and the production of referring expressions: The case of color typicality. *Frontiers in Psychology* 6. 935.

Appendix

Table 2: Full list of target items showing the unusual properties, normal properties, and the head noun.

Properties		Head noun
blue	brown	dog
fresh	wilted	flowers
rotten	red	apples
soggy	crisp	crackers
polluted	blue	lake
curdled	white	milk
rubbery	savory	chicken
soap	shoe	shop
melted	frozen	ice cream
blurry	glossy	photographs
explicit	meaningful	lyrics
green	orange	fire
stale	soft	buns
yellow	green	grass
clown	public	school
carrot	sweet	candy
snake	soft	slippers
cold	hot	shower
sour	sweet	honey
monster	school	friends
crashing	flying	planes
candy	boreal	forests
gravy	train	station
plastic	coloured	chalk
purple	morning	sunshine
plaid	school	pants
smokeless	smoky	cigarettes
pickle	juice	pitcher

Chapter 6

Can you reach for the planets or grasp at the stars? – Modified noun, verb, or preposition constituents in idiom processing

Eva Smolka & Carsten Eulitz
University of Konstanz, Germany

Idioms are a special case of multiword expressions in that their meaning cannot be compositionally constructed from the meaning of the single constituents. The question of how the idiomatic meaning is assembled remains an unsettled issue in psycholinguistic research. The present study examines whether the figurative meaning of an idiom is recognized if critical idiomatic constituents, such as the noun, verb, or preposition, are modified. In three paraphrase experiments, participants saw (a) the canonical idiomatic phrase (e.g., *She reached for the stars*), (b) the idiomatic phrase with a modified constituent (e.g., *She reached/grasped for/at the stars/planets*), or (c) a matched literal control sentence (e.g., *She reached for the sweets*) and rated on a scale from 1 (not at all) to 7 (completely) how strongly the sentence reflected the meaning of a paraphrase of the idiom (e.g., *She has always aspired to unattainable goals*).

Sentence type and constituent type strongly affected paraphrase ratings with highest ratings for canonical idiomatic phrases, lowest ratings for control sentences, and ratings in between for idioms with modified constituents. Further, idioms with modified verbs were rated higher in matching the figurative meaning than idioms with modified prepositions or nouns. Overall these findings indicate that the figurative meaning was assembled in spite of the modifications. We conclude that idioms are not fully 'semantically fixed' but allow for some flexibility in the processing of idioms. Modified constituents that activate meanings similar to those of the canonical constituents will co-activate the figurative meaning of the idiom together with the other idiomatic constituents. We discuss psycholinguistic models on idiom comprehension.

Eva Smolka & Carsten Eulitz. 2019. *Can you reach for the planets or grasp at the stars?* – Modified noun, verb, or preposition constituents in idiom processing. In Sabine Schulte im Walde & Eva Smolka (eds.), *The role of constituents in multiword expressions: An interdisciplinary, cross-lingual perspective*, 143–166. Berlin: Language Science Press. DOI:??

1 Introduction

Idioms like *nach den Sternen greifen* (literal, L, and figurative, F, translation: 'reach for the stars') represent a special type of multi-word expressions. As with other semantically opaque word formations, the figurative meaning of idioms is not derived compositionally from the meaning of the constituents and their syntactic assembly. For example, the figurative meaning of the idiom *She spilled the beans* cannot be derived by combining the meaning of the individual constituents (*'she', 'spilled', 'the', 'beans'*) and their syntactic combination ('an agent spilling some object') as would be the case in *She spilled the coffee*, despite the parallel syntactic structure. Hence, one of the aims of linguistic theory (e.g.,Grice 1975, Grice 1978) has been the formulation of distinguishing criteria for idiomatic as compared to literal multi word expressions. The most important of these are 'semantic fixedness' and 'syntactic anomaly'. Semantic fixedness specifies that the figurative meaning does not allow the replacement of any of the constituents (e.g., **she dropped the beans; *she spilled the seeds/pellets*), while syntactic fixedness indicates that the figurative meaning restricts the syntactic transformations that an idiomatic expression may undergo (e.g., **the beans were spilled by her; *she spilled the secret beans*).

Linguistic and psycholinguistic researchers are thus puzzled by the question of how idiomatic meaning is processed and stored in lexical memory (Burger 2003 Burger 2004; Cacciari & Glucksberg 1994; Gibbs Jr 2002; Gibbs 1994; Swinney & Cutler 1979; for a review see, Titone & Connine 1999; Titone & Libben 2014). In particular, it remains an unresolved question whether the meaning of an idiom is represented separately from the meaning of its parts, and how the figurative meaning is assembled. Seminal studies argued for a 'non-compositional' representation in which the whole figurative meaning of an idiomatic phrase is stored as a distinct entry, the "idiom word" in the mental lexicon similar to the representation of a complex word like *Finanzmarktaufsichtsbehörde* ('financial market supervisory authority'). "Idiomatic processing", the process by which figurative meaning is retrieved is thus assumed to be independent from the process by which literal meaning is computed (Bobrow & Bell 1973; Gibbs 1980; Swinney & Cutler 1979).

In contrast, 'hybrid' approaches assume that idioms are both unitary (i.e. each idiom possesses a distinct lexical entry for its figurative meaning) and compositional (i.e. composed of the single word lemmas of the constituents). The constituents are first processed literally until the "idiom key" or something akin to a unitary entry that carries the idiomatic concept is reached and activated (Cacciari & Tabossi 1988; Caillies & Butcher 2007; Connine et al. 1992; Cutting &

Bock 1997; Gibbs Jr & Nayak 1989; Holsinger & Kaiser 2013; Sprenger et al. 2006; Titone & Connine 1999). For example, even though idioms are syntactically analyzed similar to literal sentences, Cutting & Bock 1997 postulate a distinct "lexical concept node" that is activated by the idiomatic concept. Similarly, Sprenger et al. 2006 assume a "superlemma" like [spill-the-beans] that specifies the information relating to that idiom, such as the single constituents (i.e. *spill*, *the*, and *beans*), their syntactic functions (subject, direct object), syntactic categories (noun phrase, prepositional phrase), and parts of speech (noun, verb). Other hybrid models assume that the literal meanings of the constituents are activated only before the unitary entry is reached. For example, the "configuration hypothesis" (e.g., Cacciari & Tabossi 1988) postulates a so-called "idiom-key" – the point at which the specific word configuration renders an idiom with figurative meaning. Words of a sentence are processed in a literal way until the idiom key is reached and the word formation is recognized as expressing figurative meaning. As soon as the idiom key has been hit, only the figurative meaning of the idiom is processed and remains activated, while the literal activations disappear.

However, Smolka and colleagues (Rabanus et al. 2008; Smolka et al. 2007) observed that the literal meaning of verbs remains accessible even after the idiom key has been hit. In two sentence priming experiments, participants read an idiomatic sentence, such as *"Sie hat ihm gründlich den Kopf gewaschen"* (word-by-word, W: *'she has him thoroughly the head washed'*; literal, L: *She thoroughly washed his head; figurative*, F: *"She gave him a piece of her mind"*) and made lexical decisions about words associated with the figurative meaning (e.g., *Standpauke*, 'telling-off'), about associations with the literal meaning of the verb (e.g., *Kleidung*, 'clothes'), and about matched unrelated words.

Because all sentences were highly predictable (i.e., with cloze probabilities, on average, higher than 87%), the idiom key – the point at which the constituents are recognized to form an idiom – should occur before the sentence-final word (e.g., *gewaschen*, 'washed'). The sentences were presented visually and targets were presented 500 ms after the presentation of the verb participle to make sure that the figurative meaning was available. Under these experimental conditions, the configuration hypothesis (Cacciari & Tabossi 1988) predicts figurative meaning activations only. However, the results of both studies showed that associations with the literal meaning of the verb were activated to the same degree as were associations with the figurative meaning.

The authors concluded that (1) the literal meaning of single word constituents is accessed during figurative processing and that (2) the literal meaning, at least that of verbs, remains activated even after the figurative meaning of the idiom has been recognized (e.g., Cacciari & Tabossi 1988; Cutting & Bock 1997; Sprenger et

al. 2006). Note that hybrid models, assuming an idiom key, specify that the literal meaning of the constituents is not recalled as soon as the figurative meaning of the idiom is recognized, and (3) described a model on idiom comprehension that incorporates the complexity of idiom processing: the meaning of the single constituents is activated, and the joint co-activation of the single constituents activates the figurative meaning at the conceptual level.

The above findings give rise to the following questions: If a single idiomatic constituent activates its literal meaning alongside the figurative meaning, and if the joint activation of idiomatic constituents triggers the figurative meaning, will a close associate of the idiomatic constituent (that activates a similar meaning) contribute to the activation of the figurative meaning of the idiom? For example, will the word *planets* in the configuration *"reach for the planets"* activate the figurative meaning of *"reach for the stars"*? A positive finding would indicate that idioms are not as semantically fixed as current models on idiom processing assume (e.g., Sprenger et al. 2006). Furthermore, are some constituents of the idiom more susceptible to modification than others? That is, does the word category of an idiomatic constituent – whether it is the verb, the noun, or a preposition – influence whether the constituent can be modified without losing the figurative meaning?.

Indeed, in a recent study, Geeraert et al. 2017 observed that noun constituents of idioms may be modified to some degree. Participants rated the acceptability of idioms in their canonical form (e.g., *Although these were new stocks, they went through the roof*), with a lexical variation that included a near synonym (e.g.,*...they went through the ceiling*), when idioms were partial forms (e.g., *...they went through it*), when they held an integrated concept (e.g. *...they went through the investment roof*), or when they were idiom blends (e.g., *...they suddenly went through the charts*). Modifications of the idiom made it less acceptable, however, the degree of the acceptability depended on the type of the variation, indicating that modifications with near synonyms (*roof – ceiling*) or integrated concepts (*investment roof*) were more acceptable than other variations. The authors concluded that their findings challenge any theories on idiom processing that assume fixed units for the specification of the figurative meaning, be it multi-word form units (Bobrow & Bell 1973), superlemmas (e.g. Sprenger et al. 2006), or word configurations (Cacciari & Tabossi 1988).

The aim of the present study was to examine the semantic fixedness of idioms in more detail: (a) Will the figurative meaning of an idiom be retained, if an idiomatic constituent, such as the noun, verb, or preposition, is modified? (b) Will the word category of an idiomatic constituent (noun, verb, preposition) affect whether a modification will preserve the figurative meaning?.

For this purpose, we conducted three sentence paraphrase experiments. Each canonical idiomatic sentence, such as *"Sie hat immer nach den Sternen gegriffen"* (L: *'She always reached for the stars'*; F: *"She reached for the stars"*) was presented in three versions: (1) with its canonical constituent, (2) with the canonical constituent replaced by a closely associated word, or (3) with the canonical constituent replaced by an unrelated word. We manipulated the noun constituent in Experiment 1, the verb constituent in Experiment 2, and the preposition in Experiment 3. The idiomatic noun constituent (e.g., *stars*) was replaced by a closely associated noun (e.g., *planets*), as in *"Sie hat nach den Planeten gegriffen"* (L: *'She reached for the planets'*) or by an unrelated noun (e.g., *sweets*), as in *"Sie hat nach den Bonbons gegriffen"* (L: *'She reached for the sweets'*). In Experiment 2, the idiomatic verb constituent (e.g., *reach*) was substituted by a closely associated verb (e.g., *grasp*), as in *"Sie hat nach den Sternen gelangt"* (L: *'She grasped at the stars'*) or by an unrelated verb (e.g., *ask*), as in *"Sie hat nach den Sternen gefragt"* (L: *'She asked for the stars'*). In Experiment 3, the idiomatic preposition was replaced by another preposition, as in *"Sie hat zu den Sternen gegriffen"* (L: *'She reached to the stars'*) or by an unrelated prepositional phrase (that held the original preposition of the idiom), as in *"Sie hat nach den Bonbons gegriffen"* (L: *'She reached for the sweets'*). Each sentence was paired with the paraphrase of the idiomatic sentence, *"Sie hat immer etwas Unerreichbares angestrebt"* (L: *'She always strived for something unreachable'*), and participants rated on a scale from 1 to 7 how well the meanings of two sentences mirrored each other. Examples of idiomatic sentences, their modifications and paraphrases are given in Tables 1–3.

In all three experiments, we used idiomatic sentences and minimized the influence of some confounding variables by controlling the following factors: (a) the number of words in a sentence was alike, that is, each sentence was comprised of seven words; (b) all sentences had the same structure (subject-verb-prepositional phrase-participle) and all were presented in the perfect tense, so that the position of the verb was always sentence-final, and (c) all sentences had a high cloze probability (on average 90%), ensuring that the sentence-final word was highly predictable. It was thus established that the phrasal meaning was processed and the word configuration was rendered as figurative before the last word of a sentence. Finally, to provide a strong basis for the generalization of our findings, we examined between 33 and 39 different idiomatic phrases in each experiment.

If 'unitary' entries define the idiomatic constituents, then sentences whose idiomatic constituents are replaced by close associates will not be considered to hold the figurative meaning and should yield paraphrase ratings similar to sentences with unrelated constituents. If, however, the assumptions hold (a) that each idiomatic constituent activates its literal meaning, (b) that a close associate

Table 1: Examples of sentence triplets for idiomatic phrases with modified nouns and corresponding paraphrase. *Notes:* W = word by word; L = literal; F = figurative

Idiomatic/Modified/Unrelated	Paraphrase
Sie hat immer nach den Sternen/Planeten/Bonbons gegriffen. (W) She has always for the stars/planets/sweets reached. (L) She always reached for the stars/planets/sweets. (F) She always reached for the stars.	*Sie hat immer etwas Unerreichbares angestrebt.* (W) She has always something unreachable aspired. (L) She always aspired to something unreachable.
Sie haben die Katze im Sack/Beutel/Frühling gekauft. (W) They have the cat in the sack/bag/spring bought. (L) They bought the cat in the sack/bag/spring. (F) They bought a pig in a poke.	*Sie wurden beim ungeprüften Kauf übervorteilt.* (W) They were at the uninspected purchase outsmarted. (L) They were outsmarted at the uninspected purchase.
Sie hat immer aufs falsche Pferd/Roß/Spielfeld gesetzt. (W) She has always on the wrong horse/steed/playing field bet. (L) She always bet on the wrong horse/steed/playing field. (F) She always bet on the wrong horse.	*Sie hat die Lage immer falsch eingeschätzt und sich entsprechend falsch verhalten.* (W) She has the position always wrongly evaluated and herself accordingly wrongly conducted. (L) She always misgauged the situation and thus acted improperly.
Er hat sich mit fremden Federn/Daunen/Edelsteinen geschmückt. (W) He has himself with foreign feathers/down/gemstones adorned. (L) He adorned himself with foreign feathers/down/gemstones. (F) He adorned himself with borrowed plumes.	*Er hat die Verdienste anderer als die eigenen ausgegeben.* (W) He has the merits of others passed as the own ones. (L) He passed off the merits of others as his own.

Table 2: Examples of sentence triplets for idiomatic phrases with modified verbs and their corresponding paraphrase. *Notes:* W = word by word; L = literal; F = figurative

Idiomatic/Modified/Unrelated	Paraphrase
Sie hat immer nach den Sternen gegriffen/gelangt/gefragt. (W) She has always for the stars reached/grasped/asked. (L) She always reached for/grasped at/asked for the stars. (F) She always reached for the stars.	*Sie hat immer etwas Unerreichbares angestrebt.* (W) She has always something unreachable aspired. (L) She always aspired something unreachable.
Mir ist ein Stein vom Herzen gefallen/gestürzt/vom Dach gerollt. (W) To me is a stone from heart fallen/dropped/from the roof rolled. (L) A stone fell/dropped off the heart/rolled off the roof. (F) It took a load off my mind.	*Ich bin darüber sehr erleichtert gewesen.* (W) I have about that very relieved been. (L) I was very relieved about that.
Er hat mir wieder einen Korb gegeben/gereicht/geflochten. (W) He has me again a basket given/handed/woven. (L) He again gave/handed/wove me a basket. (F) He turned me down again.	*Er hat mich wieder abgewiesen.* (W) He has me again rejected. (L) He rejected me again.
Seine Frau hat ihm die Schau gestohlen/geklaut/(die Hemden gebügelt). (W) His wife has him the show stolen/lifted/(his shirts ironed). (L) His wife stole/lifted his show/(ironed his shirts). (F) His wife upstaged him.	*Seine Frau ist an seiner Stelle zum Interessensmittelpunkt geworden.* (W) His wife is on his place to the focus of attention been. (L) His wife occupied center-stage instead of him

Table 3: Examples of sentence triplets for idiomatic phrases with modified prepositions and their corresponding paraphrase. *Notes:* W = word by word; L = literal; F = figurative

Idiomatic/Modified/(Unrelated Prepositional Phrase)	Paraphrase
Sie hat immer nach/zu den Sternen/(nach den Bonbons) gegriffen. (W) She has always for/at the stars reached. (L) She always reached for/to the stars. (F) She always reached for the stars.	*Sie hat immer etwas Unerreichbares angestrebt.* (W) She has always something unreachable aspired. (L) She always aspired something unreachable.
Ich habe ihn sehr ins/ans Herz/(ins Zimmer) geschlossen. (W) I have him very into/onto heart/(into the room) locked. (L) I locked him into/onto my heart/(into the room). (F) I have grown very fond of him.	*Ich habe ihn sehr lieb gewonnen.* (W) I have him very dear won. (L) I like him very much.
Sie sind auf/an keinen grünen Zweig/(auf keinen grünen Golfplatz) gekommen. (W) They have on/at no green branch/(on no green golf course) come. (L) They didn't arrive at/on a green branch/golf course. (F) They could not come to terms.	*Trotz starker Bemühungen konnten sie sich nicht einigen.* (W) In spite of all efforts could they themselves not agree. (L) Despite all efforts, they did not come to an understanding.
Damit hat er Öl ins/aufs Feuer/(in den Tank) gegossen. (W) Thereby has he oil into/onto/(in the tank) poured. (L) By that he poured oil into/onto the fire/(in the tank). (F) By that he added fuel to the fire.	*Damit hat er den Streit und die Erregung noch verstärkt.* (W) Thereby has he the dispute and the arousal even reinforced. (L) By that he enforced the fight and the arousal even more.

of an idiomatic constituent will activate a similar literal meaning and (c) will thus contribute to the joint co-activation of the figurative meaning, sentences holding close associates of an idiomatic constituent will be rated as higher in reflecting the figurative meaning than those with unrelated constituents. Furthermore, if the assumption holds that the verb is the structural center of the phrase (as assumed by Rabanus et al. 2008 and Smolka et al. 2007, the modification of the verb constituent will differ from that of the noun or preposition constituents.

2 Experiment 1

2.1 Method

2.1.1 Participants

Thirty-six university students, all native speakers of German, participated in the experiment for course credit or payment.

2.1.2 Materials

Thirty-nine idiomatic phrases were selected for the sentence paraphrase test. We defined an 'idiomatic phrase' as a verb phrase (a) where both the verb and its complement are used in a nonliteral way to produce an overall idiomatic interpretation, (b) that shows some kind of morphosyntactic anomaly, and (c) whose figurative meaning is lexicalized. In the light of these three properties, the idiomaticity of the phrases selected was agreed upon by three independent judges and further verified by reference to an idiomatic phrase dictionary (Redewendungen 2002).

Each idiomatic sentence consisted of seven words and was cast in the perfect tense, rendering the past participle form of a verb in sentence-final position. All idiomatic sentences were chosen from the pool of sentences tested in the sentence-completion experiment described in Smolka et al. 2007. To assure that their figurative meaning was the dominant reading, only idiomatic phrases with high sentence completion rates were selected. That is, these sentences were completed with words that produced the figurative meaning in 93% of the cases (range 52% to 100%).

2.1.2.1 Sentence completion task

More than 1,100 sentences in literal and figurative meaning were tested in a sentence completion task (for a more detailed description see Smolka et al. 2007).

For the completion task, the last word of a sentence (i.e. the past participle) was omitted and completed by between 25 and 32 monolingual native speakers of German in an online portal (Language Experiments Portal by Keller et al. 1998). For each sentence, the number of sentence completions with a specific verb was counted. For example, 19 of the 25 participants who saw the sentence *"Sie hat immer nach den Sternen _____"* (L: *She always _____ for the stars*) completed it with the participle *gegriffen* ('reached') and thus finalized the sentence in its figurative meaning *"She always reached for the stars"*. The other 6 participants completed the sentence with the verb *geschaut* ('looked at') and thus yielded the literal meaning "She always looked at the stars".

2.1.2.2 Noun association test

Each idiomatic sentence, such as *"She reached for the stars"*, was cast in three versions, holding either (a) the canonical idiomatic noun constituent (I), such as *Sterne* ('stars'), (b) an associated noun (A), such as *Planeten* ('planets'), or (c) an unrelated noun (U), such as *Bonbons* ('sweets'). Tables 1–3 provides examples of idiomatic sentences and the noun modifications; Table 4 provides the stimulus characteristics of the idiomatic sentences and their corresponding noun constituents.

Table 4: Idiomatic sentences and stimulus characteristics of the idiomatic, modified, and unrelated noun constituents in Experiment 1. *Note.* N = number of items, Lemma = mean lemma frequency per one million, taken from CELEX (Baayen et al. 1993), Closure = mean sentence completion in %.

		Type of noun		
	Idiomatic phrase	Idiomatic	Modified	Unrelated
Example (translation)	*Sie hat nach den Sternen gegriffen* (L: She reached for the stars')	*Sternen* ('stars')	*Planeten* ('planets')	*Bonbons* ('sweets')
N		39	39	39
Lemma		57.1	25.5	39.6
Surface		32.3	16.8	22.9
Association		–	5.8	–
Closure		93.1	–	–

To find close associates, two noun associations (e.g., *planets* and *moons*) were selected for each of the idiomatic noun constituents that should be modified (e.g., *stars* in the idiomatic phrase *"She reached for the stars"*). Care was taken that the

associations were unrelated with the figurative phrasal meaning, and that the gender and the number inflections of the noun associations fitted the original idiomatic sentence. To avoid episodic effects, the same noun occurred only once in the whole experiment.

The strength of the associations was assessed in a pre-test that comprised two lists. The two noun associates of the idiomatic constituents were allocated to two lists; in both lists, the idiomatic noun constituent was paired with its associated noun and with an unrelated noun. For example, in List 1, the idiomatic constituent *Sterne* ('stars') was presented with the association *Planeten* ('planets') and the unrelated noun *Praxis* ('practice'); in List 2, *Sterne* ('stars') was presented with *Monde* ('moons') and *Praxis* ('practice').

Forty participants (who did not participate in the paraphrase experiment) rated on a scale from 1 (not at all) to 7 (strongly) how strongly the two nouns (e.g., *stars – planets*) are meaning related. Noun associations were selected as modifications of the original idiomatic noun, if they received high ratings (mean rating 5.8), and if their lemma and surface frequencies (taken from CELEX, see Baayen et al. 1993) were well matched with those of the idiomatic constituent.

2.1.2.3 Paraphrases

For each idiomatic phrase, we constructed a paraphrase by looking up the definition of the idiom in the idiomatic phrase dictionary (Redewendungen 2002). For a similar appearance as the idiomatic sentence, the paraphrase was cast in the past perfect tense and with the same subject as that of the idiomatic phrase.

2.1.3 Apparatus and procedure

Three lists were constructed, each included one of the sentence triplets of an idiomatic phrase. The three versions of an idiomatic phrase were rotated over the three lists by Latin square in such a way that a list contained the idiomatic sentence either with the canonical idiomatic (I), the associated (A), or the unrelated (U) noun. Each of the three sentence triplets was paired with the same paraphrase of the idiomatic sentence (see Table 1 for examples). Altogether, each list comprised 39 sentence pairs.

Participants were randomly assigned to a list. Paraphrase tests were distributed via email. Participants rated on a scale from 1 to 7 how strongly the two sentences reflected each other's meaning. The instructions included two examples: one sentence pair with high meaning relatedness, the other sentence pair with low meaning relatedness.

2.2 Results

In this and the following experiments, we used R (Venables & Smith 2012) and lme4 (e.g., Bates 2005; Bates et al. 2012; Baayen et al. 2008) to perform linear mixed effects modeling (LMM). As random effects, we had intercepts for participants and items (i.e. sentences). As fixed effects, we included the factor Modification Type (idiomatic/associated/unrelated), the Sentence Closure of the idiomatic phrase, and the Frequency of the constituent. The absolute and normalized lemma frequencies were taken from the CELEX lexical database (Baayen et al. 1993) and were log-transformed and centered (e.g. Winter 2013). All *p*-values were calculated on the basis of Satterthwaite approximation by using the *lmerTest* package (Kuznetsova et al. 2015). In this and the following experiments, we applied a forward procedure for the model selection, starting with a minimal model and adding additional predictors only when they improved the model fit. The best model fit was obtained by comparing the Akaike Information Criterion (AIC) statistics between models, with a difference between models > 4 (Sakamoto et al. 1986).

Table 5: Fixed effects of the predictors in the linear mixed-effect model for the paraphrase ratings in Experiment 1. *Note.* significance code: *** < 0.0001.

	Estimate	Std. Error	df	*t*-value	*p*	
(Intercept: Idiomatic)	6.246	0.179	123.1	34.83	$<2.00 \times 10^{-16}$	***
Constituent (Modified)	−2.115	0.210	112.8	−10.08	$<2.00 \times 10^{-16}$	***
Constituent (Unrelated)	−3.722	0.210	112.8	−17.73	$<2.00 \times 10^{-16}$	***

The LMM analysis of Experiment 1 indicated that the best model fit included the fixed-effect factor Modification Type, no other fixed-effect factors were significant. Table 5 summarizes the effects; the left panel of Figure 1 depicts the ratings. Results were straightforward: Paraphrase ratings were highest to idiomatic phrases that held the canonical idiomatic noun constituent (mean = 6.25, SD = 1.81), lower to phrases in which the canonical noun was modified by a closely associated noun (mean = 4.13, SD = 2.33), and lowest to phrases with unrelated nouns (mean = 2.52, SD = 2.21).

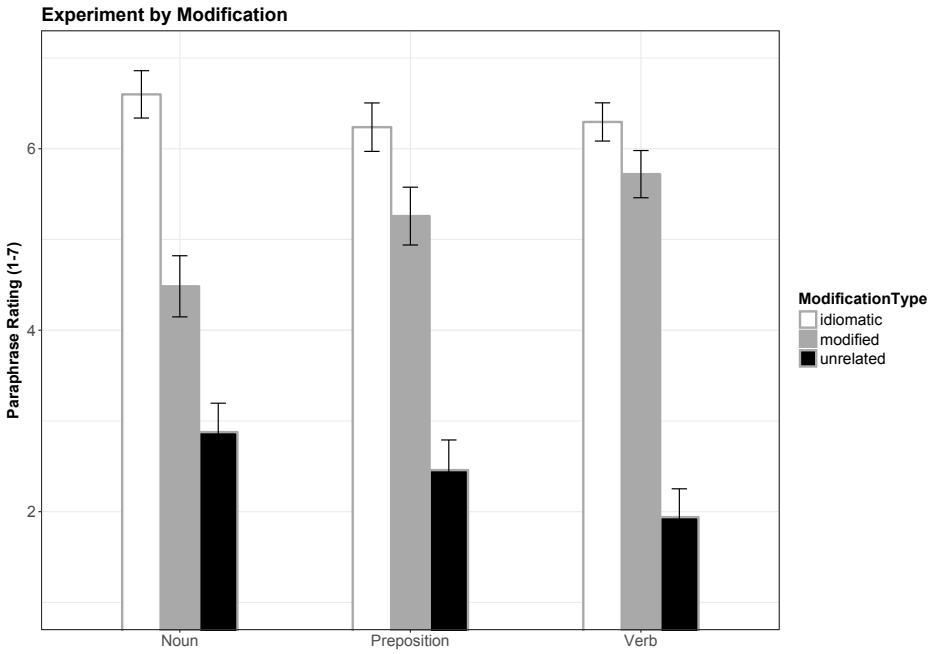


Figure 1: Paraphrase ratings on a scale from 1–7 for idiomatic sentences holding idiomatic, modified, or unrelated constituents. Noun constituents were manipulated in Experiment 1 (left panel), prepositions in Experiment 3 (mid panel), and verb constituents in Experiment 2 (right panel). Y-bars indicate standard errors of the mean.

3 Experiment 2

3.1 Method

3.1.1 Participants

Fifty-eight university students who had not participated in the previous experiment participated in the experiment for course credit or payment. All were native speakers of German.

3.1.2 Materials

Thirty-three idiomatic phrases were selected for the sentence paraphrase test according to the same principles as in Experiment 1: They were fully idiomatic phrases as defined in Experiment 1 and were selected from the sentence pool

described in Experiment 1. To ensure that their figurative meaning was the dominant reading, all had high sentence completion rates, that is, they were completed with verbs that produced the figurative meaning in 91% of the cases (range 52% to 100%).

Each idiomatic sentence, such as *"Sie hat nach den Sternen gegriffen"* (F: *"She reached for the stars"*), was cast in three versions, holding either (a) the canonical idiomatic verb (I), such as *gegriffen* ('reached'), (b) an associated verb (A), such as *gelangt* ('grasped'), or (c) an unrelated verb (U), such as *gefragt* ('asked'). In 26 of the 33 unrelated verbs, also the noun constituent that precedes the verb was modified to create a meaningful sentence, such as *"Sie hat nach den Sternzeichen gefragt"* (L: *"She asked for the zodiacs"*). See Tables 1–3 for examples of idiomatic sentences and their verb modifications. Table 6 provides the stimulus characteristics of the idiomatic sentences and the corresponding modifications.

Table 6: Idiomatic sentences and stimulus characteristics of the idiomatic, modified, and unrelated verb constituents in Experiment 2. *Note.* N = number of items, Lemma = mean lemma frequency per one million, taken from CELEX (Baayen et al. 1993), Closure = mean sentence completion in %.

		Type of noun		
	Idiomatic phrase	Idiomatic	Modified	Unrelated
Example (translation)	*Sie hat nach den Sternen gegriffen* (L: 'She reached for the stars')	*gegriffen* ('reached')	*gelangt* ('grasped')	*gefragt* ('asked')
N		33	33	33
Lemma		135	69	440
Surface		18.5	7	24.5
Association		–	5.3	–
Closure		90.9	0.46	–

3.1.2.1 Verb association test

Two verb associations (e.g., *fassen*, 'grip' and *langen*, 'grasp') were selected for each of the idiomatic verbs that should be modified (e.g., *greifen*, 'reach'). It was taken care of that the verb associations were unrelated with the figurative phrasal meaning and that they generated a meaningful sentence. To avoid episodic effects, the same verb occurred only once in the whole experiment.

The strength of the associations was assessed in a pre-test. The two associates of an idiomatic verb were allocated to two lists; in both lists, the idiomatic verb

was paired with an associated and an unrelated verb. For example, in List 1, the idiomatic verb *greifen* ('reach') was presented with the association *fassen* ('grip')and the unrelated verb *kleben* ('stick'); in List 2, *greifen* ('reach') was presented with *langen* ('grasp') and *kleben* ('stick'). Each list tested 112 verb pairs.

Thirty participants (who did not participate in the paraphrase experiment) rated on a scale from 1 (not at all) to 7 (strongly) how strongly the meanings of the two verbs (e.g., *greifen – langen*) are related. Verb associations were selected as associations of the original idiomatic verb, if they received high ratings (mean rating 5.8), and if their lemma and surface frequencies (taken from CELEX, see (Baayen et al. 1993) were well matched with those of the idiomatic verb.

3.1.2.2 Paraphrases and fillers

The same procedure as in Experiment 1 was used to construct the paraphrases for each idiomatic phrase (see also Tables 1–3). In addition to the 33 idiomatic sentences, 22 literal sentences with the same sentence structure were used as fillers and were paired with unrelated paraphrases.

3.1.3 Apparatus and procedure

Participants were randomly assigned to a list. Paraphrase tests were distributed via email. Participants rated on a scale from 1 to 7 how strongly the two sentences reflected each other's meanings. The instructions included two examples, one sentence pair with high meaning relatedness, the other sentence pair with low meaning relatedness. As in Experiment 1, three lists were constructed in such a way that each included one of the sentence triplets of an idiomatic phrase, either with the idiomatic (I), associated (A), or unrelated (U) verb. Each of the three sentence triplets was paired with the same paraphrase of the idiomatic sentence (see Tables 1–3 for examples). The same 22 filler sentence pairs were added to each list, so that, altogether, each list comprised of 55 sentence pairs. The number of fillers ensured that 60% of the sentences in a list were not meaning related with their paraphrase.

3.2 Results

We applied the same LMM analyses as described in Experiment 1. The best model fit included the fixed-effect factor Modification Type and is summarized in Table 7; the right panel of Figure 1 depicts the paraphrase ratings. As in Experiment 1, paraphrase ratings were the highest to idiomatic phrases that held the canonical idiomatic verb (mean = 6.54, SD = 1.46), lower to phrases in which the canonical

Table 7: Fixed effects of the predictors in the linear mixed-effect model for the paraphrase ratings in Experiment 2. *Note.* significance code: *** < 0.0001, * < 0.05.

	Estimate	Std. Error	df	*t*-value	*p*	
(Intercept: Idiomatic)	6.538	0.1389	112.2	47.07	$<2.00 \times 10^{-16}$	***
Constituent (Modified)	−0.580	0.1852	93.61	−3.13	0.00234	*
Constituent (Unrelated)	−4.373	0.1853	93.68	−23.61	$<2.00 \times 10^{-16}$	***

verb was modified by a closely associated verb (mean = 5.96, SD = 1.80), and the lowest to phrases with unrelated verbs (mean = 2.18, SD = 2.16).

4 Experiment 3

4.1 Method

4.1.1 Participants

Fifty university students, all native speakers of German, participated in the experiment for course credit or payment.

4.1.2 Materials

Thirty-three idiomatic phrases were selected for the sentence paraphrase test according to the same principles as in Experiment 1: They were fully idiomatic phrases and selected from the same sentence pool as described in Experiment 1. To ensure that their figurative meaning was the dominant reading, all had high sentence completion rates, that is, they were completed with words that produced the figurative meaning in 90.2% of the cases (range 52% to 100%).

Each idiomatic sentence, such as *"Sie hat nach den Sternen gegriffen"* (F: *"She reached for the stars"*), was cast in three versions, holding either (a) the canonical idiomatic preposition (I), such as *nach* ('after'), (b) a modified preposition (A), such as *zu* ('to'), or (c) an unrelated prepositional phrase that held the same preposition as the idiomatic phrase (U), such as *nach den Bonbons* ('for the sweets'). See Tables 1–3 for examples of idiomatic sentences and their prepositional modifications. Table 8 provides the stimulus characteristics of the idiomatic sentences and the corresponding modifications.

Table 8: Idiomatic sentences and stimulus characteristics of the idiomatic and modified preposition, and unrelated prepositional phrase in Experiment 3. *Note.* N = number of items, Lemma = mean lemma frequency per one million, taken from CELEX (Baayen et al. 1993), Closure = mean sentence completion in %.

		Type of Preposition		
	Idiomatic phrase	Idiomatic	Modified	Unrelated PP
Example (translation)	*Ich habe ihn ins Herz geschlossen* (L: 'I locked him into my heart' F: 'I am fond of him')	*ins* ('into')	*ans* ('onto')	*ins Zimmer* ('into the room')
N		33	33	33
Lemma		6.7	4.2	–
Closure		90.2	–	–

4.1.2.1 Preposition substitution

Since prepositions may take many different meanings, so that association tests are not applicable, two native speakers selected a preposition (e.g., *zu*, 'to') that best matched the meaning of the idiomatic preposition (e.g., *nach*, 'after'). We made sure that the modified preposition fitted the sentence frame and generated a meaningful sentence. As 'unrelated' control condition, we used the idiomatic preposition and combined it with an unrelated noun phrase.

4.1.2.2 Paraphrases and fillers

The same procedure as in Experiment 1 was used to construct the paraphrases for each idiomatic phrase. In addition to the 33 idiomatic sentences, 22 literal sentences with the same sentence structure were used as fillers and were paired with unrelated paraphrases.

4.1.3 Apparatus and procedure

Three lists were constructed in such a way that each included the idiomatic phrase with either the canonical idiomatic preposition (I), the modified preposition (A), or the unrelated prepositional phrase (U). Each of the three sentence triplets was paired with the same paraphrase of the idiomatic sentence (see Tables 1–3 for examples). Twenty-two fillers in each list reduced the relatedness proportion (between the sentences of a sentence pair) in a list to 40%. The rest of the procedure was the same as in the previous experiments.

4.2 Results

We applied the same LMM analyses as described in Experiment 1. The best model fit included the fixed-effect factor Modification Type only and is summarized in Table 9; the mid panel of Figure 1 depicts the paraphrase ratings. As in the previous experiments, paraphrase ratings were highest to idiomatic phrases that held the canonical idiomatic preposition (mean = 6.23, SD = 1.85), lower to phrases with a modified preposition (mean = 5.28, SD = 2.21), and lowest to phrases with an unrelated prepositional phrase (mean = 2.48, SD = 2.3).

Table 9: Fixed effects of the predictors in the linear mixed-effect model for the paraphrase ratings in Experiment 3. *Note.* significance code: *** < 0.0001

	Estimate	Std. Error	df	t-value	p	
(Intercept: Idiomatic)	6.260	0.1768	133.04	35.42	$<2.00 \times 10^{-16}$	***
Constituent (Modified)	−0.973	0.2099	96.36	−4.63	1.13×10^{-5}	***
Constituent (Unrelated)	−3.797	0.2099	96.40	−18.09	$<2.00 \times 10^{-16}$	***

Table 10: Fixed effects of the predictors in the linear mixed-effect model for the paraphrase ratings combining Experiments 1–3. *Note.* significance code: *** < 0.0001, * < 0.05.

	Estimate	Std. Error	t-value	p	
(Intercept: Idiomatic, Noun)	6.245	0.159	39.18	$<2.00 \times 10^{-1}$	***
Constituent (Modified)	−2.115	0.193	−10.96	$<2.00 \times 10^{-1}$	***
Constituent (Unrelated)	−3.895	0.183	−21.25	$<2.00 \times 10^{-1}$	***
Experiment (Preposition)	0.033	0.161	0.20	0.8396	
Experiment (Verb)	0.381	0.161	2.38	0.0182	*
Constituent (Modified) x Experiment (Prep.)	1.122	0.233	4.82	2.09×10^{-6}	***
Constituent (Unrelated) x Experiment (Prep.)	0.248	0.177	1.40	0.1613	
Constituent (Modified) x Experiment (Verb)	1.454	0.233	6.24	1.18×10^{-9}	***
Constituent (Unrelated) x Experiment (Verb)	−0.443	0.214	−2.07	0.0386	*

4.2.1 Post-hoc analysis on experiments 1–3

The results of all three experiments showed that idiomatic constituents may be modified by a close associate and still yield the figurative meaning. A visual inspection of Figure 1 suggests that modified verbs are better in yielding the figurative meaning than either noun or prepositions. The following LMM analysis was conducted to test whether the word category of a constituent (noun, verb, preposition) affects whether its modification is considered as reflecting the figurative meaning better.

We applied the same LMM analysis as in the previous experiments. As random effects, we had intercepts for participants and items (i.e. sentences). In addition to the previously used fixed effects – Modification Type (idiomatic/associated/unrelated), the Sentence Closure of the idiomatic phrase, and the Frequency of the constituent (log-transformed and centered, absolute lemma frequencies from CELEX) – we included the factor Experiment (corresponding to the tested constituent). We applied a forward procedure for the model selection, and obtained the best model fit by comparing the Akaike Information Criterion (AIC) statistics between models.

The best model fit included the fixed-effect factors Modification Type and Experiment, and an interaction between the two. Table 10 summarizes the effects. The results reflect the findings depicted in Figure 1. Overall speaking, as in each of the Experiments 1–3, paraphrase ratings were highest to idiomatic phrases that held the canonical constituent, lower to phrases in which the canonical constituent was modified by a closely associated constituent, and lowest to phrases with unrelated constituents. Across experiments though, sentences with modified preposition or verb constituents received higher ratings and were thus perceived as better representing the figurative meaning than sentences with modified nouns. Further, sentences holding unrelated verbs received lower paraphrase ratings than sentences holding unrelated nouns, indicating that unrelated verbs are perceived as lowest in representing the figurative meaning.

5 General discussion

The present study investigated whether idioms are semantically fixed, as suggested by established linguistic and psycholinguistic models on the processing and production of idioms. We asked first, whether idiomatic constituents may be modified while retaining the figurative meaning, and second, whether some idiomatic constituents are more susceptible to modification than others in keeping the figurative meaning.

Previous studies observed that idiomatic verb constituents activate their literal meaning while they contribute to the activation of the figurative meaning (e.g., Rabanus et al. 2008; Smolka et al. 2007). In the present study, we thus asked whether not only the constituent itself but also a close associate of the constituent (that activates a similar literal meaning) will contribute to the activation of the figurative meaning. We compared the processing of canonical idiomatic phrases like *Sie hat immer nach den Sternen gegriffen* ('She always reached for the stars') with sentences in which one of the idiomatic constituents (i.e., the noun, verb, or preposition) was modified by a close semantic associate, as in *Sie hat immer nach/zu den Sternen/Planeten gegriffen/gelangt* (L: *'She always reached/grasped for/to the stars/planets'*). The results of the paraphrase ratings indicated that the figurative meaning of the idiom is recognized even when a semantic associate replaces the canonical idiom constituent. These findings indicate that modified idiomatic constituents may contribute to the generation of the figurative meaning of the idiom.

Our findings confirm the findings by Geeraert et al. 2017 that the figurative meaning is accepted when idiomatic noun constituents are modified by near synonyms or semantic associates (e.g., *they went through the ceiling*). We have extended the finding on noun constituents to other idiomatic constituents, such as the verb and the preposition, and have shown that they may be modified as well. Indeed, the modifications of all types of constituents (nouns, verbs, and prepositions) were rated as better reflecting the figurative meaning than unrelated constituents.

We further asked whether a particular type of constituent (noun, verb, or preposition) more strongly preserves the figurative meaning than others. Indeed, our results show that modified verbs are stronger than modified nouns or prepositions at activating the figurative meaning. This finding fits well with the assumption by Hamblin & Gibbs 1999 that the meaning of the verb in idiomatic phrases may influence the meaning of the idiom. When a verb such as *kick* in *"kick the bucket"* was replaced by a verb that expressed the fast and sudden action, such as *punt*, this substitution was rated as better preserving the meaning of the idiom than a verb that did not represent the inherent meaning of the verb, such as *nudge*. Hamblin and Gibbs concluded that the verb-inherent action was transferred to the meaning of the whole idiomatic phrase.

In the following paragraphs we are searching for a plausible reason why the modified verb more strongly activates the figurative meaning than a modified noun or preposition does: Since there are, to our knowledge, no studies that directly compare the processing of different idiomatic constituents (nouns, adjectives, verbs, prepositions) and how each contributes to the overall figurative

meaning, we are allowing ourselves to speculate why verb constituents of idioms are differently processed than noun or prepositional constituents.

The processing of modified verbs similar to canonical ones may have been further facilitated by the fact that in the present study verbs occupy the sentence-final position. From a semantic perspective, the verb is thus partly processed even before it has been encountered. Consider the German idiom *"Ich habe ihn sehr ins Herz geschlossen"* (L: *'I locked him into my heart';* F: *"I am very fond of him").* The German preposition *in(s)* governs both the dative case for locations (indicating the semantic feature [+static]) and the accusative case for directions (indicating the feature [−static]; Gansel 1992). Because the above example assigns an accusative, the semantic feature [−static] of the participle *geschlossen* ('locked') can be anticipated. Hence, certain semantic properties of the verb are processed before it is realized.

Also from a syntactic perspective, the verb is partially processed even before it has been encountered. According to valency theory (e.g., Tesnière 1959), the verb controls the syntactically obligatory complements.[1]. These complements, in turn, are dependent on the subcategorization properties of the verb and are predictable as soon as the verb has been processed. In our sentences, where the verb occupies the sentence-final position, the direction of predictability is reversed: The number and type of complements that occur in the sentence constrain the choice of possible verbs in the last position, so that the verb is partially processed even before it has been encountered.

Also the high sentence closures in the present experiments indicate that participants expect the meaning of a specific verb in sentence-final position. Hence, the meaning of the idiomatic verb constituent was activated before it was encountered, so that the modified verb, which activates a similar literal meaning, is stronger in activating the figurative meaning than other (noun or preposition) constituents that are not as expected.

To summarize, if we assume that (a) the verb-inherent action is transferred to the figurative meaning of the idiom (see Hamblin & Gibbs 1999), and (b) the literal meaning of the verb remains activated even after the figurative meaning of the idioms has been recognized (see Rabanus et al. 2008; Smolka et al. 2007), (c) the syntactic and semantic properties of the verb in the sentence-final position

[1]With respect to literal language, the verb's valency (i.e. the number of complements it requires) was shown to affect both language production (e.g., Thompson et al. 1997) and language comprehension (Shapiro et al. 1987). However, the verb's valency did not affect the processing of figurative language: Idiomatic sentences holding transitive verbs (that require one obligatory complement) and idiomatic sentences holding ditransitive verbs (that require two obligatory complements) were processed equally fast (Dörre & Smolka n.d.)

are partly processed before it is encountered, the possibility that a close associate of the verb (that activates a literal meaning similar to that of the canonical verb) will trigger the figurative meaning of the idiom.

Overall, the present findings provide evidence against any type of model on idiom comprehension or production that assumes some kind of 'fixed' lexical entry of the idiomatic constituents that generate the figurative meaning, including fixed "idiom words" (Bobrow & Bell 1973). The present findings also disagree with 'hybrid' models that assume a unitary or fixed representation to capture the idiosyncratic meaning of an idiom, such as the fixed word configuration in form of an "idiom key" (e.g., Cacciari & Tabossi 1988), fixed "superlemmas" (e.g., Sprenger et al. 2006), or fixed "lexical concept nodes" (e.g., Cutting & Bock 1997). For example, according to the configuration hypothesis (Cacciari & Tabossi 1988), the Italian sentence *"Dopo l'ottima prestazione, il tennista era al settimo cielo"* (F: *"After the excellent performance, the tennis player was in seventh heaven"*) is processed literally until the specific word configuration *"to be in seventh heaven"* is recognized to form the figurative meaning. As soon as the figurative meaning is hit, the literal meaning activation is dropped and no longer active. Accordingly, the presentation of the noun constituent *cielo* ('heaven') did not activate its literal association *stelle* ('stars'). Because the configuration hypothesis assumes that only the very specific word configuration – the idiom key – renders the figurative meaning, a sentence with a modified word configuration such as *"The tennis player was in seventh sky"* should not be able to activate the figurative meaning.

A similar assumption underlies the concept of the "superlemma" (Sprenger et al. 2006): A 'superlemma' such as [hit-the-road] specifies the single constituents of the idiom (i.e., *hit, the, road*) as well as their syntactic features and functions. It engages morphosyntactic constraints on the idiomatic configuration to discriminate idiomatic from literal word configurations (Sprenger et al. 2006). Hence, the morphosyntactic constraints of the superlemma [hit-the-road] could not apply for modified constituents such as *'hit the street'* or *'strike the road'* and would not retrieve a figurative meaning.

Overall, the present findings provide evidence against any noncompositional lexical representation of the figurative meaning of idioms. By contrast, the present findings fit well with the recent study on idiom variation referred to above (Geeraert et al. 2017). Baayen and colleagues modelled their findings in a naïve discrimination learning account (Baayen et al. 2013; 2011; 2016) that entails sublexical orthographic units such as letter trigrams that are mapped onto meaning units in form of "lexomes". The lexome of an idiom corresponds to a pointer to its semantic vector like "die" that is activated by the different letter triplets that the idiom holds. Importantly, the many different inputs may activate the same

lexome, so that *to die, pass away,* and *kick the bucket* will all activate the same lexome "die". This may explain why idioms with modifications may be acceptable to some degree. However, given that the NDL account does not recognize abstract linguistic categories, such as nouns, verbs, or prepositions, it is unclear how it could account for the finding that the modification of verbs is more effective than nouns or prepositions at activating the figurative meaning.

Finally, the present findings fit well with the "stem-based account" (Günther et al. 2018; Rabanus et al. 2008; Smolka et al. 2015; Smolka & Libben 2017; Smolka et al. 2019; 2014; 2007), which is a unitary system for the processing of literal and figurative language: Stems of multi-word expressions – ranging from complex verbs and compounds to idioms – activate the literal meanings of the stems, and together the stems co-activate their joint figurative meaning.[2]. This holds for the meanings of semantically transparent and opaque complex verbs (e.g., *understand*) and compounds (e.g., *hogwash*) just as for the opaque meaning of idioms (*kick the bucket*). Because the literal meaning of a constituent is activated alongside the figurative meaning of the multi-word expression, semantically associated words that activate a similar meaning as well as that of the idiomatic constituent will contribute to the figurative meaning assembly.

6 Conclusion

The present findings indicate that lexical representations of idioms are not as "semantically fixed" as has been assumed so far: Modified constituents that activate meanings similar to those of the canonical constituents will co-activate the figurative meaning of the idiom together with the other idiomatic constituents. Modified verb constituents more strongly activate the figurative meaning than modified noun or prepositions do. Future studies will be necessary to examine how many idiomatic constituents may be modified at once (e.g., *She grasps at the planets*) while keeping the figurative meaning of the idiom (e.g., *She reaches for the stars*).

Acknowledgements

This study was supported by Grant FP561/11 by the Volkswagen Foundation to Eva Smolka. Experiment 1 was part of Sarah Baumann's M.A. thesis, we thank

[2]Even though the literal meaning of a constituent is assumed to be activated, figurative meanings are not second-level interpretations that necessitate complete literal interpretations of the utterances on the first level. Rather, figurative interpretations do not block the activation of literal associations (see Gibbs Jr 2002)

her for conducting the experiments.

References

Baayen, R Harald, Douglas J Davidson & Douglas M Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4). 390–412.

Baayen, R Harald, Peter Hendrix & Michael Ramscar. 2013. Sidestepping the combinatorial explosion: An explanation of n-gram frequency effects based on naive discriminative learning. *Language and Speech* 56(3). 329–347.

Baayen, R Harald, Petar Milin, Dusica Filipović Đurđević, Peter Hendrix & Marco Marelli. 2011. An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review* 118(3). 438–481.

Baayen, R Harald, Richard Piepenbrock & Rijn van H. 1993. *The CELEX Lexical Database (on CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvaniam.

Baayen, R Harald, Cyrus Shaoul, Jon Willits & Michael Ramscar. 2016. Comprehension without segmentation: A proof of concept with naive discriminative learning. *Language, Cognition and Neuroscience* 31(1). 106–128.

Bates, D, M Maechler & B Bolker. 2012. Lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0. *Google Scholar*.

Bates, Douglas. 2005. Fitting linear mixed models in R. *R News* 5(1). 27–30.

Bobrow, Samuel A & Susan M Bell. 1973. On catching on to idiomatic expressions. *Memory & Cognition* 1(3). 343–346.

Burger, Harald. 2003. *Phraseologie. Eine Einführung am Beispiel des Deutschen (rev. ed.)* Berlin: Erich Schmidt.

Burger, Harald. 2004. Phraseologie - Kräuter und Rüben? Traditions und Perspectives der forschung. *In K. Steyer (Ed.),Wortverbindungen - mehr oder weniger fest. Berlin / New York*. 19–40.

Cacciari, Cristina & Sam Glucksberg. 1994. Understanding figurative language. In M. Gernsbacher (ed.), in, 447–477. US : Academic Press.

Cacciari, Cristina & Patrizia Tabossi. 1988. The comprehension of idioms. *Journal of Memory and Language* 27(6). 668–683.

Caillies, Stéphanie & Kirsten Butcher. 2007. Processing of idiomatic expressions: Evidence for a new hybrid view. *Metaphor and Symbol* 22(1). 79–108.

Connine, Cynthia M, Dawn Blasko, Russell Brandt & Jody Kaplan. 1992. Idiomatic processing: Syntactic frozenness and Subjective familiarity. *Psychological Research* 54(3). 225–232.

Cutting, J Cooper & Kathryn Bock. 1997. That's the way the cookie bounces: syntactic and semantic components of experimentally elicited idiom blends. *Memory & cognition* 25(1). 57–71.

Dörre, Laura & Eva Smolka. N.d. *When the stars are reached for–The Effects of Transitivity and Constituent Adjacency on the Processing of Passivized Idiomatic Sentences.*

Gansel, Christina. 1992. *Semantik deutscher Verben in kognitions-psychologischer Sicht.* Vol. 5. Frankfurt am Main: Peter Lang.

Geeraert, Kristina, John Newman & R Harald Baayen. 2017. Idiom variation: Experimental data and a blueprint of a computational model. *Topics in Cognitive Science* 9(3). 1–17.

Gibbs Jr, Raymond W. 2002. A new look at literal meaning in understanding what is said and implicated. *Journal of Pragmatics* 34(4). 457–486.

Gibbs Jr, Raymond W & Nandini P Nayak. 1989. Psycholinguistic studies on the syntactic behavior of idioms. *Cognitive Psychology* 21(1). 100–138.

Gibbs, R. 1994. Figurative thought and figurative language.in M. Gernsbacher (ed.) *Handbook of Psycholinguistics* (pp. 411-446). In. San Diego, CA, US: Academic Press.

Gibbs, Raymond W. 1980. Spilling the beans on understanding and memory for idioms in conversation. *Memory & Cognition* 8(2). 149–156.

Grice, H Paul. 1975. Logic and conversation. *Syntax and semantics* 3. 41–58.

Grice, H Paul. 1978. Further notes on logic and conversation. *In P. Cole (Ed.),Syntax and Semantics* vol. 9. 113–128.

Günther, Fritz, Eva Smolka & Marco Marelli. 2018. Understanding differs between English and German: capturing systematic language differences of complex words. *Cortex*. https://doi.org/10.1016/j.cortex.2018.09.007.

Hamblin, Jennifer L & Raymond W Gibbs. 1999. Why you can't kick the bucket as you slowly die: Verbs in idiom comprehension. *Journal of Psycholinguistic Research* 28(1). 25–39.

Holsinger, Edward & Elsi Kaiser. 2013. Processing (non) compositional expressions: mistakes and recovery. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 39(3). 866–878.

Keller, Frank, Martin Corley, Steffan Corley, Lars Konieczny & Amalia Todirascu. 1998. Webexp: A Java Toolbox for Web-Based Psychological Experiments. Technical Report HCRC/TR-99. Human Communication Research Centre, University of Edinburgh. Available from [http://www.hcrc.ed.ac.uk/web_-exp/].

Kuznetsova, Alexandra, Per Bruun Brockhoff & Rune Haubo Bojesen Christensen. 2015. Lmertest: Tests in Linear Mixed Effects Models. R package version 2.0-20. *Vienna: R Foundation for Statistical Computing*.

Rabanus, Stefan, Eva Smolka, Judith Streb & Frank Rösler. 2008. Die mentale Verarbeitung von Verben in idiomatischen Konstruktionen. *Zeitschrift für Germanistische Linguistik* 36(1). 27–47.

Redewendungen, Duden. 2002. Wörterbuch der deutschen idiomatik, 2., neu bearbeitete und aktualisierte auflage. *Dudenverlag: Mannheim, Leipzig, Wien, Zürich* 11. 956.

Sakamoto, Yosiyuki, Makio Ishiguro & Genshiro Kitagawa. 1986. *Akaike Information Criterion Statistics*. D. Reidel Publishing Company.

Shapiro, Lewis P, Edgar Zurif & Jane Grimshaw. 1987. Sentence processing and the mental representation of verbs. *Cognition* 27(3). 219–246.

Smolka, Eva, Matthias Gondan & Frank Rösler. 2015. Take a stand on understanding: Electrophysiological evidence for stem access in German complex verbs. *Frontiers in Human Neuroscience* 9. 62.

Smolka, Eva & Gary Libben. 2017. 'Can you wash off the hogwash?'–semantic transparency of first and second constituents in the processing of German compounds. *Language, Cognition and Neuroscience* 32(4). 514–531.

Smolka, Eva, Gary Libben & Wolfgang U Dressler. 2019. When Morphological Structure Overrides Meaning: Evidence from German Prefix and Particle Verbs. *Language, Cognition and Neuroscience* 34(5). 599–614.

Smolka, Eva, Katrin H Preller & Carsten Eulitz. 2014. 'Verstehen' ('understand') primes 'stehen'('stand'): Morphological Structure Overrides Semantic Compositionality in the Lexical Representation of German Complex Verbs. *Journal of Memory and Language* 72. 16–36.

Smolka, Eva, Stefan Rabanus & Frank Rösler. 2007. Processing verbs in German idioms: Evidence against the Configuration Hypothesis. *Metaphor and Symbol* 22(3). 213–231.

Sprenger, Simone A, Willem JM Levelt & Gerard Kempen. 2006. Lexical access during the production of idiomatic phrases. *Journal of Memory and Language* 54(2). 161–184.

Swinney, David A & Anne Cutler. 1979. The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior* 18(5). 523–534.

Tesnière, Lucien. 1959. Elements of structural syntax, translated by timothy osborne and sylvain kahane. In. John Benjamins, Amsterdam.

Thompson, Cynthia K, KL Lange, Sandra L Schneider & Lewis P Shapiro. 1997. Agrammatic and non-brain-damaged subjects' verb and verb argument structure production. *Aphasiology* 11(4-5). 473–490.

Titone, Debra A & Cynthia M Connine. 1999. On the compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics* 31(12). 1655–1674.

Titone, Debra & Maya Libben. 2014. Time-dependent effects of decomposability, familiarity and literal plausibility on idiom priming: A cross-modal priming investigation. *The Mental Lexicon* 9(3). 473–496.

Venables, William N & David M Smith. 2012. R core team (2012). *An Introduction to R [Notes on R: A Program]*.

Winter, Bodo. 2013. Linear models and linear mixed effects models in R with linguistic applications.: tutorial 11. *arXiv:1308.5499. [http://arxiv.org/pdf/1308.5499.pdf]*.

Name index

The role of constituents in multiword expressions

Multiword expressions (MWEs), such as noun compounds (such as *nickname* in English, and *Ohrwurm* in German), complex verbs (such as *give up* in English, and *aufgeben* in German) and idioms (such as *break the ice* in English, and *das Eis brechen* in German), may be interpreted literally but often undergo meaning shifts with respect to their constituents. Theoretical, psycholinguistic as well as computational linguistic research remain puzzled by when and how MWEs receive literal vs. meaning-shifted interpretations, what the contributions of the MWE constituents are to the degree of semantic transparency (i.e., meaning compositionality) of the MWE, and how literal vs. meaning-shifted MWEs are processed and computed.

This edited volume presents an interdisciplinary selection of six papers on recent findings across linguistic, psycholinguistic, corpus-based and computational research fields and perspectives, discussing the interaction of constituent properties and MWE meanings, and how MWE constituents contribute to the processing and representation of MWEs. The collection is based on a workshop at the 2017 annual conference of the German Linguistic Society (DGfS) that took place at Saarland University in Saarbrücken, Germany.

ISBN 978-3-96110-184-9