Constituents in multiword expressions: What is their role, and why do we care?

Sabine Schulte im Walde
Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart

Eva Smolka
Fachbereich Linguistik, 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) have remained an unsettled issue over the past 20+ years.

Our research question concerns semantically transparent MWEs as well as MWEs 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 *Nudelsuppe* ('noodle soup') to semantically opaque, as in *Spitzname* ('nickname', lit. 'pointy name'), *Geduldsfaden* ('patience', lit. 'patience thread'), or *Zwickmühle* ('dilemma', lit. 'pinch mill'), which contain a modifier (i.e. the left constituent) and/or a head (i.e. the right

Sabine Schulte im Walde & Eva Smolka. 2020. Constituents in multiword expressions: What is their role, and why do we care? In Sabine Schulte im Walde & Eva Smolka (eds.), *The role of constituents in multiword expressions: An interdisciplinary, cross-lingual perspective*, iii–xix. Berlin: Language Science Press. DOI:10.5281/zenodo.3598548

constituent) that render 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 seven contributions in this volume discuss multiword expressions 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 in existing and novel compounds (e.g., *nickname*, and *campeel*, respectively), the combination of stem+stem+suffix in deverbal compounds (e.g., *budget assessment*), the combination of stem+preposition+stem in noun compounds (e.g., *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 manipulations 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 effect, 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; 2009; Meunier & Longtin 2007; Marslen-Wilson et al. 2008; 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 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; 2019). 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), and 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 figurative 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), *nach* (Haselbach 2011) and *an* (Springorum 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 characteristics 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 & 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 (i.e., 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 verbal 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

Compounds also lie on a continuum between relatively transparent and rather opaque with respect to the meanings 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 compounds' 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 1975; 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:

1. 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

2. 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 under 2. 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 and productivity (Bell & Schäfer 2016; Schulte im Walde et al. 2016).

In this collection, **Pezzelle & 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 & 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.

In the third chapter on compounds, **Libben** investigates English compounds from a psycholinguistic perspective. He uses novel compounds such as *ankle-cob* and *clampeel*, the former being unambiguous, the latter being ambiguous in the way they can be parsed (i.e. *ankle-cob* versus *clam-peel* or *clamp-eel*, respectively). A typing experiment shows that the typing latencies indeed peak at the morpheme boundary of non-ambiguous compounds. Equivalent latencies at the critical letters of ambiguous compounds indicate that they are parsed in both possible reading ways. Libben refers to the heuristics of his Fuzzy Forward Lexical Activation account, which assumes that MWEs are parsed from left to right for any possible word combination. He concludes that complex words are not static representations but rather patterns of actions.

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 & 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, 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 literal vs. idiomatic interpretations. A large number of automatic classification approaches addressed idioms as non-literal language 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 the constituents of MWEs and their meaning contributions.

The last paper by **Smolka & Eulitz** deals with idioms and how the meaning of the constituents contributes 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 (University of Sheffield, UK / Federal University of Rio Grande do Sul, Brazil)

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

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.

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

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. Boston.

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(3). 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. San Diego, CA: Academic Press.

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

Caillies, Stéphanie & 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. Cambridge, MA, USA.

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. 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.

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.

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: MIT Press.

Günther, Fritz, Eva Smolka & Marco Marelli. 2018. ''Understanding" differs between English and German: Capturing systematic language differences of complex words. *Cortex* 116(10). 158–175. DOI: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(2-3). 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. (Doctoral dissertation).

Holsinger, Edward & Elsi Kaiser. 2013. Processing (non)compositional expressions: Mistakes and recovery. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 39(3). 866–878.

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). 277–288.

Kamp, Hans & Uwe Reyle. 1993. *From discourse to logic*. Dordrecht: 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.

Lapata, Maria. 2002. The disambiguation of nominalizations. *Computational Linguistics* 28(3). 357–388.

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

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

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(1). 50–64.

Longtin, Catherine-Marie, Juan Segui & Pierre Hallé. 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(8). 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 2014 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 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 977–983. Denver, CO.

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 & 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*, vol. 1, 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 (Article 62). DOI:10.3389/fnhum.2015.00062

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. 2019. When morphological structure overrides meaning: Evidence from German prefix and particle verbs. *Language, Cognition and Neuroscience* 34(5). 599–614. DOI: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 Association for Computational Linguistics*, 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 expressions. *Journal of Verbal Learning and Verbal Behavior* 18(5). 523–534.

Taft, Marcus & Kenneth Forster. 1975. Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior* 14. 638–647.

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 A. & Cynthia M. Connine. 1999. On the compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics* 31(12). 1655–1674.

Titone, Debra A. & 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.

Turney, Peter D., Yair Neuman, Dan Assaf & Yohai Cohen. 2011. Literal and metaphorical sense identification through concrete and abstract context. In *Proceedings of the 2011 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(1). 141–188.

Zwitserlood, Pienie. 1994. The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes* 9(3). 341–368.