Tagungsbericht / Conference report

"Wen wurmt der Ohrwurm? An Interdisciplinary, Cross-Lingual Perspective on the Role of Constituents in Multi-Word Expressions"
(Workshop at the 39th DGfS Annual Conference „Information und sprachliche Kodierung", Saarbrücken, 08.–10.03.2017)

Organized by Sabine Schulte im Walde (University of Stuttgart) and Eva Smolka (University of Konstanz) as part of the 39th Annual Conference of the German Linguistic Society (DGfS) held at the Saarland University in Saarbrücken, Germany, the workshop aimed "to shed light on the interaction of constituent properties and compound transparency across languages and disciplines integrating linguistic, psycholinguistic, corpus-based and computational studies". The workshop brought together researchers from linguistics, psycholinguistics, and natural language processing and comprised 11 contributed talks, framed by two invited talks by Gary Libben and Marco Marelli. Most of the slides are available from the workshop's homepage at "http://www.ims.uni-stuttgart.de/events/dgfs-mwe-17/program.html".

With the majority of talks centering on compounding, the wide variety of approaches presented soon proved an asset of the workshop. The very fundamental issue of the nature of compositionality in multi-word expressions implicitly or explicitly played a role in all contributions. Translated into a wide variety of tasks and study designs, it was consequently tackled in a multitude of different ways. Thus, while a number of the presentations focused on the effects associated with individual constituents and their properties, others focused on the best quantitative correlate of compositionality. The relationship between compositional expression meanings and semantically transparent meanings, often implicitly or explicitly equated, also came under scrutiny and stimulated the discussion.

In the first of two invited talks, Gary Libben's (Brock University) *Morphological superposition and the nature of the mental lexicon* argued for a reconceptualization of the mental lexicon, focusing particularly on the dynamic and interconnected nature of lexical representations. In what he dubbed as Lexical Field Theory, it is not so much the single and static word, but rather the lexical system as a whole that accounts for the measurable psycholinguistic features. Linking his view to early proposals by the psychologist William James in the late 19th century, he presented three sets of supporting

evidence: A study by Libben et al. (2016) on the string *-ational*, experiments using ambiguous novel compounds (e.g. *feedraft*: either *fee-draft* or *feed-raft*), and a study by Teddiman & Libben (2015) assessing the effect of splitting compounds not on their morpheme boundaries. Based on these findings, he argued for the notion of morphological superstates: lexical representations are indefinite until their actual usage creates a specific morphological state.

Stefanie Rößler, Thomas Weskott and Anke Höller (University of Göttingen) discussed three experiments exploring *N1-accessibility as a matter of compound processing*. Two sentence completion tasks showed facilitatory effects for animacy and compounds requiring spatio-temporal contiguity of the referents of the two constituents (cf. *car insurance* versus *car accident*: only the denotation of the latter must encompass an actual car): in both cases, participants are more likely to pick up the first constituent's referent with a following personal pronoun. Differences between synthetic and root compounds were also tested, hypothesizing that the different status of the N1 as either a root or a noun in the framework of distributional morphology should facilitate pronominal pickup of the latter. This was not the case, leading to a final eye-tracking study honing in on the difference between these two compound types. Synthetic compounds led to shorter total reading times than root compounds, reflecting a processing advantage in line with the authors' predictions.

Serkan Uygun (Yeditepe University) and Ayşe Gürel (Boğaziçi University) presented *Factors affecting the processing of compounds in the second language*. Using a lexical decision task in combination with masked priming, they compared the differences between native speakers of Turkish with intermediate and advanced L2 speakers of Turkish. Compounds were distinguished as either transparent-transparent or partially opaque. Whereas no priming effect was obtained for intermediate or advanced second language learners, the native speakers showed priming effects for both compound categories. Interestingly, while for the partially opaque compounds priming either the first or the second constituent both led to faster reaction times, the transparent-transparent combinations showed a significant priming effect only for the second constituent, although descriptively the first constituent showed a similar pattern.

Anna Hätty (Robert Bosch GmbH) and Michael Dorna (University of Stuttgart) presented *Exploring the impact of transparency and productivity of multiword term constituents on single-word term identification*, adapting Kaguero and Umino's (1996) characterization of terms as "linguistic units which characterize specialized domains". This talk presented an interesting twist to the common approach to compounds and multi-word expressions where the multi-word expressions are the phenomena under investigation. Here, instead, the question was in how far multi-word terms can be exploited

to identify single-word terms. Exploring the similarity between the multi-word term and the embedded single word terms as well as the variance of the embedded single-word terms, they found that the similarity measure showed little effect, but the variance of the head showed an effect.

In *More than words. A discriminative learning model with lexical bundles,* Saskia E. Lensink (Leiden University) and Harald Baayen (University of Tübingen) pointed out that many studies on multi-word units prominently focus on frequency, although it is known that it is an impoverished measure that collapses word senses and disregards context. Further, they noted that neighborhood density effects are well known to influence the processing of words, but are typically disregarded when investigating multi-word units. Using the naïve discriminative learning model proposed in Baayen et al. (2011), they aimed to capture neighborhood density effects through the inclusion of lexical bundles, units of three words, into the training of their naïve discriminative learning model. They showed that the inclusion of these measures into regression models of eye-tracking data yield a number of significant predictors for several eye-tracking measures. Similary, lexical-bundle derived predictors were also found in a production experiment.

In the first day's last talk, *Semantic entropy measures and the semantic transparency of noun noun compounds,* Melanie J. Bell (Anglia Ruskin University) and Martin Schäfer (University Jena) reported on the modeling of human semantic transparency judgments with the help of semantic entropy measures. Based on a large annotated database of compound constituent families created for Bell & Schäfer (2016), their relation entropy measure reflects the uncertainty with respect to the semantic relation between constituents, while the synset entropy measure reflects the uncertainty with regard to different readings of the constituents in compounds. They reported an interaction of the two head-based entropy measures: when the synset entropy is low, the relational entropy does not make much of a difference, but when the synset entropy is high, relation entropy is negatively correlated with compound transparency. The most transparent compounds have high synset entropy but low relation entropy.

The second day began with *The Role of the Head in the Interpretation of English Deverbal Compounds* by Gianina Iordăchioaia (University of Stuttgart), Lonneke van der Plas (University of Malta), and Glorianna Jagfeld (University of Stuttgart). Starting with the observation that deverbal compounds are often hard to interpret, they identified the pervasive ambiguity of the deverbal head noun as a possible source for this difficulty. In a corpus study, they followed the hypothesis that deverbal nouns in compounds have the same preference for a specific reading as they have outside of compounds, using Grimshaw's (1990) distinction between argument structure and resultative interpretations for a classification of out-of-compound usages of the

deverbal heads. They found that the frequency of realization of internal arguments of the head noun outside of a compound correlates with compound interpretations in which the first noun serves as the object. Furthermore, the number of occurrences of a given head noun in different compounds similarly correlates with object interpretations.

The talk by Inga Hennecke (University of Tübingen) on *Semantic transparency and variation in nominal syntagmatic compounds in Romance languages* started with the observation that Romance nominal syntagmatic compounds, exemplified by noun-preposition-noun constructions, are either analyzed as products of word formations or of syntax, and given a host of different names. Extracting noun-preposition-noun constructions from French, Spanish, and Portuguese corpora, and combining quantitative analysis in terms of frequency and productivity with qualitative analysis in terms of the semantic relations between the nouns, she found that Spanish exhibits the lowest amount of prepositional variation between nouns and also the fewest types of semantic relations. French takes the middle place, interestingly showing a divergence between corpus data and human acceptability judgements. Finally, the pattern is productive and fully acceptable in Portuguese. She concluded that the preposition is more than a pure linking element, being frequently linked to specific semantic relations, with the specific link and its productivity varying across the three languages.

In *Compounding in Context*, Melanie J. Bell (Anglia Ruskin University) explored the hypothesis that compounds can often be disambiguated on the basis of their immediate sentential context, usually by disambiguating the head noun. Based on preliminary results from a study contrasting free paraphrases of novel compounds context-free and in their sentential contexts, she reports that grouped readings can indeed be linked to specific constituent senses. She also showed that presenting a compound in its sentential context does not necessarily clarify its interpretation but still allows for considerable variation in its interpretation. This finding, she pointed out, links up nicely with the ideas on morphological indeterminacy presented in Libben's talk.

In *Approximating compound compositionality based on Word alignments*, Fabienne Cap (Uppsala University) explored the use of word alignments in assessing compound compositionality. Aligning the constituents of German compounds with their translation counterparts in English, she calculated the translational entropy of the compounds, assessing the variance in its translation equivalents. Next, she compared the correlation of this measure with human ratings on a set of compounds, and against the correlation of the vector-based approaches presented in Schulte im Walde, Hätty, and Bott (2016) for these very same ratings. With alignment variance correlating weakly with compositionality, head variance is a better indicator then

modifier variances. Data sparsity remains a challenge for this approach, but might be overcome by incorporating alignments across multiple languages.

Marco S. G. Senaldi (Scuola Normale Superiore, Pisa), Gianluca E. Lebani, and Alessandro Lenci (both University of Pisa) presented a dense talk titled *Exploring Idiomaticity with Variant-based Distributional Measures and Shannon Entropy*. In the first part, they report on an idiom type identification task for Italian verb-noun and adjective-noun combinations. Using synonyms of the tokens making up the respective combinations, they constructed variants and measured the similarity between the variants and the target construction using distributional semantic models. From these measurements, they derived compositionality indices capturing the idiomaticity of the target constructions. In the second part, they reported on predicting human ratings of idiom syntactic flexibility with the help of the best performing index, a formal entropy measure and frequency data. Their best model included all predictors, with entropy and the compositionality index positively correlating with perceived flexibility, and frequency negatively correlating with perceived flexibility.

In *Evaluating [the] semantic composition of German compounds*, Corina Dima, Jianqiang Ma, and Erhard Hinrichs (University of Tübingen) first summarized the steps taken to arrive at the best distributional semantics composition function for German noun-noun compounds (cf. Dima 2015). This best performing composition function uses masks, or weights, sensitive to the position of the nouns to be combined as either modifier or head of the compounds. In a second step, they used the shared subset of compounds of their dataset and the dataset in Schulte im Walde, Müller, and Roller (2013) in order to compare the performance of their composition function against the human transparency ratings on the compounds. This allows them to identify ambiguous heads and modifiers as a major source of problems of the composition models, suggesting the usage of sense-aware word representations.

The workshop closed with the second keynote by Marco Marelli, who spoke about *Understanding compound words: A new perspective from compositional systems in distributional semantics*. He introduced a distributional semantics approach called CAOSS: Compounding as Abstract Operation in Semantic Space. In this approach, the semantic representations for the constituents of a given compound are updated via weights derived from previous training on the corpus, similar to the method explored in Guevara (2011). Just as the best performing function of the Dima et al. talk, these updates are position-sensitive, that is, the weights used for modifier and head are different. This model is then evaluated against four sets of experimental data, first on the effects found for novel compounds in Gagné & Shoben (1997) and Gagné (2001), secondly on familiar compounds, using the lexical decision data from the English Lexicon Project and the results of an eye-tracking experiment.

While the experimental data on novel compounds could be straightforwardly mirrored in CAOSS models, two different models were built for the familiar compounds, each involving complex interactions. Surprisingly, there was an interesting task effect involving the CAOSS stand-in for compound compositionality, that is, the similarity between the compound vector derived by the composition function and the actual vector of the compound found in the corpus: while for the off-line lexical decision data compound compositionality, in interaction with the semantic contribution of the head, led to longer response time, it led to faster first gaze duration in the eye-tracking experiment. Interpreting this difference, Marelli argued that reading requires the selection of a specific sense and hence any conflicts between compositional meaning and lexicalized meaning need to be resolved, leading to an advantage for compositional compounds. In contrast, when the task just requires to decide whether something is a word or not, all activation helps and no final resolve is required.

All in all, the wide variety of different approaches provides fruitful avenues for further research. Consider e.g. the usage of word alignments across languages to assess compositionality in Fabienne Cap's talk as opposed to language internal distributional semantics measures and human ratings of transparency: investigating in more detail where and how these measures diverge is an important step towards a better understanding of the notion of compositionality. A further point highlighted by the workshop are the differentiated roles of modifiers and heads, leading to the positional weighting in distributional models, but also in line with recent psycholinguistic evidence for differentiated constituent effects of heads and modifiers (cf. Smolka & Libben 2017). The importance of theoretical accounts of multi-word constructions was underscored by their usefulness in deriving carefully set-up hypothesis-driven experiments. Finally, and perfectly illustrated by the two framing invited talks, the indeterminacy of meaning in multi-word constructions and the task-dependency of the usage of specific strategies to access and construct meanings is still in need of much more detailed investigation.

References

Baayen, R. Harald, Peter Milin, Dusica Filipovic Durdević, 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.

Bell, Melanie J. & Martin Schäfer. 2016. Modelling semantic transparency. *Morphology* 26(2). 157–199.

Dima, Corina. 2015. Reverse-engineering Language: A Study on the Semantic Compositionality of German Compounds. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (EMNLP 2015), Lisbon,

Portugal, 1637-1642. Association for Computational Linguistics. https://aclweb. org/anthology/D/D15/D15-1188.pdf (accessed 26 May 2017).

Gagné, Christina L. 2001. Relation and lexical priming during the interpretation of noun-noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 236–254.

Gagné, Christina L. & E. 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.

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

Guevara, Emiliano. 2011. Computing semantic compositionality in distributional semantics. *Proceedings of the Ninth International Conference on Computational Semantics*. 135-144. Association for Computational Linguistics. http://dl.acm.org/citation.cfm?id=2002669.2002684 (accessed 26 May 2017).

Kageura, Kyo & Bin Umino. 1996. Methods of automatic term recognition. A review. *Terminology*. 3(2). 259–289.

Libben, Gary, Gonia Jarema, Bruce Derwing, Alessandra Riccardi & Danuta Perlak. 2016. Seeking the –ational in derivational morphology. *Aphasiology*. 30(11). 1304–1324.

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* (*SEM). Atlanta, GA. Association for Computational Linguistics. https://*www.aclweb.org/anthology/S13-1038 (accessed 26 May 2017)*.

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* (*SEM). Berlin, Germany. Association for Computational Linguistics. *https://aclweb.org/anthology/S/S16/S16-2020.pdf (accessed 26 May 2017)*.

Smolka, Eva, & Libben, Gary. 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.

Teddiman, Laura & Libben, Gary. 2015. Segmented binaural presentation as a means to examine lexical substructure. *The Mental Lexicon*. 10(3). 435–457.

Dr. Martin Schäfer

Friedrich-Schiller-Universität Jena

Institut für Anglistik/Amerikanistik

Ernst-Abbe-Platz 8

D-07743 Jena

post@martinschaefer.info