



WORD FORMATION AND DISCOURSE STRUCTURE

LEIPZIG
MAY 5 & 6, 2022



BOOK OF ABSTRACTS



UNIVERSITÄT
LEIPZIG

INSTITUT FÜR
GERMANISTIK





Contents

| | |
|--|-----------|
| Timetable | 2 |
| Thursday, 5 of May | 2 |
| Friday, 6 of May | 3 |
| Abstracts | 4 |
| Thursday, 5 of May | 4 |
| Anaphoric connectives: Discourse processing at its best! (M. Stede) | 4 |
| Word-formation of light verb-constructions and its role in German texts. A corpus-based study (M. Werner, S.M. Beiter) | 4 |
| The discourse status of verbal gerunds (Z. Huang) | 6 |
| Im Einklang mit der Konstruktion: a diachronic corpus analysis of deverbal nominalizations in prepositional phrases (F. Fleißner, R. Ruf, E. Smirnova) | 8 |
| (Con)textaholic: a corpus-driven study on Italian blends among textuality and word formation (M. Le Donne) | 10 |
| Personal name compounding in German. On the relationship between productivity and pragmatic functions across discourse domains and text types (M. Belosevic, S. Arndt-Lappe) | 12 |
| Thursday, 5 of May | 14 |
| A Proposal for Explicit Word Formation Annotation in Discourse Corpora (M.-C. Müller) | 14 |
| Improving Coreference Resolution with Word Formation (W. Zhao, K. Mathews, H. Chai) | 16 |
| Affixes and their role in discourse structure (A. Baltuttis) | 17 |
| Word formation and cohesion in specialized languages (K. Menzel) | 19 |
| Complex nouns as markers of academic register in L1-and L2-authored essays (J. Lukassek, A. Lüdeling, A. Shadrova, S. Wan) | 21 |
| Partners & Sponsors | 23 |
| Organizers | 23 |
| Contact | 23 |
| Sponsors | 23 |
| Space for your notes | 24 |

Timetable

Thursday, 5 of May

| | | |
|-------------|--|---|
| 8:30–9:00 | | Registration |
| 9:00–9:15 | | Introduction |
| | <i>Chair: Barbara Schlücker</i> | |
| 9:15–10:15 | Manfred Stede Potsdam | Anaphoric connectives: Discourse processing at its best! |
| 10:15–11:00 | Martina Werner & Sophia M. Beiter Vienna | Word-formation of light verb-constructions and its role in German texts. A corpus-based study |
| 11:00–11:30 | | Coffee |
| | <i>Chair: Maximilian Frankowsky</i> | |
| 11:30–12:15 | Zi Huang Barcelona | The discourse status of verbal gerunds |
| 12:15–13:00 | Fabian Fleißner, Regina Ruf & Elena Smirnova Neuchâtel | Im Einklang mit der Konstruktion: a diachronic corpus analysis of deverbal nominalizations in prepositional phrases |
| 13:00–14:30 | | Lunch |
| | <i>Chair: Adele Baltuttis</i> | |
| 14:30–15:15 | Mauro Le Donne Perugia | (Con)textaholic: a corpus-driven study on Italian blends among textuality and word formation |
| 15:15–16:00 | Milena Belosevic & Sabine Arndt-Lappe Trier | Personal name compounding in German. On the relationship between productivity and pragmatic functions across discourse domains and text types |
| 16:00–16:30 | | Coffee |
| 19:30 | | Conference dinner |

Friday, 6 of May

Chair: Anna Bliß

| | | |
|------------|--|---|
| 9:00–9:45 | Mark-Christoph Müller Heidelberg | A Proposal for Explicit Word Formation Annotation in Discourse Corpora |
| 9:45–10:30 | Wei Zhao, Kevin Mathews & Haixia Chai Heidelberg | Improving Coreference Resolution with Word Formation |

10:30–11:00 **Coffee**

Chair: Barbara Schlücker

| | | |
|-------------|--|---|
| 11:00–11:45 | Adele Baltuttis Leipzig | Affixes and their role in discourse structure |
| 11:45–12:30 | Katrin Menzel Saarbrücken | Word formation and cohesion in specialized languages |
| 12:30–13:15 | Julia Lukassek, Anke Lüdeling, Anna Shadrova & Shujun Wan Berlin | Complex nouns as markers of academic register in L1-and L2-authored essays |

End

Abstracts

Thursday, 5 of May

Anaphoric connectives: Discourse processing at its best!

M. Stede

Invited speaker

University of Potsdam, Potsdam, Germany

In computational linguistics, the two most central subtasks of “discourse processing” are coreference resolution and the identification of coherence relations (e.g., as building blocks for a representation of discourse structure). While generally regarded as two clearly distinct tasks, they meet in the presence of anaphoric connectives: lexical items that prompt the addressee to recover an antecedent and form a coherence relation holding between two arguments. This phenomenon is widespread in German, which offers a range of adverbial connectives like “trotzDEM”, “DEMzufolge” or “DEShalb”. After a general overview of the various ambiguity problems introduced by these items, the talk will present observations from a corpus study on “demzufolge” and its disambiguation, as well as results from a recent controlled experiment on antecedent selection for “trotzdem”.

Word-formation of light verb-constructions and its role in German texts. A corpus-based study

M. Werner^{1,2}, S.M. Beiter²

¹ ÖAW, Vienna, Austria

² University of Vienna, Vienna, Austria

Light-verb constructions (LVCs) are a frequent phenomenon of present-day German (PDG). LVCs are traditionally defined as verbal periphrases with a semantically bleached verb (such as *bringen*, *kommen*), an abstract noun (i.e., a nominalization), and a preposition (cf. e.g. Polenz 1963 und 1987, So 1990, Kamber 2008). Examples are given in (1).

- (1) a. *zur Aufführung bringen* ‘to be performed’
b. *ins Rollen kommen* ‘to start rolling’

Though intensive investigation for the last 30 years (e.g. Polenz 1963, 1987), the morphological patterning of LVCs in PDG has many unsolved questions such as selection of light-verbs (e.g. *bringen* vs. *kommen*) and prepositions (*zu(r)/in(s)/...*), morphological representation of abstract nouns

(e.g., *-ung*-nouns like *Aufführung* in (1-a) or nominalized infinitives like *Rollen* in (1-b), predicate modification (e.g., via adverbs), and putative noun-modification (via compounding, adjectives or attributive phrases) (cf. Yuan 1986, So 1990, Pottelberge 2001, Kamber 2008, Körösi 2010). We present two corpus-based studies on LVCs and their interactions with text in the Austrian Media Corpus (AMC¹).

Our first study serves the identification of LVCs and the role of textual reference by focusing on light-verbs as well as on PDG deverbal nominalization patterns involved in LVC (as in (1)). Our data suggest that a contextual identification of LVCs in PDG is inalienable given that a) the light-verbs in question are formally indistinct from full lexemes (*bringen* 'bring', *kommen* 'come'), b) a major part of the involved nominals, especially *-ung*-nouns, tend to lexicalization (Demske 2000, Werner 2020, among others), and c) a strictly clause-based interpretation is often insufficient. An example illustrating the ambiguity between a LVC-interpretation in (a) vs. a lexeme-based interpretation (in b) is given in (2). Our talk identifies the over-individual factors which trigger the proper interpretation of LVCs in the individual context vs. which are responsible for lexeme- or reference-based interpretation.

(2) *Er bringt sie zur Kreuzung.*

- a. 'He begins cross-breeding them (e.g. animals, plants).'
- b. 'He brings her/them to the crossroads.'

LVCs are considered to only allow very restricted modification which correlates with the degree of lexicalization of the LVC (Polenz 1987: 176; So 1990: 46-53). For proofing this, the LVCs received from study 1 were morphologically analyzed w.r.t. the occurrence of compounding and adjectival modification of the LVC-nominal. Study 2 shows, however, that the majority of (non-)lexicalized LVCs contains modifiers.

Furthermore, we investigated the textual function of modifiers by focusing on underinvestigated adjectival modification (noticed in Lipka 1987:60–61). It will be shown that adjectival modification is constitutive for cohesion and coherence, which confirms the findings of other studies on the impact of word-formation processes related to modification in texts (see e.g. Dressler & Mörth 2012 on compounding).

References:

- Demske, Ulrike. 2000. Zur Geschichte der *-ung*-Nominalisierung im Deutschen. Ein Wandel morphologischer Produktivität. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 122. 365-411.
- Dressler, Wolfgang U. & Mörth, Karlheinz. 2012. Produktive und weniger produktive Komposition in ihrer Rolle im Text anhand der Beziehungen zwischen Titel und Text. In Gaeta, Livio & Schlücker, Barbara (eds.), *Das Deutsche als kompositionsfreudige Sprache*. Berlin/New York: de Gruyter. 219-234.

¹See: <https://amc.acdh.oeaw.ac.at/about-amc/> [11/15/2021]. AMC is one of the largest corpora of present-day Austrian Standard German, currently comprising more than 45 million of all press articles (11 billions of words) released in Austria since 1990 until today.

- Kamber, Alain. 2008. *Funktionsverbgefüge: empirisch*. Tübingen: Niemeyer.
- Körösi, Veronika. 2010. Datenbank zur Analyse deutscher Sätze an der Grenze zwischen regulären Stützverbkonstruktionen und idiomatischen Satzrahmen. In *Studien zur Informations- und Sprachverarbeitung* 13. München.
- Lipka, Leonhard. 1987. Word-formation and text in English and German. In Asbach- Schnitker, Brigitte & Roggenhofer, Johannes (eds.), *Neuere Forschungen zur Wortbildung und Historiographie der Linguistik*. Tübingen: Narr. 59-67.
- Polenz, Peter von. 1963. Funktionsverben im heutigen Deutsch. Sprache in der rationalisierten Welt. In *Beiheft 5 „Wirkendes Wort“*.
- Polenz, Peter von. 1987. Funktionsverben, Funktionsverbgefüge und Verwandtes. Vorschläge zur satzsemantischen Lexikographie. *Zeitschrift für Germanistische Linguistik* 15(2). 169- 189.
- Pottelberge, Jeroen van. 2001. *Verbonominale Konstruktionen, Funktionsverbgefüge: vom Sinn und Unsinn eines Untersuchungsgegenstandes*. Heidelberg: Winter.
- So, Man-Seob. 1990. *Die deutschen Funktionsverbgefüge in ihrer Entwicklung vom 17. Jahrhundert bis zur Gegenwart. Eine sprachhistorische Untersuchung anhand von populärwissenschaftlichen Texten*. Trier: WVT Wissenschaftlicher Verlag.
- Werner, Martina. 2020. Korpuslinguistische Perspektiven auf die sprachhistorische Entwicklung der nominalisierten Infinitive im Deutschen. In *Linguistik online* 102 (2) 2020. 155-181.
- Yuan, Jie. 1986. *Funktionsverbgefüge im heutigen Deutsch: eine Analyse und Kontrastierung mit ihren chinesischen Entsprechungen*. Heidelberg: Groos.

The discourse status of verbal gerunds

Z. Huang

Universität Pompeu Fabra, Barcelona, Spain

There are two verbal gerunds in English: **POSS-ing** (1) and **ACC-ing** (2).

- (1) Clay's/his winning the game
- (2) Clay/him winning the game

There is debate about whether the subject of *POSS-ing*, assigned genitive case, is semantically a possessor. This study attempts to see whether they tend to have different discourse status due to *POSS-ing* being a possessive structure and *ACC-ing* not.

In the semantic literature on English gerunds, it has been claimed that *POSS-ing* is definite and *ACC-ing* indefinite (Portner 1992). It follows that *POSS-ing* should be familiar in the discourse. In the literature on possessives, however, there are contradictory predictions as to whether possessee

referents should be given. A corpus study by Willemse et al. (2009) shows that the discourse status of possessive structures ranges from given to brand new, with most of the event nominals being brand new in the discourse.

I collected a sample of 205 cases from all the POSS-*ing* from a dependency parsed version of the British National Corpus (2007), and took the first 200 cases of ACC-*ing* from the same corpus. I annotated the verbal gerunds with a scheme based on Baumann & Riester (2012): referential givenness of the event described by the gerund, and both referential and lexical givenness of the subject and object of the gerund. I also annotated for whether the gerund refers to an event type or a token (see Grimm & McNally 2015). Below is an example:

- (3) [...] “Naylor tells me you’re one of the best executives in your particular line,” Cicely Hepwood, a neat and gentle woman, remarked at one point during the meal.
Leith shot a glance at Naylor, seated next to her, who wasn’t even a tinge pink around the ears at his aunt’s revealing what, since there were others in her particular line who were far more senior, must surely be a lie.

| Type/token | Gerund | Subj | Subj | Obj | Obj | Pred |
|------------|-------------|-------------|-------------|-------------|-------------|------------------|
| | r-givenness | r-givenness | l-givenness | r-givenness | l-givenness | r-givenness |
| token | token-text | given | new | given | new | accessible-other |

The results show that POSS-*ing* (100/205 not new) occurs more in context where it is given in some ways than ACC-*ing* (61/200 not new). However, in more than half of the cases (105/205), POSS-*ing* is still hardly inferable from the context. Among the discourse new cases, more POSS-*ing* contain given elements (object or predicate) than ACC-*ing* and slightly fewer POSS-*ing* cases are entirely new. Most POSS-*ing* have a given subject (196/205), implying that it relies on its possessor as an anchor to introduce new information, while ACC-*ing* has much fewer given subjects (104/200). ACC-*ing* when selected by *with* mostly introduces new information and holds an Elaboration relation to its matrix clause.

The discourse status of POSS-*ing* and ACC-*ing* only show a tendency, not a categorical distinction. In comparison to Willemse et al. (2009), it seems that verbal gerunds are more likely to be given than deverbal nominalizations, which is an unintuitive result. I would like to compare the discourse function of verbal gerunds with that of deverbal nouns and nominal gerunds (*Clay’s winning of the game*) in future research.

References:

- Baumann, S. & Riester, A. 2012. Referential and lexical givenness: Semantic, prosodic and cognitive aspects. In G. Elordieta & P. Prieto (eds.), *Prosody and meaning*. Berlin, Boston: De Gruyter Mouton. 119-162.
- Grimm, S. & McNally, L. 2015. The -*ing* dynasty: Rebuilding the semantics of nominalizations. In S. D’Antonio, M. Moroney & C. R. Little (eds.), *Proceedings of the 25th Semantics and Linguistic Theory Conference (SALT)*. Vol. 25. Ithaca, NY: LSA and CLC Publications. 82-102.

Portner, P. 1992. *Situation Theory and the semantics of propositional expressions* (Unpublished doctoral dissertation). University of Massachusetts at Amherst, Amherst, MA.

The British National Corpus, version 3 (BNC XML Edition). 2007. Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium, URL <http://www.natcorp.ox.ac.uk/>.

Willemse, P., Davidse, K. & Heyvaert, L. 2009. English possessives as reference-point constructions and their function in the discourse. In W. B. McGregor (ed.), *The expression of possession*. Berlin: De Gruyter Mouton.

Im Einklang mit der Konstruktion: a diachronic corpus analysis of deverbal nominalizations in prepositional phrases

F. Fleißner, R. Ruf, E. Smirnova

Universität de Neuchâtel, Neuchâtel, Switzerland

The increasing use of nominalizations in the recent history of German has been repeatedly emphasized in the literature and often described in the context of a general tendency towards the *Nominalstil* (cf. Eggers 1962, Möslein 1981, Khadijeva 2017). In this connection, the textual function of information condensation (Lipka 1987) has been assumed to be the driving force behind this tendency. It remains however to be seen how precisely information condensation is achieved at the text/discourse level. Furthermore, the question remains open as to why this tendency becomes apparent only in the 19th century, even though most word formation patterns have existed since the oldest periods of German. The present study aims for a better understanding of the role of deverbal nominalizations on the level of text and discourse. It reports the first results of a corpus study based on the data from the 19th century (www.deutschestextarchiv.de). Expanding the focus from extensively studied *ung*-nominalizations (*Beendigung* 'ending', *Auswechslung* 'replacement', *Entwicklung* 'development'; see Demske 2000, Hartmann 2016) to all types of deverbal nominalizations, including implicit derivations (*Bezug* 'reference') and conversions (*Erscheinen* 'appearance'), the study looks at their behavior in particular syntagmatic patterns, namely in prepositional phrases with postnominal modifiers, as illustrated in (1) –(3):

- (1) *Er habe erfahren, dass die Botschafter zu **Auswechslung** der Ratificationen über Pe-taŋ nach der Hauptstadt reisen wollten, und deshalb beim **Erscheinen** der Schiffe die Besatzung von da entfernt.* [Berg: Die preussische Expedition, 1873]
'He had learned that the ambassadors wanted to travel to the capital via Pe-taŋ to exchange ratifications and therefore [he] removed the crew at the appearance of the ships.'
- (2) *So kam erst nach glücklicher **Beendigung** des Siebenjährigen Krieges ein frischer Hauch **in die industrielle **Entwicklung** des Landes.*** [Beck: Die Geschichte des Eisens, 1897]
'Thus, only after the fortunate ending of the Seven Years' War did a breath of fresh air come into the industrial development of the country.'
- (3) *Die höhere Bestätigung bleibt vorbehalten **in Bezug** auf die stimmführenden Mitglieder des Magistrats und des Stadtgerichts.* [Dahlmann: Die Politik, 1835]

'The higher confirmation is reserved with regard to the voting members of the Magistrate and the Municipal Court.'

The study is guided by two general hypotheses:

1. The textual function of nominalizations is not limited to information condensation. In addition, it comprises other functions such as referential cohesion and textual connectivity.
2. The textual function of nominalizations is not achieved in isolation, by using a newly coined word in a text. Instead, they have first to be integrated into a particular syntagmatic configuration (= construction).

It will be argued that textual functions arise via a complex interplay between the (elements of the syntactic) construction and the nominalization itself. Over time, certain patterns conventionalize and become independent from the original syntactic configurations, which has been the case for the present-day complex prepositions such as *im Einklang mit* and *in Bezug auf* (cf. Stefanowitsch et al. 2020). More generally, the contribution is meant as a plea for a constructional-based approach, which is seen as bridging discourse-pragmatic distributional patterns and morphosyntactic regularities in mutual dependence.

References:

- Demske, Ulrike. 2000. Merkmale und Relationen. Diachrone Studien zur Nominalphrase im Deutschen. Berlin / New York: De Gruyter.
- Eggers, Hans. 1962. Zur Syntax der deutschen Sprache der Gegenwart. *Studium Generale* 15. 49–59.
- Hartmann, Stefan. 2016. Wortbildungswandel. Eine diachrone Studie zu deutschen Nominalisierungsmustern. Berlin / Boston: De Gruyter.
- Lipka, Leonhard. 1987. Word-formation and text in English and German. In Brigitte Asbach-Schnitker & Johannes Roggenhofer (eds.), *Neuere Forschungen zur Wortbildung und Historiographie der Linguistik: Festgabe für Herbert E. Brekle zum 50. Geburtstag*. Tübingen: Narr. 59–67.
- Möslein, Kurt. 1981. Einige Entwicklungstendenzen in der wissenschaftlich-technischen Literatur seit dem Ende 18. Jahrhunderts. In Walther von Hahn (ed.), *Fachsprachen*. Darmstadt: Wissenschaftliche Buchgesellschaft. 276–319.
- Khadjieva, Shaira. 2017. Nominalisierungstendenzen im Deutschen [PhD Thesis, Ludwig-Maximilians-Universität München]. <https://edoc.ub.uni-muenchen.de/24349/>.
- Stefanowitsch, Anatol, Elena Smirnova & Matthias Hüning. 2020. Complex Adpositions in three West Germanic Languages. In Elena Smirnova & Benjamin Fagard, Dejan Stosic, José Pinto de Lima (eds.), *Complex Adpositions in European Languages*. Berlin / Boston: De Gruyter. 65-138.

Primary sources:

- Beck, Ludwig. 1897. Die Geschichte des Eisens. Bd. 3: Das XVIII. Jahrhundert. Braunschweig: Vieweg.
- Berg, Albert. 1873. Die preussische Expedition nach Ost-Asien. Bd. 3. Berlin: Decker.

Dahlmann, Friedrich Christoph. 1835. Die Politik, auf den Grund und das Maaß der gegebenen Zustände zurückgeführt. Bd. 1: Staatsverfassung. Volksbildung. Göttingen: Dieterich.

(Con)textaholic: a corpus-driven study on Italian blends among textuality and word formation

M. Le Donne

Università per Stranieri di Perugia, Perugia, Italy

This study is a corpus-driven analysis on Italian lexical blends. More specifically, the central focus of this paper is blends' potential to form neologisms and/or *hapaxes* and their textual dimension inside Italian corpora. Lexical blends are lexemes formed by two (or more) source words, provided that at least one source word has been reduced phonologically, e.g., *diversabile* < *diversa(mente)* + *abile*; *fantastiliardo* < *fantasti(co)* + *(m)iliardo* (Castagneto & Parente, 2020: 353-354). The interest in Lexical Blending increased notably in the last two decades. However, in the Italian tradition, lexical blends appear to be only occasionally studied still nowadays. Few studies (Thornton, 1993, 2004; Bertinetto, 2001; Cacchiani, 2011, 2016; Mattiello, 2013, 2019; Bombi, 2015; Castagneto & Parente, 2020) turned their attention on such formations, probably because considered "marginal" or "extra-grammatical" with regard to other word formation processes (Scalise, 1983; Dressler, 1987). Conversely, recent studies adopting different methodological approaches and frameworks have proved that blends' coinage is neither random, nor unpredictable (Kemmer, 2003; Gries, 2004; Piñeros, 2004; Bauer, 2012; Renner & Lalić-Krstin, 2011). Moreover, there is now evidence of this word formation process in typologically diverse languages (Renner, 2015: 121-122). Nevertheless, blends' textual dimension has always been scarcely explored or subsumed under the general domain of abbreviations, i.e., together with other neighbouring word formation processes, such as acronyms and clippings (McArthur, 1992; Fandrych, 2004; Kjellander, 2015). In this paper, blends' occurrences are computed scrutinising different (mostly web-based) corpora consulted through the software *Sketch Engine* (Kilgariff et al., 2014). Following a suggestion from the past literature (Lehrer, 1996, 2007; Ronneberger-Sibold, 2006; Kjellander, 2015), namely that blends are better decoded out of their context, the occurrences of 10 lexical blends are examined to shed light on their textual functions (viz. how cohesion is achieved), and to better comprehend affinities and divergencies with nominal compounds on this respect (Halliday & Hasan, 1976; Dederding, 1983: 63). In this study it is argued that blends can function as lexical "contextual cues", defined as "any feature of linguistic form that contributes to the signalling of contextual presuppositions" (Gumperz, 1982: 131; Levinson, 2003). Doing so, it is expected that the context in which they appear will serve to decode blends' constituents and to situate their "intralinguistic" meaning that otherwise relies solely on the encyclopaedic knowledge of the reader/hearer (Kjellander, 2015: 9). The examination of blends' occurrences, then, will determine how and in which direction cohesion is achieved (anaphoric or cataphoric). From the perspective of word formation, two main neological patterns are taken into account: blends' combination with affixes (in particular, suffixes); reinterpretation of blends' splinters as formatives – often involving splinters' subsequent desemantization (Kemmer, 2003). The identification of suffixes as markers of register and text style (Plag et al., 1999: 225; Seiffert, 2015: 2181) will support the recognition of those special languages where blends often occur (Mattiello, 2019), and to determine which affixes are most common in our sample.

References:

- Bauer, L. 2012. Blends: Core and periphery. In V. Renner, F. Maniez, P. Arnaud (eds.), *Cross-Disciplinary Perspectives on Lexical Blending*. Berlin / Boston. De Gruyter Mouton. 11-22.
- Bertinetto, P. M. 2001. Blends and syllabic structure: A four-fold comparison. In L. Mercé, N. Alturo, E. Boix, M. Lloret, L. Payrató (eds.), *La gramàtica i la semàntica en l'estudi de la variació*. Barcelona. Promociones y Publicaciones Universitarias S.A. 59-112.
- Bombi, R. 2015. Il contatto anglo-italiano e i riflessi nel lessico e nei processi di "formazione delle parole". In C. Consani, *Il contatto interlinguistico fra presente e passato*. Milano. LED. 379-396.
- Cacchiani, S. 2011. On unfamiliar Italian lexical blends from names and nouns. *Linguistica* 51. 105-120.
- Cacchiani, S. 2016. On Italian Lexical Blends: Borrowings, Hybridity, Adaptations, and Native Word Formations. In S. Knospe, A. Onysko, M. Goth (eds.), *Crossing languages to play with words: Multidisciplinary perspective*. Berlin/Boston. De Gruyter. 305-336.
- Castagneto, M., Parente, E. 2020. Ti va una apericena al ristorante? I blend come lessicalizzazioni complesse. In I. Valenti (ed.), *Lessicalizzazioni complesse. Ricerche e teoresi*. Roma. Aracne editrice. 345-372.
- Dederding, H.-M. 1983. Wortbildung und Text: Zur Textfunktion (TF) von Nominalkomposita (NK). *Zeitschrift für germanistische Linguistik* 11. 49-64.
- Dressler, W. U. 1987. *Leitmotifs in natural morphology*. Amsterdam / Philadelphia. John Benjamins Publishing Company.
- Fandrych, I. M. 2004. *Non-morphematic Word-Formation Processes: A Multi-Level Approach to Acronyms, Blends, Clippings and Onomatopoeia*. Ph.D. Thesis. Bloemfontein. University of the Free State.
- Gries, S. T. 2004. Shouldn't it be breakfunch? A quantitative analysis of blend structure in English. *Linguistics* 42(3). 639-667.
- Gumperz, J. J. 1982. *Discourse strategies*. Cambridge. Cambridge University Press.
- Halliday, M. A. K., Hasan, R. 1976. *Cohesion in English*. London. Longman.
- Kemmer, S. 2003. Schemas and lexical blends. In H. Cuyckens, T. Berg, R. Dirven, K.-U. Panther (eds.), *Motivation in language: From case grammar to cognitive linguistics. A Festschrift for Gunter Radden*. Amsterdam / Philadelphia. Benjamins. 69-97.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., Suchomel, V. 2014. The Sketch Engine: ten years on. In *Lexicography* 1. 7-36.
- Kjellander, D. 2015. *Beauty and the Blend: Implications of Cognitive Constraints and Word Class Distribution in Lexical Blending*. MA dissertation. Stockholms universitet, Stockholm. Available at: <http://www.divaportal.org/smash/get/diva2:821138/FULLTEXT01.pdf>
- Lehrer, A. 1996. Identifying and Interpreting Blends: An Experimental Approach. *Cognitive Linguistics* 7(4). 359-390.
- Lehrer, A. 2007. Blendalicious. In J. Munat (ed.), *Lexical Creativity, Texts and Contexts*. Amsterdam / Philadelphia. John Benjamins Publishing Company. 115-133.

- Levinson, S. C. 2003. Contextualizing 'contextualization cues'. In S. L. Eerdmans, C. L. Prevignano, P. J. Thibaut, *Language and interaction: discussions with John J. Gumperz*. Amsterdam / Philadelphia. John Benjamins Publishing Company. 31-39.
- Mattiello, E. 2013. "Extra-Grammatical morphology in English". *Abbreviations, blends, reduplicatives and related phenomena*. Berlin / Boston. De Gruyter.
- Mattiello, E. 2019. A corpus-based analysis of new English blends. In I. Balteiro, L. Bauer (eds.), *Lexis 14*. DOI: <https://doi.org/10.4000/lexis.3660>
- McArthur, T. (ed.). 1992. *The Oxford Companion to the English Language*. Oxford. Oxford University Press.
- Piñeros, C. E. 2004. The creation of portmanteaus in the extragrammatical morphology of Spanish. *Probus 16(2)*. 203-240.
- Plag, I., Dalton-Puffer, C., Baayen, R. H. 1999. Morphological productivity across speech and writing. *English Language and Linguistics 3(2)*. 209-228.
- Renner, V., Lalić-Krstin, G. 2011. *Predicting stress assignment in lexical blends: the case of English and Serbian*. ELLSIIR International Conference. 1. 2009. Serbia. Belgrade. HAL. 265-273.
- Renner, V. 2015. Lexical blends as wordplay. In A. Zirker, E. Winter-Froemel (eds.), *Wordplay and Metalinguistic / Metadiscursive Reflection. Authors, Contexts, Techniques and Meta-Reflection*. 1. Berlin / Boston. De Gruyter. 119-134.
- Ronneberger-Sibold, E. 2006. Lexical Blends: Functionally Tuning the Transparency of Complex Words. *Folia Linguistica 40(1/2)*. 155-181.
- Scalise, S. 1983. *Morfologia Lessicale*. Padova. Clesp.
- Seiffert, A. 2015. Word-formation and text. In P. O. Müller, I. Ohneiser, S. Olsen, F. Rainer, *Word-Formation*. 3. Berlin / Boston. De Gruyter. 2178-2191.
- Thornton, A. M. 1993. Italian blends. In L. Tonelli, W.U. Dressler (eds.), *Natural Morphology: Perspectives for the Nineties*. Padova. Unipress. 143-155.
- Thornton, A. M. 2004. Parole macedonia. In M. Grossmann, F. Rainer (eds.), *La formazione delle parole in italiano*. Tübingen. Niemeyer. 567-571.

Personal name compounding in German On the relationship between productivity and pragmatic functions across discourse domains and text types

M. Belosevic, S. Arndt-Lappe

Trier University, Trier, Germany

This paper examines German personal name compounds (PN compounds) by comparing their productivity with regard to external and structural factors (cf. Plag 2006). PN compounds are determinative compounds with a personal name as a second element and a lexical unit as a modifier

(*Helikopter-Cem*). In contrast to compounds with a proper name as a first element (*Merkel-Besuch*, cf. Schlücker 2017, 2020), they have received little attention in the literature. So far, they have been cursorily mentioned as an unproductive word-formation pattern that bears an evaluative text function (cf. Wildgen 1981, Kürschner 2020).

By drawing on the newspapers from the German Reference Corpus (DeReKo)¹, the blog corpus of the German Digital Dictionary (DWDS)² and Twitter³, we test the hypothesis that the productivity of PN compounds depends on textual, discursive, and structural factors. Although productivity measures proposed by Baayen (2009) depend on the corpus size and pose a problem, especially for Twitter data, we consider the type frequency (cf. Baayen 2009: 901) a first step towards measuring the productivity of PN compounds.

A corpus of 450 types, retrieved from the list of 60 names from three discourse domains (politics, sports and celebrities), has been annotated in *Maxqda* for external (discourse domain and text types) and structural properties (the semantics of a modifier and the co-occurrence of other lexemes in the context) that may influence their productivity. The text types include newspapers, blogs and tweets (tweets are similar to journal articles or volume contributions, cf. Hausendorf et al. 2017: 174). Furthermore, the data yields eventive, partitive and qualifying readings that emerge from different aspects of extralinguistic knowledge about the name bearer. Whereas eventive readings refer to a social event where the name bearer participates (*Helikopter-Cem*), partitive readings evoke the knowledge about the membership of a name bearer to some entity (*Polenklose*). Qualifying readings comprise a property of a name bearer (*Brillen-Dobrindt*).

The analysis indicates that eventive readings comprising the names of politicians are generally more frequent than attestations with the names of celebrities and sportspeople. Furthermore, eventive and qualitative readings are more frequent in tweets and blogs, whereas the number of types in partitive readings is higher in newspapers. These findings can be explained by taking a closer look at the text function of PN compounds. Partitive readings usually bear a reference function, whereas eventive and qualitative readings do not primarily refer to the name bearer but contribute to evaluating his actions or decisions. It is, therefore, no surprise that evaluative readings are more frequent in social media than in newspapers. Regarding the linguistic context, PN compounds are usually not used meta linguistically, and they often co-occur in the same context. The analysis sheds new light on the productivity of PN compounds and their functions in different text types and domains.

References:

- Baayen, Harald R. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics. An international handbook*. Vol. 2. Berlin, New York: De Gruyter. 899–919. DOI: <https://doi.org/10.1515/9783110213881.2.899>.
- Hausendorf, Heiko, Wolfgang Kesselheim, Hiloko Kato & Martina Breitholz. 2017. *Textkommunikation: Ein textlinguistischer Neuanatz zur Theorie und Empirie der Kommunikation mit und durch Schrift*. Berlin: de Gruyter.

¹<https://cosmas2.ids-mannheim.de/cosmas2-web/faces/home.xhtml> We searched for all newspaper corpora within Archive W-public - all public corpora of Archive W.

²<https://www.dwds.de/d/korpora/blogs>

³<https://twitter.com/>

- Kürschner, Sebastian. 2020. Nickname formation in West Germanic: German Jessi and Thomson meet Dutch Jess and Tommie and English J-Bo and Tommo. In Gunther de Vogelaer, Dietha Koster, & Torsten Leuschner (eds.), *German and Dutch in Contrast. Synchronic, Diachronic and Psycholinguistic Perspectives*. Berlin: de Gruyter. 15–47. DOI: <https://doi.org/10.1515/9783110668476-002>.
- Plag, Ingo. 2006. Productivity. In Bas Aarts & April M. S. McMahon (eds.), *The handbook of English linguistics*. Malden, Mass.: Blackwell. 537–557.
- Schlücker, Barbara. 2017. Eigennamenkomposita im Deutschen. In Johannes Helmbrecht, Damaris Nübling, & Barbara Schlücker (ed.), *Namengrammatik*. Hamburg: Buske. 59–93.
- Schlücker, Barbara. 2020. Von Donaustrom zu Donauwelle. Die Entwicklung der Eigennamenkomposition von 1600–1900. *Zeitschrift für germanistische Linguistik* 48(2). 238–268. DOI: <https://doi.org/10.1515/zgl-2020-2002>.
- Wildgen, Wolfgang. 1981. *Grundstrukturen und Variationsmöglichkeiten bei Eigennamenkomposita: Komposita mit den Eigennamen Schmidt und Strauß als Konstituenten in Wahlkampfberichten des Spiegels*. Trier: Linguistic Agency University of Trier.

Thursday, 5 of May

A Proposal for Explicit Word Formation Annotation in Discourse Corpora

M.-C. Müller

Heidelberg Institute for Theoretical Studies (HITS), Heidelberg, Germany

Meaningful empirical linguistic analysis requires both machine-readable, annotated corpora of sufficient size and computational methods for processing them. For modern annotated corpora, especially those that cover higher-level phenomena like coreference, discourse connectives, rhetorical structure, etc., *stand-off*, *multilevel annotation* (Gries & Berez, 2017) is the de-facto standard representation format. It supports the coexistence of different annotations on the same textual data, which makes it a prerequisite for the analysis of the interplay of phenomena on different linguistic levels.

We propose to enhance annotated text corpora by adding dedicated WORD FORMATION annotation levels. We argue that this will greatly improve the way that complex words can be analysed in context, including, but not limited to, the role they play in establishing and maintaining discourse structure and coherence.

Many corpora share a limitation, viz. the *lack of recognizing linguistic structure below the level of the orthographic word*, which – at least for German – includes the entire domain of word formation. The reason is that text is usually tokenized, i.e. split into the smallest units that can be annotated, on the basis of whitespace and punctuation. As a result, complex nominals like (synthetic) compounds, derivations, etc. are treated as unanalysed monoliths, just because they happen to be spelled in one word, even though their discourse function could be described much more accurately on the basis of their constituents.

Consider the following two examples, which are the result of a cursory inspection of the Potsdam Commentary Corpus (PCC) 2.2 (Bourgonje & Stede, 2020) (bracketing and emphasis M.-C.M.):

- (1) “Eine der modernsten Produktionslinien ist gestern offiziell gestartet worden, mit deren Hilfe frische **Milch** [...] auf dem Frühstückstisch [...] landet. Zu Recht ist die Chefetage des [**Milch**]giganten zufrieden. [...] Und das macht Mut. Mut, mit dem die [**Milch**]verarbeiter [...] weitere Investitionen anvisieren.” (MAZ-11544)

Of the three occurrences of **Milch**, all of which contribute to lexical cohesion by repetition, two occur in *one-word* compounds, which prevents them from being independent tokens. Also, the analysis of the anaphoric expression **Milchverarbeiter** as a synthetic compound is inaccessible.

- (2) “Diepensee siedelt um. [...] Dessen konnten sich die vor dem möglichen **Ausbau** des Flughafens Schönefeld weichenden Dorfbewohner gewiss sein. Bis eine Presseinformation der [**Ausbau**]-Planer einschlug [...].” (MAZ-6993)

Ausbau-Planer is the first mention of a discourse-new referent. While one could argue that **Planer** alone would require a *bridging* interpretation, the synthetic combination with **Ausbau** renders the whole expression unambiguous.

If there is consensus that, *in principle*, analyses like the above are useful, realisation is straightforward. In practical terms, what is required is a flexible multi-level annotation framework like our tool MMAX2 (Müller & Strube, 2006; Müller, 2020), which supports both manual annotation and automated analysis.

Our proposal is linguistically unbiased, or even naive: In essence — at least when implemented in an adequate corpus data representation framework — it may be merely a technicality, but one that we argue is able to support new, informative analyses.

References:

- Bourgonje, Peter & Manfred Stede. 2020. The Potsdam Commentary Corpus 2.2: Extending Annotations for Shallow Discourse Parsing. *Proceedings of the 12th Language Resources and Evaluation Conference, LREC 2020, Marseille, France, May 11-16, 2020*, ed. by Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asunción Moreno, Jan Odijk, and Stelios Piperidis, 1061–1066. European Language Resources Association. Online: <https://aclanthology.org/2020.lrec-1.133/>.
- Gries, Stefan Th. & Andrea L. Berez. 2017. Linguistic annotation in/for corpus linguistics. In Nancy Ide & James Pustejovsky (eds.), *Handbook of linguistic annotation*. Dordrecht, The Netherlands: Springer Netherlands. 379–409.
- Müller, Mark-Christoph. 2020. pyMMAX2: Deep access to MMAX2 projects from python. *Proceedings of the 14th linguistic annotation workshop*. Barcelona, Spain: Association for Computational Linguistics. 167–173. Online: <https://aclanthology.org/2020.law-1.16>.
- Müller, Mark-Christoph & Michael Strube. 2006. Multi-level annotation of linguistic data with MMAX2. *Corpus technology and language pedagogy: New resources, new tools, new methods*, ed. by Sabine Braun, Kurt Kohn, and Joybrato Mukherjee. Frankfurt a.M., Germany: Peter Lang. 197–214.

Improving Coreference Resolution with Word Formation

W. Zhao, K. Mathews, H. Chai

Heidelberg Institute for Theoretical Studies (HITS), Heidelberg, Germany

In natural language processing (NLP), coreference resolution, aiming for the identification of a collection of expressions referring to the same real-world entity, allows for a better understanding in natural language, and as such has received attention for long. Given the unprecedented success of contextualized language models such as BERT (Devlin et al., 2019), employing them in NLP tasks requiring deep semantic understanding has achieved extraordinary performances, as it is the case in coreference resolution (Joshi et al., 2019). However, previous work does not outline the current state of limitations in coreference resolution systems, such as where and why they cannot resolve coreference.

In discourse linguistics, word formation has a long-lasting connection with coreference resolution. Research has shown that employing syntactic function of word formation, such as pronominalizing compounds – a coreference-like approach, can decrease “readers” cognitive load, and accordingly, increase text coherence (Schröder, 1978; Dederding, 1983; Lipka, 1987; Eichinger, 1995; Schlienz, 2004; Peschel, 2011). But previous studies are limited in scope to nominal compounds as the only pattern of word formation, and more prominently, they do not contribute to the current state of coreference resolution systems.

In this work, we aim for employing word formation to understand the deficiencies of coreference resolution systems and address them. To this end, we intend to assemble noun phrases from the largest coreference dataset CoNLL (Pradhan et al., 2012) and its extension (Chai et al., 2020), and relate the noun phrases to word formation patterns, such as compounds, word abbreviations, noisy words, metaphor and metonymy. Next, we will resolve coreference with the current state-of-the-art coreference resolution system (Joshi et al., 2020), and then analyze the impact of word formation patterns on coreference resolution. Finally, we will propose a simple approach to improve the coreference resolution system with rewriting the noun phrases for which the corresponding coreference is challenging to resolve. In particular, we intend to mask these noun phrases and employ BERT to perform a fill-in-the-blank task, i.e., predicting masked words that fit best in the context. In the following, we elaborate on our approach.

Our Approach. First, we mask each target word, and then employ BERT to retrieve a shortlist of candidates pertaining to the target word, *v.i.z.*, retaining a set of words with the probabilities above a threshold. We illustrate how to choose per target word a substitution from shortlisted candidates:

- **Novel word sense:** we use cosine similarity to measure distances between synset (word sense) definitions of a target word and of each word in shortlisted candidates, based on their embeddings. We replace a target word with its nearest neighbor in the shortlist according to synset similarities. We acknowledge that this approach requiring a word sense disambiguation tool to identify word synset, which is prone to wrong identification.
- **Novel words** are not present in the BERT vocabulary and WordNet, thereby not associated with word embeddings and synset definitions. This means we cannot find per novel word a substitution, based on their semantic and synset similarities. To this end, we use the word with the highest probability in the vocabulary to replace a target word. Such substitutions, although contextually related, may be not semantically related to (or synonymous with) target words.

References:

- Chai, Haixia, Wei Zhao, Steffen Eger & Michael Strube. 2020. Evaluation of coreference resolution systems under adversarial attacks. In *Proceedings of the First Workshop on Computational Approaches to Discourse*. Online. Association for Computational Linguistics. 154–159.
- Dederding, Hans-Martin . 1983. Wortbildung und Text. *Zeitschrift für Germanistische Linguistik* 11(1). 49-64. Zur Textfunktion (TF) von Nominalkomposita (NK).
- Devlin, Jacob , Ming-Wei Chang, Kenton Lee, & Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1 (Long and Short Papers)*. Minneapolis, Minnesota. Association for Computational Linguistics. 4171–4186.
- Eichinger, Ludwig M. 1995. *Wegweiser durch Textwelten. Wozu komplexe Substantive gut sind.*
- Joshi, Mandar, Danqi Chen, Yinhan Liu, Daniel S. Weld, Luke Zettlemoyer & Omer Levy. 2020. SpanBERT: Improving pre-training by representing and predicting spans. *Transactions of the Association for Computational Linguistics* 8. 64–77.
- Joshi, Mandar, Omer Levy, Daniel S. Weld & Luke Zettlemoyer. 2019. BERT for coreference resolution: Baselines and analysis. In *Empirical Methods in Natural Language Processing (EMNLP)*.
- Lipka, Leonhard. 1987. Word-formation and text in English and German.
- Peschel, Corinna. 2011. *Zum Zusammenhang von Wortneubildung und Textkonstitution*. Vol. 237. Walter de Gruyter.
- Pradhan, Sameer, Alessandro Moschitti, Nianwen Xue, Olga Uryupina & Yuchen Zhang. 2012. CoNLL- 2012 shared task: Modeling multilingual unrestricted coreference in OntoNotes. In *Joint Conference on EMNLP and CoNLL - Shared Task*. Jeju Island, Korea. Association for Computational Linguistics. 1–40.
- Schlienz, Michael. 2004. *Wortbildung und Text: eine Untersuchung textverknüpfender Wortbildungselemente*. Palm und Enke.
- Schröder, Marianne. 1978. Über textverflechtende Wortbildungselemente. *Deutsch als Fremdsprache. Zeitschrift für Theorie und Praxis des Deutschunterrichts für Ausländer Leipzig* 15(2). 85–92.

Affixes and their role in discourse structure

A. Baltuttis

Leipzig University, Leipzig, Germany

This talk considers the possibility that affixes as word formation elements without a semantic information can still play a vital role for the constitution of discourse.

Until now relatively few studies of word formation elements have been conducted. The majority of research in that field took place in the late 1970s and 1980s (E.g. Schröder 1978, Viehweger 1978, Dederding 1983). Since then a few more works on this topic have been published (E.g. Wolf 1996, Matussek 1994, Peschel 2002, Schlienz 2004). But most of them exclusively focus on compound forms and their function within text constitution, while ignoring derivational or conversational patterns.

Although a large corpus study is yet to be done, smaller samples of texts I collected already suggest that affixes might play a larger role in forming text constitution and coherence in discourses. Plag's et al. work on productivity of derivational affixes (1999) shows that the use of affixes depends on register, text type and subject of the text. This use is explained – among other things – with the affixes's embedding in the context.

Furthermore, it is assumed that topic relations are achieved by linking anaphoric elements that are semantically recurrent. Those *topic elements* (Koch 1966, Agricola 1972) are primarily cohesion markers, but function as coherent elements when they are used to link a sequence of sentences. The problem is that only recurrent elements that form a semantic connection are considered to be able to fulfill a coherent function. Other cohesive markers such as affixes are excluded because they carry no semantic information (Wolf 1996).

However, Lim (2004: 118) briefly shows – based on the following example – that word formation elements such as the affix *zer-* are capable of forming topic relations, since two identical affixes carry the same semantic information, like two basic morphemes do. This suggests that affixes can also function as text linking elements and possibly even as coherence markers.

Am liebsten **zerstört** die Masse Häuser und Gegenstände. Da es sich oft um **Zerbrechliches** handelt, wie Scheiben, Spiegel, Töpfe, Bilder, Geschirr, neigt man dazu zu glauben, daß es eben diese **Zerbrechlichkeit** von Gegenständen sei, die die Masse zur **Zerstörung** anreizt. Es ist nun gewiß richtig, daß der Lärm der **Zerstörung**, das **Zerbrechen** von Geschirr, das Klirren von Scheiben zur Freude daran ein Beträchtliches beiträgt.

Therefore, the main question that needs to be discussed is the following: Do affixes only function as cohesive markers – as believed by most scientists (e.g. Wolf 1996, Averintseva- Klisch 2018) – or are they capable of fulfilling a larger function, specifically by establishing coherence in a broad sense?

References:

- Agricola, Erhard. 1972. *Semantische Relationen im Text und im System*. 2., bearbeitete Aufl. Mouton/Paris.
- Averintseva-Klisch, Maria. 2018. *Textkohärenz*. Heidelberg.
- Dederding, Hans-Martin. 1983. Wortbildung und Text. Zur Textfunktion (TF) von Nominalkomposita (NK). *Zeitschrift für Germanistische Linguistik* 11(1). 49-64.
- Koch, Sabine. 1972. Semantische Relationen in sprachlichen Texten (Topikrelationen). In Nüdel, S. (ed.), *Automatische Sprachübersetzung II*. Berlin. 129-158.
- Lim, Seong Woo. 2004. *Kohäsion und Kohärenz: Eine Untersuchung zur Textsyntax am Beispiel schriftlicher und mündlicher Texte*. Dissertationsschrift. Würzburg.
- Matussek, Magdalena. 1994. *Wortneubildungen im Text*. Hamburg.
- Peschel, Corinna. 2002. Zum Zusammenhang von Wortneubildung und Textkonstitution. Tübingen: Niemeyer.
- Plag, Ingo, Christiane Dalton-Puffer & R. Harald Baayen. 1999. Morphological productivity across speech and writing. *English Language and Linguistics* 3(2). 209–228.
- Schlienz, Michael. 2004. *Wortbildung und Text. Eine Untersuchung textverknüpfender Wortbildungselemente*, Erlangen & Jena.
- Schröder, Marianne. 1978. Über textverflechtende Wortbildungselemente. In *DaF* 15. 85-92.

Viehweger, Dieter. 1978. Struktur und Funktion nominativer Ketten im Text. In Motsch, W. (ed.), *Kontexte der Grammatiktheorie (studia grammatica XVII)*. Berlin. 149-168.

Wolf, Nibert R. 1996. Wortbildung und Text. *Sprachwissenschaft* 21. 241-261.

Word formation and cohesion in specialized languages

K. Menzel

Saarland University, Saarbrücken, Germany

This talk addresses the relationship between selected word formation patterns and cohesion in English as well as implications of this relationship for translation (English<->German). This topic was recognized as essential by Lipka (1987), but has not received much coverage in the literature on morphology, text linguistics and translation studies.

Lexical cohesion and language- and text-type specific word formation strategies interact. The textual relations between lexical items can be made explicit via cohesive chains of interrelated items sharing similar morphological structures. These items function as cross-clausal links signalling semantic relationships or evoking associations with formally similar words in the same text. It is beneficial for language users to extend their knowledge of word formation processes in their first (and second) language to increase their discourse competence (Menzel 2018).

The morphological level is of particular importance in languages for specific purposes (Halliday 1964:127f). I will therefore discuss word formation patterns in specialized English and English/German corpora from a diachronic, synchronic and cross-linguistic perspective.

These patterns include:

- Formations with Graeco-Latinate combining forms (e.g. *photolysis*) in scientific English, patterns that in many cases serve as internationalisms (Menzel/Degaetano-Ortlieb 2017).
- Suffixed nouns denoting actions, qualities or states (e.g. German nouns with *-heit*, *-keit*). Lexical items in specialized registers contain a high degree of word-internal complexity, and English often makes use of root morphemes with Romance/Latin/Greek etymology (e.g. *environmentality*) and suffixes from different etymological sources (*-acy*, *-hood*, *-ity*, *-iety*, *-ment*, *-ness*). There is a certain tendency in specialized English to avoid suffixed complex nouns with Germanic components or with several lexical morphemes. German specialized texts show a higher number of affixes and/or root morphemes combined within such nouns. Difficulties may therefore arise when translating specialized texts (Menzel 2018).
- Eponyms derived from personal names, a productive resource for components of new single-word and multi-word expressions in English for specific purposes (Menzel 2021). I will discuss adjective+noun and proper noun+noun eponyms and structures containing a possessive marker (*Brownian movement*, *Basset force*, *Hadley's quadrant*). From a diachronic perspective, English multiword eponyms with possessive markers have become less productive as they are potentially ambiguous and may not easily be distinguishable from true possessives indicating ownership.
- Multimorphemic adjectives (adjective compounds such as *arsenic-bearing* and adjectives negated with

prefixes, e.g. *non-linear*, cf. Menzel 2020, Menzel et al. Forthcoming). Such adjectives are often used in languages for specific purposes as temporary formations for the densification and adjectivization of phrasal or clausal structures. English makes use more regularly of adjective compounds than German. We also observed an increased usage of English adjectives with affixal negation markers in scientific writing occurring in increasingly diverse discourse structures. The variation between adjective negation via affixes vs. analytic negation marker is interesting from a contrastive perspective as in German the status of *nicht(-)* (and also in French *non(-)* in front of adjectives as affix or free morpheme is less clear than that of the English negation prefix *non-*. The wordhood status of adjective compounds also needs to be addressed.

The patterns and findings discussed in this talk illustrate why expert language users and translators should consider the productivity and limits of word or lexeme formation patterns in the languages they are working with and the contribution of word formation choices to discourse coherence.

References:

- Halliday, M.A.K. 1964. Comparison and Translation. In M.A.K. Halliday, Angus McIntosh & Peter Strevens. (eds.), *The Linguistic Sciences and Language Teaching*. London: Longman. 111-134.
- Lipka, L. 1987. Word-formation and text in English and German. In B. Asbach-Schnitker & Johannes Roggenhofer (eds.), *Neuere Forschungen zur Wortbildung und Historiographie der Linguistik: Festgabe für Herbert E. Brekle zum 50. Geburtstag*. Tübingen: Narr. 59–67.
- Menzel, K. 2018. Morphological awareness and translation studies. In T. Tinnefeld (ed.) with the collaboration of R. Kresta, J. Szakos, *Challenges of modern foreign language teaching – Reflections and analyses*. Saarbrücken: htw saar 2018. 181-206.
- Menzel, K. 2020. Morphologische Strukturen und Wortbildungsaspekte aus übersetzungswissenschaftlicher Perspektive. In B. Ahrens, M. Beaton-Thome, M. Krein- Kühle, R. Krüger, L. Link, U. Wienen (eds.), *Interdependenzen und Innovationen in Translation und Fachkommunikation – Interdependence and Innovation in Translation, Interpreting and Specialised Communication*. 45-70.
- Menzel, K. 2021. Scientific eponyms throughout the history of English scholarly journal articles. In H. Van de Velde and F. T. Dolezal (eds.), *Broadening Perspectives in the History of Dictionaries and Word Studies*. Newcastle upon Tyne: Cambridge Scholars Publishing. 159-193.
- Menzel, K. & S. Degaetano-Ortlieb. 2017. The diachronic development of combining forms in scientific writing. In *Lege artis. Language yesterday, today, tomorrow, The Journal of University of SS Cyril and Methodius in Trnava*. Warsaw: De Gruyter Open. Vol. 2(2) 2017. 185-249.
- Menzel, K., M. Krielke & S. Degaetano-Ortlieb. Forthcoming. Negated adjective constructions in English scientific journal articles. *Lege artis. Language yesterday, today, tomorrow, The Journal of University of SS Cyril and Methodius in Trnava*. Warsaw: De Gruyter Open. Vol. 1 2022.

Complex nouns as markers of academic register in L1-and L2-authored essays

J. Lukassek, A. Lüdeling, A. Shadrova, S. Wan

Humboldt-Universität zu Berlin (HU), Berlin, Germany

How do advanced learners of German acquire nominal style? This question is interesting for several reasons. First, nominal style (Biber & Gray 2010; Hennig 2020) plays a central role for discourse structuring in argumentative registers written in academic contexts in so far as it allows for a high concentration of information in a small number of words. Furthermore, terminology in academic contexts is to a large extent encoded in nouns. This is due to the fact that nouns are considered to convey precision and clarity, cf. Roelcke (2020). The key to mastering argumentative registers in German is therefore mastering the condensation of information into complex nouns that can then be heads of complex nominal phrases.

Second, nominal style has been shown to pose a prominent challenge for learners of German as a foreign language (GFL) (see Berman & Nir-Sagiv, 2007 for parallel effects in English). For advanced learners, acquiring a native-like command of a given register comes with the challenge to recognize the **adequate usage** and **distribution** of morphological patterns and possibilities for productive noun formation in that register. In our talk, we want to investigate RQ1 that results from the described situation:

RQ: Do advanced learners of German as a foreign language use nominal style in the same way as L1 speakers in argumentative essays?

A careful comparison of morphological patterns between L1 and L2 speakers in a tightly controlled setting helps us understand one facet of register knowledge in learners.

We discuss our research question using data from the Falko Essay corpus (Reznicek et al. 2012), a deeply annotated learner corpus of academic essays including a control group of L1- authored texts on the same topics. We manually annotated all nouns in Falko for the highest word formation process involved (derivation, compounding, conversion, non-productive nominalizations, transpositions, etc.) and a number of other morphological and lexical properties (origin, complexity), cf. the annotation guidelines (Lukassek et al. 2021). Based on these annotations, we present and discuss the following findings: (1) Overall, complex nouns are used in a similar way in the L2 and the L1 subcorpora. This implies that the learners cope well with identifying the adequate register features in the nominal domain. (2) A closer look at the distributions over the annotated noun types shows that the variance between texts is very high in both subcorpora. This variance can partially be explained by topic effects, but even within the same topic distributions over the noun types can vary massively (for the L1 speakers, see Shadrova et al., accepted). This finding entails new questions: What exactly is the target register that advanced learners are aiming at? How can we even compare L1 and L2 speakers in this domain if deviations from the mean are high? Is the distribution over complex noun classes a question of subjective preferences rather than register?

References:

- Berman, Ruth A. & Bracha Nir-Sagiv. 2007. Comparing narrative and expository text construction across adolescence: A developmental paradox. *Discourse Processes* 43(2). 79–120. <https://doi.org/10.1080/01638530709336894>. <https://www.tandfonline.com/doi/abs/10.1080/01638530709336894>.
- Biber, Douglas & Bethany Gray. 2010. Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes* 9(1). 2–20.

- Hennig, Mathilde. 2020. *Nominalstil: Möglichkeiten, Grenzen, Perspektiven*. Narr Francke Attempto Verlag.
- Lukasek, Julia, Rudabe Akbari & Anke Lüdeling. 2021. *Guidelines for the morphological annotation of nouns in Falko / Richtlinie zur morphologischen Annotation von Nomina in Falko*. Vol. 1 (REALIS). Berlin: REALIS 1.
- Reznicek, Marc, Anke Lüdeling, Cedric Krummes, Franziska Schwantuschke, Maik Walter, Karin Schmidt, Hagen Hirschmann & Torsten Andreas. 2012. *Das falko-handbuch. Korpusaufbau und Annotationen version 2.01*. Tech. rep. Humboldt-Universität. https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/falko/FalkoHandbuchV2/at_download/file.
- Roelcke, Thorsten. 2020. *Fachsprachen*. 4th edn.
- Shadrova, Anna, Pia Linscheid, Julia Lukasek, Anke Lüdeling & Sarah Schneider. Accepted. A challenge for contrastive L1/L2 corpus studies: Large inter- and intra-individual variation across morphological, but not global syntactic categories in task-based corpus data of a homogeneous L1 German group. *Frontiers in Psychology* 12. <https://doi.org/10.3389/fpsyg.2021.716485>.

Partners & Sponsors

Organizers

Adele Baltuttis, Anna Bliß, Maximilian Frankowsky & Barbara Schlücker

Contact

🏠 <https://home.uni-leipzig.de/grammatik-des-deutschen/WFDS/>

✉ wfds@uni-leipzig.de

Sponsors

We thank *Universitätsgesellschaft - Freunde und Förderer der Universität Leipzig* for financial support of the workshop.



Universitätsgesellschaft
Freunde und Förderer
der Universität Leipzig

Space for your notes



This template originates from [LaTeXTemplates.com](https://www.latextemplates.com) and is based on the original version at:
https://github.com/maximelucas/AMCOS_booklet

