

---

**Sektion Computerlinguistik | *Section Computational Linguistics***

Raum | *Room*: Tagungsraum, Seminargebäude

---

**Poster Session**

**Rainer Osswald**

*Heinrich-Heine-Universität Düsseldorf*

rainer.osswald@hhu.de

The poster session is organized by the Special Interest Group on Computational Linguistics of the DGfS ([dgfs.de/cl](https://dgfs.de/cl)).



---

## Creating Nonsensical Dependency Treebanks for Multiple Languages

---

**David Arps**

*Heinrich-Heine-Universität Düsseldorf*

david.arps@hhu.de

I present a resource of nonsensical syntactic dependency treebanks. The resource is generated by an algorithm which takes as input a standard dependency treebank in the Universal Dependencies format (de Marneffe et al. 2021). The result is a treebank that contains the same syntactic structures as the input, but nonsensical sentences. The algorithm has a random component, thereby creating an arbitrary number of parallel nonsensical treebanks. The general idea is that content words in the input sentences are replaced by other forms that appear in the same syntactic context elsewhere in the input treebank. In the following example, all words printed in boldface are replaced using this mechanism:

- (1) a. We have an entire new set.
- b. We **kidnap** a **pregnant international fire**.

The syntactic context consists of POS tag, morphosyntactic features and dependency relations. A language-specific component deals with phenomena that are not covered in the syntactic context but compromise the syntactic acceptability of the generated sentences in other ways. For instance, *an* is replaced with *a* because the adjective following the determiner has changed. I apply this algorithm to treebanks in several languages from different language families.

The resource can be put to use in various scenarios. In particular, I present one use case that quantifies the implicit syntactic information in neural language models (LMs). While the exact nature of these linguistic structures is not easily identifiable, several papers have tested how well syntactic dependency structure can be reconstructed from an LM’s internal representations (see Müller-Eberstein et al., 2022 for an overview). For multiple languages, I test if experimental results from the literature still hold when using the data created in this project. The use of nonsensical data diminishes the role of semantic cues which potentially distort previous experiments.

**References:** • de Marneffe, Marie-Catherine & C. Manning & J. Nivre & D. Zeman (2021). Universal Dependencies. *Computational Linguistics* 47(2), 255-308. • Font, S. (2008). *Font Survey*. Cologne: Quick Press. • Müller-Eberstein, Max & R. van der Goot & B. Plank (2022). Probing for Labeled Dependency Trees. *Proceedings of ACL 2022*. Dublin, Ireland: Association for Computational Linguistics.

---

## Spoken Language Corpora at the Leibniz Institute for the German Language: Established Tools and New Developments

---

**Mark-Christoph Müller, Elena Frick & Henrike Helmer**

*Leibniz-Institut für Deutsche Sprache, Mannheim*

{mark-christoph.mueller, frick, helmer}@ids-mannheim.de

While corpora of **written text** in machine-readable form are readily available for researchers in (computational) linguistics, the situation for transcribed **spoken language** is radically different, because the peculiarities of spontaneous spoken language make the creation of high-quality, standardized, and versatile corpora a complex task. As the leading research data centre for German spoken language corpora, the **Archive for Spoken German** (*Archiv für Gesprochenes Deutsch* (AGD), [agd.ids-mannheim.de](http://agd.ids-mannheim.de)), located at the IDS Mannheim, is dedicated to making corpora of spoken German available to the international research community in a sustainable way.

The **Database for Spoken German** (*Datenbank für Gesprochenes Deutsch* (DGD), Schmidt 2017, [dgd.ids-mannheim.de](http://dgd.ids-mannheim.de)), launched in 2012, is an established search and browsing platform providing access to (currently) 40 corpora from the AGD (approx. 4700 hours of audio and video recordings and manually created time-aligned transcripts with 20 mio. tokens). The DGD uses a rich data model which is specifically tailored to the requirements of spoken language. Each *transcribed* token has annotations for the corresponding *normalized* and *lemmatized* forms, which captures and allows to query e.g. dialectal pronunciation variants of the same lemma. Each token also has an entry from a set of POS-Tags specifically enhanced for spoken language, with tags for e.g. hesitation and response particles, discourse markers, and tag questions. These primary data are enriched by corpus *meta* data like e.g. speaker demographics, interactional settings, and conversation topics.

In addition, the **ZuMult** project (Fandrych et al. 2022, [zumult.org](http://zumult.org)) has developed a suite of web applications to complement the DGD with new features for searching and browsing transcripts. By extending the DGD data model with CQP and ISO/TEI standards, they allow to query time-based span annotations as well as typical spoken language phenomena, e.g. speaker changes and overlaps or paraverbal events like laughter, coughing, and pauses.

**References:** • Schmidt, T. (2017): DGD – die Datenbank für Gesprochenes Deutsch. Mündliche Korpora am Institut für Deutsche Sprache (IDS) in Mannheim. In *Zeitschrift für germanistische Linguistik* 45.3., Berlin/Boston: de Gruyter, 451–463. • Fandrych, C., Frick, E., Kaiser, J., Meißner, C., Portmann, A., Schmidt, T., Schwendemann, M., Wallner, F. & Wörner, K. (2022): ZuMult: Neue Zugangswege zu Korpora gesprochener Sprache. In Kämper, H. & Plewnia, A. (Hgg.): *Sprache in Politik und Gesellschaft. Perspektiven und Zugänge*. Berlin/Boston: de Gruyter (Jahrbuch des Instituts für Deutsche Sprache 2021), 305–312.

---

## CORLiCo – Empirische Analysen im Nähe-Distanz-Kontinuum

---

**Sarah Broll & Roman Schneider**

*Leibniz-Institut für Deutsche Sprache, Mannheim*

broll@ids-mannheim.de, schneider@ids-mannheim.de

Gesprochene und geschriebene Sprachäußerungen unterscheiden sich hinsichtlich Repertoire und Verwendung von syntaktischen Konstruktionen oder Vokabular, stilistischen Merkmalen und einigem mehr. Typisch lautliche Ausdrucksmöglichkeiten und Strategien auf der einen Seite kontrastieren mit einer (zumindest angenommenen) größeren Nähe zu Sprachstandards und Konventionen. Eine Überwindung des medialen Binarismus erlaubt der von Koch und Oesterreicher (1985) eingeführte Ansatz eines mehrdimensionalen Kontinuums zwischen den beiden Polen “Sprache der Nähe” und “Sprache der Distanz”. Dieses Kontinuum erforschen wir empirisch anhand einer heterogen stratifizierten Datengrundlage.

Wir stellen das im Aufbau befindliche CORLiCo (*Corpus for the Oral-Literate Continuum*) vor, das sich mit einer Zielgröße von 100 Millionen Wort-Tokens zu ungefähr gleichen Teilen auf knapp 20 textsortenspezifische Subkorpora aufteilt. Abgedeckt werden nicht nur die Pole des Kontinuums, sondern auch solche Sprachäußerungen, die sich nicht eindeutig einem dieser Extreme zuordnen lassen: Wissenschaftskommunikation, Interviews, Reden, Liveticker, Songtexte, Social Media, E-Mails, Online-Diskussionen, Podcasts etc. Das Korpus repräsentiert damit eine sehr breit gefächerte Datengrundlage und füllt eine Lücke als aggregierende Ressource für die vergleichende Erforschung schriftlicher und mündlicher Diskurse. Sämtliche Texte sind mit Metadaten und Annotationen angereichert, die zum einen anhand der Textoberfläche, zum anderen auf der Basis des Outputs maschineller Tagger berechnet werden.

Wir implementieren und evaluieren ein automatisiertes Klassifikationsverfahren, das unter Nutzung von Random-Forest-Entscheidungsbäumen Aggregationen einzelner Vorhersagen durchführt (Broll & Schneider, o. J.). Für die Identifizierung der Pole definieren wir einen Merkmalskatalog aus Sprachphänomenen, die als Markierer für Nähe/Mündlichkeit bzw. Distanz/Schriftlichkeit diskutiert werden. Basierend auf der sehr guten Klassifikationsgüte verorten wir eine Reihe weiterer Textsorten. Die Ergebnisse werden zur besseren Interpretierbarkeit visuell aufbereitet.

**References:** • Broll, S., & Schneider, R. *Empirische Verortung konzeptioneller Nähe/Mündlichkeit inner- und außerhalb schriftsprachlicher Korpora*. (in Vorbereitung) • Koch, P., & Oesterreicher, W. (1985). Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch*, 36(1), 15–43. <https://doi.org/10.1515/9783110244922.15>.

---

## Ein Annotationsschema für Politeness und Preference Detection in mündlichen Konversationen

---

Clara Lachenmaier<sup>1</sup>, Sina Zariß<sup>1</sup> & Pia Bergmann<sup>2</sup>

<sup>1</sup>Universität Bielefeld., <sup>2</sup>Friedrich-Schiller-Universität Jena

clara.lachenmaier@uni-bielefeld.de, sina.zarriess@uni-bielefeld.de,

pia.bergmann@uni-jena.de

Soziale Praktiken wie Höflichkeit prägen Diskursgestaltung und Wortwahl. Ein Einbezug von Höflichkeit in moderne Dialogsysteme und -agenten können deren Interaktionen natürlicher wirken lassen (z.B. de Jong et al. 2008). Die computer-gestützte Detektion von Höflichkeit steckt jedoch noch in den Kinderschuhen. Die wenigen Versuche, Höflichkeit vorherzusagen, existieren vor allem fürs Englische und arbeiten auf Texten anstelle von Gesprächsdaten (Danescu-Niculescu-Mizil et al. 2013; Kumar 2014). Darüberhinaus leiden sie unter meist binären Klassifizierungsschemata, die vor allem auf gesprochensprachlichen Daten die Komplexität von Höflichkeit nicht abbilden können (Danescu-Niculescu-Mizil et al. 2013).

In dieser Arbeit wird ein Multilayer-Annotationsschema für Frage-Antwort-Paare basierend auf einer qualitativen Analyse des CallHome Korpus (Canavan 1997) entwickelt, das als Basis für einen komplexeren Politeness-Classifizierer für deutsche Gesprächsdaten dienen soll. Durch die Betrachtung von Adjazenzpaaren und Präferenzgestaltung (Schegloff 2007) wird das Konzept von Höflichkeit dekonstruiert. Das dreilagige Annotationsschema besteht aus folgenden Komponenten: In einem ersten Analyse-Schritt werden Fragen und Antworten in Gesprächsdaten identifiziert, in einer zweiten Layer werden Dispräferenzmarker festgehalten, die sich auf der Sprachoberfläche der Antworten befinden. Eine dritte Annotationsschicht bezieht den weiteren Gesprächskontext in Form von Vorlaufsequenzen und Rezeption an dritter Position mit ein. In einer Pilotstudie wird die Annotation erprobt und evaluiert.

**Literatur:** • Canavan, A., Graff, D., & Zipperlen, G. (1997). Callhome german speech. Linguistic Data Consortium. • Danescu-Niculescu-Mizil, C., Sudhof, M., Jurafsky, D., Leskovec, J., & Potts, C. (2013). A computational approach to politeness with application to social factors. arXiv preprint arXiv:1306.6078. • De Jong, M., Theune, M., & Hofs, D. (2008, May). Politeness and alignment in dialogues with a virtual guide. In Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1 (pp. 207-214). • Kumar, R. (2014, May). Developing Politeness Annotated Corpus of Hindi Blogs. In LREC (pp. 1275-1280). • Schegloff, E. A. (2007). Sequence organization in interaction: A primer in conversation analysis I (Vol. 1). Cambridge university press.

---

## Empirical evidence for pejoration of ‘woman’-terms: Challenges of estimating inter-annotator agreement with complex annotation

---

Carla Sökefeld & Heike Zinsmeister

*Universität Hamburg*

carla.soekefeld@uni-hamburg.de, heike.zinsmeister@uni-hamburg.de

This poster presents a small dataset of manually semantically annotated historical dictionary entries and discusses the challenges of estimating inter-annotator agreement given a complex and – to some extent – subjective annotation task. The annotation study is part of a PhD thesis on semantic change: Many German terms denoting women have historically undergone a pejoration, that is, their meaning has changed to have a more negative bias (e.g., *Frau* has changed from ‘woman with high social status’ to just ‘woman’). To track the semantic change of 31 different German nouns denoting either women or men, a collection of 160 entries in eight historical dictionaries (16th-19th century) were annotated by two annotators (148,000 tokens; entries vary greatly in size from a minimum of two tokens to more than 21,000 tokens). Previous research has shown that women tend to be portrayed stereotypically and negatively while men are shown in a more positive light (Warnke 1993). The tagset crucially differentiates between two levels of semantic information: The *meaning components* given for the lexeme and the portrayal of men and women in *language examples*. This distinction allows for a nuanced analysis: We can trace the semantic change as reflected in the dictionaries and further gain insights into gender stereotypes conveyed in usage examples (see also Nübling 2009). The dataset will be made available on Zenodo together with detailed annotation guidelines.

We will discuss the challenges of estimating a chance-corrected inter-annotator agreement score on this data set due to the variable extensions of the annotation units as well as potential ambiguity or subjectivity (Barteld et al. 2014) and the option of multiple valid annotations (Marchal et al. 2022).

**References:** • Barteld, F.; Ilden, S.; Schröder, I.; Zinsmeister, H. (2014): Annotating descriptively incomplete language phenomena. In *Proceedings of LAW*, 99–104. • Marchal, M.; Scholman, M.; Yung, F.; Demberg, V. (2022): Establishing Annotation Quality in Multi-label Annotations. In *Proceedings of COLING*, 3659–3668. • Nübling, D. (2009): Zur lexikografischen Inszenierung von Geschlecht – ein Streifzug durch die Einträge von *Frau* und *Mann* in neueren Wörterbüchern. *ZGL* 37/3, 593–633. • Warnke, Ingo (1993): Zur Belegung von ‘Frau’ und ‘Weib’ in historischen deutschen Wörterbüchern des 16. und 17. Jahrhunderts. In B. Hufeisen (ed.), *“Das Weib soll schweigen. . .” (1. Kor.14,34): Beiträge zur linguistischen Frauenforschung*. Frankfurt a.M.: P. Lang, 127–152.

---

## Why (not) vegan? An NLP-based Investigation of Moral Sentiment and Storytelling in the Vegan Discourse

---

**Rebecca Pichler, Neele Falk & Gabriella Lapesa**

*Universität Stuttgart*

rebpic@gmail.com, neele.falk@ims.uni-stuttgart.de,

gabriella.lapesa@ims.uni-stuttgart.de

With veganism being more present in society and thus frequently debated, vegans are often criticized as do-gooders. This work employs NLP methods to investigate the moral loading of the Vegan Discourse and its interaction with the use of storytelling: **how moral is the Vegan Discourse? Which moral sentiments are expressed most? Does storytelling enhance the expression of morality in text?**

**Dataset & Methods.** We created a corpus with roughly 10'000 comments on articles about veganism published by the New York Times. To assign each comment moral labels, we implemented a BERT multilabel model trained on the Moral Foundations Twitter Corpus<sup>6</sup>, with reliable performances compared to related work. The comments were classified as containing storytelling with a classifier from our previous work. We further validated both classifiers on a manually annotated gold-standard (100 comments).

**Results.** The Vegan Discourse itself was proven to be significantly moral in the sense of the five moral foundations since more than a tenth of the corpus was classified reliably with morality. The challenge of implicitness of morality is reflected in the comments' different moral loadings throughout the corpus. The two most frequently assigned foundations are "Care-Harm" and "Purity-Degradation". We find that comments with storytelling were associated with moral loadings, but only weakly. Since these result from implicitness, storytelling is rather enhancive of morality in the corpus as a marker of detectable implicitness.

**Conclusion.** Our results show that that both concepts, moral sentiment and storytelling, are significantly observable and intertwined in the Vegan Discourse as represented in this work. Reasonably, there are many (non-)vegans that tell their personal opinions in stories to make their way of life more palpable for others.

**References:** • Pickett, S. (2021). Veganism, Moral Motivation and False Consciousness. *Journal of Agricultural and Environmental Ethics*, 34(15). • Hoover, J., Portillo-Wightman, G., Yeh, L., Havaladar, S., Davani, A. M., Lin, Y., Kennedy, B., Atari, M., Kamel, Z., Mendlen, M., Moreno, G., Park, C., Chang, T. E., Chin, J., Leong, C., Leung, J. Y., Mirinjian, A., & Dehghani, M. (2020). Moral Foundations Twitter Corpus: A Collection of 35k Tweets Annotated for Moral Sentiment. *Social Psychological and Personality Science*, 11(8), 1057–1071. • Kobbe, J., Rehbein, I., Hulpus, I., & Stuckenschmidt, H. (2020). Exploring Morality in Argumentation. *Proceedings of the 7th Workshop on Argument Mining*, 30–40. • *Reports of personal experiences and stories in argumentation: Datasets and analysis.* (2021).

**Solidarität im Migrationsdiskurs:  
Klassifizierung von Leserinnen- und Leserbriefen mit DistilBERT**

---

**Lisanne Rüh**

*Universität Hildesheim*  
ruehli@uni-hildesheim.de

Soll eine sehr große Menge an textlichen Daten klassifiziert werden, muss dies automatisiert erfolgen. Können die Daten zudem nicht nach strukturellen Merkmalen gefiltert werden, weil ein so abstraktes und schwer fassbares Konzept wie *Solidarität* untersucht werden soll (enthält ein Text *Solidarität* oder nicht?), kann ein Machine-Learning-Algorithmus wie DistilBERT (Sanh et al. 2020) für die Textklassifikation verwendet werden. In meiner Projektarbeit habe ich daher untersucht, ob DistilBERT in der Lage ist, solidarische Äußerungen im Migrationskontext in Leserinnen- und Leserbriefen zu erkennen.

Basierend auf der Begriffskonzeption von *Solidarität* nach Kneuer et al. (2021) sollte dies mithilfe zweier Klassifizierer erreicht werden, die jeweils auf einen der Solidaritätsbestandteile *Adversity* (eine Widrigkeit, die einer Person widerfährt) und *Contribution* (ein Beitrag, der zur Erreichung eines gemeinsamen Ziels geleistet wird) trainiert wurden. In Kombination sollten diese beiden Klassifizierer dann zur Identifikation der Gesamtkategorie *Solidarität* genutzt werden. Grundlage für die Trainings- und Testdaten bildete ein bereits zum Themenfeld Migration vorgefiltertes Korpus mit Leserinnen- und Leserbriefen aus verschiedenen deutschen Zeitungen.

Die Untersuchung hat gezeigt, dass die Klassifikation von kleineren Textausschnitten bereits gute bis sehr gute Ergebnisse erzielen konnte (F-Maß = 0,786 u. 0,908), aber die Gesamtkategorie *Solidarität* in vollständigen Briefen noch nicht in einem zufriedenstellenden Maße erkannt wurde (F-Maß = 0,364). Die Erfassung der Teilkategorien *Adversity* und *Contribution* funktionierte hier dagegen besser (F-Maß = 0,630 bzw. 0,625). Insgesamt wurden aber auch viele positive Beispiele, die Äußerungen der jeweiligen (Teil-)Kategorie enthalten, nicht von den Klassifizierern identifiziert. Bei der Erkennung von *Solidarität* gilt außerdem, dass Briefe, die als *Solidarität* enthaltend eingeordnet wurden, oftmals das falsche Label erhielten. Es ist hier somit noch weitere Arbeit notwendig, um die Genauigkeit der Klassifizierungsergebnisse, insbesondere bei der Bewertung gesamter Briefe, zu verbessern.

**References:** • Kneuer, M.; Corsten, M.; Schammann, H.; Kahle, P.; Wallaschek, S. & Ziegler, F. (2021). Claiming Solidarity. A multilevel discursive reconstruction of solidarity. *European Journal of Social Theory*. • Sanh, V.; Debut, L.; Chaumond, J. & Wolf, T. (2020). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. In *EMC2, the 5th Workshop on Energy Efficient Machine Learning and Cognitive Computing*.



**Fachinformationsdienst (FID) Linguistik:  
Serviceportfolio und Informationsangebote**

---

**Vanya Dimitrova & Heike Renner-Westermann**

*Goethe-Universität Frankfurt*

v.dimitrova@ub.uni-frankfurt.de,

h.renner-westermann@ub.uni-frankfurt.de

Der Fachinformationsdienst (FID) Linguistik ist eine Serviceeinrichtung für die Allgemeine und Vergleichende Sprachwissenschaft und die einzelphilologischen Sprachwissenschaften an deutschen Hochschulen und Forschungseinrichtungen. Der FID wird seit 2017 an der Universitätsbibliothek Frankfurt mit DFG-Förderung aufgebaut. Über das Lin|gu|is|tik-Portal (<https://linguistik.de/>) bietet der FID Linguistik Services in den Bereichen Recherche, Linked Open Data (LOD), E-Publishing, Erwerbung und Lizenzierung. Die Angebote sowie die Funktionalitäten des Portals werden konsequent weiter ausgebaut.

Im Rahmen der Präsentation werden die aktuellen Entwicklungen vorgestellt: z. B. die Implementierung einer semantisch gestützten Suche, die Realisierung neuer Module mit themenspezifischen Funktionalitäten sowie die Einrichtung geeigneter Exportschnittstellen. Vorangetrieben wird auch die Vernetzung des Lin|gu|is|tik-Portals mit der LOD-Cloud, die seit Jahren einen Arbeitsschwerpunkt darstellt.

Der FID Linguistik führt gezielte Maßnahmen durch, um die Sichtbarkeit, Auffindbarkeit und Verfügbarkeit linguistisch relevanter, digitaler Sprachressourcen zu erhöhen. So werden durch den Einsatz und die Weiterentwicklung ontologiebasierter NLP-Werkzeuge frei verfügbare Sprachressourcen analysiert und mit inhaltlichen Metadaten angereichert (Abromeit, 2022). Über die Verknüpfung linguistischer Veröffentlichungen mit den darin behandelten Forschungsdaten wird die Sichtbarkeit sowohl der Daten als auch ihrer wissenschaftlichen Analyse erhöht (Renner-Westermann, 2018). Darüber hinaus vermittelt der FID individuelle Lizenzen für kommerzielle Textkorpora an Forscher\*innen innerhalb Deutschlands.

Die Präsentation gibt eine Möglichkeit, die FID-Services besser kennenzulernen sowie einen Blick hinter die Kulissen zu werfen und zu erfahren, wie bibliothekarisches Know-how, Semantic-Web-Technologien und Methoden der angewandten Computerlinguistik aufeinandertreffen.

**References:** • Abromeit, F (2022). The Annohub Web Portal. In *Proceedings of the 8th Workshop on Linked Data in Linguistics (LDL-2022)*, Marseille, France, June 2022, 1-6. • Renner-Westermann, H. (2018). Fachinformationsdienst Linguistik zwischen Innovation und Tradition: Forschungsdaten in der Linguistik. *Zeitschrift für Bibliothekswesen und Bibliographie* 65 (2-3), 90-93.

---

## Multi-word expressions and language efficiency: an information-theoretic account

---

Stefan Fischer<sup>1</sup>, Peter Fankhauser<sup>2</sup> & Elke Teich<sup>1</sup>

<sup>1</sup>*Universität des Saarlandes*, <sup>2</sup>*Leibniz-Institut für Deutsche Sprache*  
stefan.fischer@uni-saarland.de, fankhauser@ids-mannheim.de,  
e.teich@mx.uni-saarland.de

Multi-word expressions (MWEs) are a cornerstone in conventionalized language use and vital for the perceived fluency of a message (Fillmore 1979). From a processing perspective, MWEs seem to have an advantage over arbitrary word sequences due to highly predictable transitions from one word to the next, or they may be perceived as wholes (see e.g. Siyanova-Chanturia et al. 2017).

The emergence and use of specific MWEs is typically context-dependent and register-specific. In our work, we investigate MWEs in the scientific domain from a diachronic perspective, asking what is the contribution of MWEs in the development of “scientific language” (here: English)? We assume that over time scientific English develops an optimal code for scientific expert communication characterized by high information density (Halliday 2004; Teich et al. 2021).

Using a large diachronic corpus of English scientific texts (Fischer et al. 2020), we work in a data-driven fashion using various established word association measures (e.g. log-likelihood, PMI) to identify and classify MWEs by time periods (e.g. 50-year periods). In a complementary step, we account for the environments of words using selected computational language models (statistical models, embeddings; cf. Fankhauser & Kupietz 2022). On this basis, we then analyse the informational characteristics of MWEs diachronically: The more conventionalized an MWE becomes, the lower its surprisal (higher predictability of the MWE) and the lower the uncertainty about an upcoming word within the MWE (entropy). We expect to see that while specific MWEs come and go over time, during their life cycles they will exhibit surprisal/entropy reduction, thus contributing to language efficiency.

**References:** • Fankhauser, P. & Kupietz, M. (2022). Count-Based and Predictive Language Models for Exploring DeReKo. *Proceedings of LREC 2022 Workshop CMLC-10 2022*, 27–31. • Fillmore, C. J. (1979). On fluency. In Fillmore, C. J., Kempler, D. & Wang, W. S.-Y. (eds.), *Individual Differences in Language Ability and Language Behavior*, 85–101. Academic Press. • Fischer, S., Knappen, J., Menzel, K. & Teich, E. (2020). The Royal Society Corpus 6.0: Providing 300+ Years of Scientific Writing for Humanistic Study. *Proceedings of LREC 2020* (online). • Halliday, M. A. K. (2004). The Language of Science. In Webster, J. (ed.), *The Collected Works of M. A. K. Halliday*. London: Continuum. • Siyanova-Chanturia, A., Conklin, K., Caffarra, S., Kaan, E. & van Heuven, W. J. (2017). Representation and processing of multi-word expressions in the brain. *Brain and Language*, 175:111–122. • Teich, E., Fankhauser, P., Degaetano-Ortlieb, S. & Bizzoni, Y. (2021). Less is More/More is Diverse: On the Communicative Utility of Linguistic Conventionalization. *Frontiers in Communication, Topic: Rational Approaches in Language Science*. 5:620275 (online).

---

**Collecting and Investigating Features of Compositionality Ratings**

---

**Sabine Schulte im Walde***Universität Stuttgart*

schulte@ims.uni-stuttgart.de

The automatic prediction of multiword expression compositionality has been a long-standing task in natural language processing. Our focus of interest is on the compositionality of noun compounds such as *climate change* and *crocodile tears* in English, and *Ahornblatt* and *Fliegenpilz* in German, whose representation, processing and modelling has received an immense attention across disciplines and languages.

Developing computational models of compositionality typically goes hand in hand with creating reliable lexical resources to utilise as gold standards for repeated intrinsic evaluation. Not much research however has looked into whether and how much both the gold standards and the prediction models vary according to properties of the compound targets within the lexical resources. Our study provides two contributions to bring forward both theoretical and computational investigations of compositionality for noun compounds: (1) We present a novel collection of compositionality ratings for 1,099 German noun compounds where – differently to previous related work – we asked the human judges to provide (a) paraphrases of the compound meanings, (b) features of constituents contributing to the compound meanings, (c) judgements on the strength of hypernymy relations between the compounds and their head constituents, and (d) judgements on the strengths of the concreteness of the compounds and their constituents, before they provided their judgements on the compounds' degrees of compositionality with regard to the respective constituents. (2) We present a series of analyses on interactions of compound and constituent properties (such as frequencies, productivities, ambiguity, hypernymy and concreteness) with regard to the compounds' degrees of compositionality, and the influence of these compound and constituent properties on computational vector-space results for predicting compositionality. As basis for our study we rely on the predominantly used lexical resources of noun compound compositionality for English (Reddy et al., 2011; Cordeiro et al., 2019) and German (Schulte im Walde et al., 2013; 2016).

**References:** • Cordeiro, S. & Villavicencio, A. & Idiart, M. & Ramisch, C. (2019). Unsupervised compositionality prediction of nominal compounds. *Computational Linguistics* 45(1), 1–57. • Reddy, S. & McCarthy, D. & Manandhar, S. (2011). An empirical study on compositionality in compound nouns. In *Proc. of IJCNLP*, 210–218. • Schulte im Walde, S. & Müller, S. & Roller, S. (2013). Exploring vector space models to predict the compositionality of German noun-noun compounds. In *Proc. of \*SEM*, 255–265. • Schulte im Walde, S. & Häty, A. & Bott, S. & Khvtisavrivili (2016). G<sub>ost</sub>-NN: A representative gold standard of German noun-noun compounds. In *Proc. of LREC*, 2285–2292.

---

## Composing noun compounds in vector spaces

---

**Chris Jenkins**

*Universität Stuttgart – Institut für Maschinelle Sprachverarbeitung*  
christopher.jenkins@ims.uni-stuttgart.de

Novel uses and meanings occur within linguistic communities over time -- and are not limited to individual words. Arbitrarily long phrases can have semantically opaque meanings, as with idioms. The variance in compositionality among multi-word expressions presents challenges for the computational representation of their meanings.

In this poster session I will present my work on comparative approaches for semantically representing noun-noun compounds using contextual word embeddings. Multi-word expressions like noun-noun compounds in English and German can vary in terms of compositionality, attested frequency in corpora, etc. These variations problematize the typical averaging of word embeddings to represent a whole compound, as these may be better suited to represent e.g. the older sense of the compound *Donnerwetter* (thunder + weather) meaning 'thunderstorm', but not the newer sense as an exclamation (approx.: 'my goodness!'). I will compare adding and multiplying embeddings (Freremann, Lapata 2016 and Mitchell, Lapata 2010) as well as an explicitly functional representation (Baroni, et al. 2010) of noun compounds' modifier affecting the representation of the head. Parameterized composition functions, upstream of a task involving some property (e.g. predicting degree of compositionality, predicting the semantic relation between head and modifier) of noun compounds will also be explored (Dima 2016).

The representation of compounds is further complicated in many contemporary implementations of word embeddings, which use sub-word tokenizer models, such as *SentencePiece* (Kudo, Richardson 2018). The tokenizer may split input words into more than one 'sub-word', which may not correspond to any recognizable morpheme. A word embedding model that operates over sub-word tokens can flexibly handle out-of-vocabulary words, but it introduces an additional composition problem which I will also endeavour to address.

**References:** • M. Baroni, R. Bernardi, and R. Zamparelli. Frege in space: A program for composition distributional semantics. In LILT, Volume 9, 2014 - Perspectives on Semantic Representations for Textual Inference. CSLI Publications, 2014. • T. Kudo and J. Richardson. Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing, 2018. • J. Mitchell and M. Lapata. Composition in distributional models of semantics. • Cognitive Science, 34(8):1388–1429, 2010. • V. Shwartz and I. Dagan. Still a Pain in the Neck: Evaluating Text Representations on Lexical Composition. *TACL*, 7:403–419, 07 2019. • C. Dima. On the compositionality and semantic interpretation of english noun compounds. Proceedings of the 1st Workshop on Representation Learning for NLP, 2016.

---

## Evaluation of neural coreference annotation of simplified German

---

Sarah Jablotschkin, Nele Benz & Heike Zinsmeister

*Universität Hamburg*

Griet.v.Koeln@severinstor-koeln.de, J.a.Rhein@eigelstein-koeln.de,  
hennes@zoo-koeln.de

Simplified variants of German (i.e. *Leichte Sprache* and *einfache Sprache*; cf. Bredel/Maaß 2016, Baumert 2019) focus on lexical and syntactic simplification in order to make written information accessible to low-literacy readers. Simplification efforts include using simple sentences and avoiding personal pronouns. However, this also affects coherence phenomena: Relations between nominal form, information status and anaphoric potential are different than we would expect them to be in Standard German (e.g. we would expect a pronoun as anaphor instead of the definite NP in example (1)).

- (1) [Die meisten Besucher] sind verkleidet. [Die Besucher] haben zum Beispiel bunte Kleidung. (LeiKo 1.5 core corpus)  
‘Most visitors are dressed up. Fore example, the visitors have colourful clothes.’

In our poster, we evaluate the effect of these differences on the performance of state-of-the-art coreference resolution developed on Standard German news texts. Our evaluation is based on the manually corrected core corpus of LeiKo 1.5 (doi: 10.5281/zenodo.6362739) (1,071 sentences / 1,401 coreference relations). The raw coreference resolution was created by the neural c2f coreference resolver (Schröder et al. 2021). We evaluate the performance quantitatively (v.8.01 of the CoNLL-2011/2012 reference scorer, Pradhan et al. 2014) and provide a detailed error analysis. C2f performs robustly on simplified German texts with respect to definite and demonstrative coreference but misses out, e.g., on example (1), which contains the indefinite expression *die meisten* (‘most’). Other error sources include speaker reference in reported speech and ambiguous reference. We will also discuss consequences for coreference annotation of a larger parallel corpus of simplified and Standard German currently under construction.

**References:** • Baumert, A. (2019). Mit einfacher Sprache Wissenschaft kommunizieren. essentials. Wiesbaden: Springer Fachmedien Wiesbaden GmbH. • Bredel, U. & C. Maaß (2016). *Leichte Sprache. Theoretische Grundlagen, Orientierung für die Praxis*. Berlin: Dudenverlag. • Pradhan, S., X. Luo, M. Recasens, E. Hovy, V. Ng & M. Strube (2014). Scoring Coreference Partitions of Predicted Mentions: A Reference Implementation. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 30–35. • Schröder, F., H. O. Hatzel & C. Biemann (2021). Neural End-to-end Coreference Resolution for German in Different Domains. In Proceedings of KONVENS 2021, 170–181.

---

## Der Genitiv schlägt zurück. Rektionsschwankungen bei Präpositionen

---

**Damaris Bendin**

*Universität Hildesheim*

bendind@uni-hildesheim.de

Aufbauend auf die Publikation von Di Meola (2009) untersucht diese korpuslinguistische Projektarbeit exemplarisch die Verbreitung „normwidriger“ Rektionen bei deutschen Primär- und Sekundärpräpositionen. Standardsprachlich regieren die „prototypischen“ Primärpräpositionen den Dativ und/oder den Akkusativ, während Sekundärpräpositionen nach Genitiv-, Dativ- und Akkusativ-Rektion gruppiert werden. Sie entstammen ursprünglich einer anderen Wortart oder syntaktischen Struktur und befinden sich in einem Grammatikalisierungs-prozesses von Inhalts- zu Funktionswort. Di Meola (2009) belegte für alle untersuchten Präpositionen Rektionsschwankungen und stellte eine besondere Anfälligkeit der Akkusativ-Präpositionen und Präpositionen im Partizip Präsens sowie eine stärkere Tendenz zur neuen Genitiv-Rektion als zur neuen Dativ-Rektion fest. Die Forschungsfragen dieses Beitrags lauten: Haben sich diese Entwicklungen fortgesetzt? Inwieweit hält sich der Sprachgebrauch heute an normierte Rektionen? Für die Untersuchung wurden aktuelle Belege für 19 Präpositionen im Umfang von insgesamt 172.857 Sätzen aus pressesprachlichen Korpora des IDS (2022), COSMAS II, Archiv Tagged-T2, exportiert und mithilfe des RNN-Taggers (Schmid 2019) u.a. nach Kasus annotiert. Anschließend wurden nach Kasus getrennte Ergebnislisten erstellt, welche auf genormte Rektionen und die von Di Meola ermittelten davon abweichenden, "normwidrigen" Rektionen der Präposition überprüft wurden. Darüber hinaus wurden für Genitiv- und Dativ-Präpositionen zusätzlich mögliche Akkusativ-Rektionen in die Analyse einbezogen, diese stellten sich allerdings eher als Ausnahmefälle anstatt als Ergebnis einer möglichen sprachlichen Entwicklung heraus. Gleich den Ergebnissen von Di Meola, werden für alle untersuchten Präpositionen Rektionsschwankungen belegt. Die neue Rektion vor allem bei Präpositionen auf, für die sie bereits als Alternative akzeptiert wurde (z.B. *laut*, *entlang* und *dank*) aber auch für Präpositionen, deren neuer Kasus standardsprachlich „verboten“ ist (z.B. *während*), ist der neue Kasus vielfach belegt. Ebenso wurde die insgesamt stärkere Tendenz zur neuen Genitiv-Rektion gegenüber einer neuen Dativ-Rektion festgestellt. Eine besondere Anfälligkeit für Kasuswechsel von Sekundären Akkusativ-Präpositionen und Präpositionen in Form des Partizip Präsens konnte jedoch anhand der Daten nicht eindeutig nachgewiesen werden.

**References:** • Di Meola, Claudio (2009): Rektionsschwankungen bei Präpositionen – erlaubt, verboten, unbeachtet. In: Konopka, Marek; Strecker, Bruno (Hrsg.): *Deutsche Grammatik – Regeln, Normen, Sprachgebrauch*. Berlin, New York: Walter de Gruyter, S. 195-221. • Schmid, Helmut (2019): *Deep Learning-Based Morphological Taggers and Lemmatizers for Annotating Historical Texts*. DATeCH, Mai 2019, Brüssel.

---

## Tagger for Intensifiers – An Automatic Identification for Intensifiers of Adjectives

---

**Imge Yüzüncüoğlu**

*Ruhr-Universität Bochum*

Imge.Yuezuencueoglu@rub.de

*Intensifiers* are rapidly changing words or wordparts which can be used to amplify the semantic meaning of a statement as in (1a) or to downtone it as in (1b) (Stratton, 2020). There is no finite amount of them and analyzing intensifiers turns out to be time-consuming as every token in a corpus has to be annotated individually.

- (1) a. *Der Film ist sehr sehenswert.* “The movie is very worth seeing.”  
 b. *Der Film ist kaum sehenswert.* “The movie is hardly worth seeing.”

Until now there have been two ways for an automatic annotation of intensifiers. One is the usage of wordlists that contain intensifiers and the other is the usage of general POS-tagger that contain the PTKIFG-tag of the POS-tagset from Empirist (2015) like the POS-tagger SoMeWeTa by Proisl (2018). But both options need to be revised manually as the wordlists cannot distinguish if a word it finds is used as an intensifier or not. Additionally, the PTKIFG-tag is a collective tag for intensifiers, degree- and focusparticles. Therefore, the goal is to implement a tagger that can identify intensifiers of adjectives in a reliable way.

The here presented tagger for intensifiers (TafIn) does exactly that. TafIn is based on the Conditional Random Field algorithm and trained with manually annotated data, consisting of 468 blogposts from the TwiBloCoP corpora (2021). Different token information based on the definition of intensifiers like POS-tags and tokens alongside of the current token were used to generate abstract features to identify intensifiers in a text contextfree. To get a clear impression of TafIn’s results, it and two other baselines, the first one (B1) being the PTKIFG-tag within SoMeWeTa and the second one (B2) being a combination of a wordlist and heuristical methods, were evaluated. The results like the macroaverage F1-scores presented in the table below show that TafIn works nearly as twice as good as the previous methods, thus it can be used as a more reliable annotation tool for intensifiers of adjectives.

	B1: SoMeWeTa	B2: Wordlist	<b>TafIn</b>
Macroaverage F1-Score	0.47	0.44	<b>0.88</b>

**References:** • EmpirIST. 2015. GSCL Shared Task: Automatic linguistic annotation of computer-mediated communication / social media. *EmpirIST* 2015. • Proisl, T. (2018). SoMeWeTa: A part-of-speech tagger for German social media and web texts. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 665–670. • Stratton, J. M. (2020). Adjective Intensifiers in German. *Journal of Germanic Linguistics* 32(2), 183–215.

---

## Zweckentfremdung von mathematischen alphanumerischen Symbolen zur Textgestaltung auf Twitter und ihre Probleme bei der Verwendung von Screenreadern

---

Selina Becker-Gierling & Kilian Evang

Heinrich-Heine-Universität Düsseldorf

segie102@uni-duesseldorf.de, evang@hhu.de

Twitter ist als textbasierte Social-Media-Plattform bei Menschen mit Sehbehinderung eine beliebte Austausch-Plattform. Die Möglichkeit zur Textgestaltung durch Auswahl verschiedener Schriftarten und -schnitte besteht nicht. Dies wird teilweise umgangen, indem spezielle, für mathematische Variablen gedachte Zeichen innerhalb Unicodes verwendet werden. Jedoch sind Tweets, in denen solche Zeichen zur Textgestaltung verwendet werden, für Screenreader-Nutzer\*innen mit hoher Wahrscheinlichkeit unverständlich (siehe Abbildung 1).

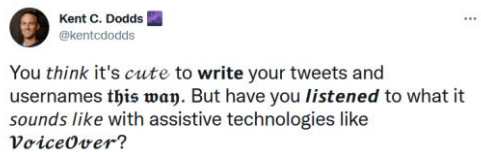


Abbildung 1: Tweet-Beispiel zu zweckentfremdeten mathematische alphanumerische Symbolen auf Twitter

Diese Benachteiligung steht im Widerspruch zu Artikel 21 der UN-Behindertenrechtskonvention. Um das Ausmaß der Einschränkung einzuschätzen, wurde zunächst eine Korpusanalyse an 400 deutschen, zufällig ausgewählten Tweets, die mathematische alphanumerische Symbole enthalten, durchgeführt. Dabei wurde untersucht, welche kommunikativen Funktionen diese Symbole erfüllen, und sie anhand dessen kategorisiert. Die Analyseergebnisse zeigen, dass die Symbole am häufigsten in den Funktionskategorien ‚Werbung‘ und ‚Abkürzungen‘ vorkamen. Des Weiteren wurde die Qualität von drei systemintegrierten, kostenlosen Screenreadern (Narrator von Microsoft, VoiceOver von Apple und TalkBack von Google) im Umgang mit mathematischen alphanumerischen Symbolen an exemplarisch ausgesuchten Beispielen vergleichend untersucht und danach bewertet, wie verständlich sie sie vorlesen: Stufe 0 (nicht ausgelesen), Stufe 1 (teilweise ausgelesen), Stufe 2 (vollständig ausgelesen) oder Stufe 3 (vollständig ausgelesen, mit veränderter Betonung).

Die Ergebnisse zeigen, dass Narrator am häufigsten nur Stufe 0 erreicht, VoiceOver am häufigsten Stufe 1 und TalkBack am häufigsten Stufe 2.

**References:** • Dodds, K. C. (2019). Tweet. <https://twitter.com/kentcodds/status/1083073242330361856> [20.09.22]