Research Statement

Nathan Schneider

May 2022

My research is in natural language processing (NLP). NLP is a well-established field which represents the intersection of computational linguistics and AI. Encompassing many different computational problems for intelligent processing of textual language data, NLP drives developments in human language technologies such as machine translation and dialogue systems (think: Siri or Alexa).

In many ways, language capacity is the crown jewel of human cognition. Language likewise occupies a central place in artificial intelligence research. Computers, by default, are very good at memorizing and number-crunching, but very bad at making human-like abstractions and inferences necessary for understanding or producing language in a meaningful way. The field of NLP, therefore, is vast, and the hardest problems are far from solved, especially for less-researched languages and language varieties.

My own work contributes novel resources and algorithms for problems in the **linguistic foundations** of NLP, with an emphasis on **meaning**. I am fortunate to be able to approach these problems both as a linguist and as a computer scientist. On the linguistics side, I work on formalizing meaning in a way that can be done on a large scale in corpus data. On the computer science side, I develop algorithms that try to emulate human linguists in analyzing text. The questions of how to formalize linguistic analyses of particular sentences, and how to analyze sentences automatically, are closely intertwined.

The chief contributions of my work are **resources and methods for analyzing text in linguistically-motivated ways**—including human annotation, automatic analysis algorithms (taggers, parsers), and approaches to evaluation. Much of this research contributes scientific/theoretical/methodological advances as well as data and/or software resources (released under permissive licenses). Nearly all of it is highly collaborative, in most cases with a student as lead author.

Rather than focus on one particular problem or topic in the linguistic foundations of NLP, I work on a variety of topics connected by common themes. These themes include: broad-coverage approaches to meaning representation; linguistic annotation of corpora; meaning at the intersection of lexicon and grammar (e.g., prepositions, multiword expressions); and the so-called "long tail" of language (most individual words/ constructions/meanings are infrequent).

1 Lexical Semantic Disambiguation

To a first approximation, understanding the meaning of a sentence involves understanding the meanings of individual words, and how they relate to one another as signaled by grammar. But attempting to formalize this algorithmically raises many challenges. For example: What counts as a word with its own meaning? Some sentences contain **multiword expressions** (MWEs)—including idioms like "kick the bucket" meaning 'die', and complex nouns and verbs ("hot dog", "pay attention", "give in") where the individual words would have other meanings in other contexts. Detecting such multiword units automatically in context would therefore aid semantic interpretation. However, some of these expressions are not completely fixed—they can contain **gaps** with intervening words, as in "**pay** really close **attention**", or "little **attention** was **paid**"—which makes detection challenging. The sheer number and variety of MWEs in the vocabulary is an obstacle as well. MWEs are a topic of NLP research, but such research typically focuses on particular kinds (e.g., noun-noun compounds). To get a better understanding of the scope of the problem in English, I developed a method for *comprehensive annotation* of MWEs, and released a corpus with such annotations (**Schneider** et al., 2014b). Then, I introduced a new formalization of tagging for MWEs containing gaps, and evaluated statistical models for doing this tagging automatically (**Schneider** et al., 2014a). This work has been widely cited and built upon by other MWE researchers.[1]

A second challenge in understanding word meanings is **sense ambiguity**, which affects not just single-word expressions but also MWEs (e.g., "holding up" could refer to endurance, imposing a delay, or a bank robbery). Whereas a traditional approach to word sense disambiguation requires dictionary definitions and many annotated examples for every word in the vocabulary, **supersense** disambiguation is based on a general coarse-grained set of semantic class labels which are not specific to the vocabulary of any particular language, but help disambiguate any noun or verb (**Schneider** et al., 2012, *inter alia*). I saw an opportunity to connect the MWE identification and supersense tagging tasks, introducing a unified dataset and model (**Schneider** and Smith, 2015) and organizing a SemEval challenge competition in which a number of teams built their own systems (**Schneider** et al., 2016).

After working on supersense disambiguation for nouns and verbs, I became fascinated with the ambiguity of **prepositions**. Though they do not contribute semantic *content*, prepositions can be key signals of the semantic *relation* that holds between two other words in the sentence: for example, "a dog **with** a hat" is a relation between an entity and some characteristic or part of that entity (in this case, an article of clothing). Yet this is different from the relations in "dining **with** friends", "dining **with** chopsticks", or "dining **with** style". Since 2014, joined principally by Jena Hwang (Allen Institute for AI) and Vivek Srikumar (University of Utah) along with many other collaborators, I have been spearheading an effort to make sense of the lexical semantics of such connectives by defining supersense categories for the relevant semantic relations and annotating corpora in English and other languages. Our framework is called the Semantic Network of Adposition and Case Supersenses (SNACS). It is a comprehensive approach—all types and tokens of adpositions are annotated in a corpus, including rare and nonprototypical usages. **Schneider** et al. (2018) presents the main approach and English datasets/disambiguation classifiers; we have also written an extensive annotation manual (**Schneider** et al., 2020). SNACS is probably the most comprehensive unified analysis of preposition meanings to date, with respect to coarse-grained disambiguation; and it employs a novel two-layer annotation scheme (which we call the *construal analysis*) to capture richer generalizations. Members of my lab have picked up numerous research threads relating to possessives (Blodgett and **Schneider**, 2018), crowdsourcing annotations (Gessler et al., 2020), and integration with other meaning representations (Prange et al., 2019). As part of an NSF-supported collaboration, my lab annotated prepositions in a corpus of Reddit posts by nonnative English speakers, which will support the study of crosslinguistic influences (Kranzlein et al., 2020). Most ambitiously, I have worked with many collaborators to adapt SNACS to corpora in other languages.[2] This multilingual annotation has targeted translations of *The Little Prince*, and with this parallel corpus we are now in a position to examine crosslinguistic similarities/differences in adposition use in systematic fashion. We have also developed a website called Xposition for browsing adpositions and their meanings in multiple languages (Gessler et al., 2022).[3]

An integral part of the development of the aforementioned approaches to lexical semantic analysis is the STREUSLE corpus of annotated English web reviews,[4] which I maintain as the annotation policies evolve.

In addition to continuing to expand and refine SNACS as an annotation framework, it opens up possibil-

[1]I continue to be interested in improving both our linguistic conceptualization and algorithms for identifying MWEs (more recent contributions in Qu et al., 2015; Walsh et al., 2018; Liu et al., 2021).

[2]In addition to English, we have also made significant progress for adpositions/case in 4 languages and counting: Mandarin Chinese (Peng et al., 2020), Korean (Hwang et al., 2020), Hindi (Arora et al., 2022), and German (Prange and **Schneider**, 2021).

[3]`http://xposition.org`

[4]`https://github.com/nert-nlp/streusle`

ities for answering new research questions. A recent collaboration (with the University of Utah and HUJI) examined learned embedding representations of prepositions in the BERT language model. The question was whether BERT representations approximate semantic senses for prepositions and other highly ambiguous words. We developed new techniques to explore the embedding space as a more direct alternative to training probing classifiers (Karidi et al., 2021).

2 Structured Meaning Representations

Moving beyond the word level, computational linguists have developed a number of different symbolic formalizations of sentence-level meaning, including logic-based approaches and graph-based approaches.[5] My research on structured meaning representation design and parsing has targeted 3 graph-based frameworks: Abstract Meaning Representation (AMR), frame semantics in FrameNet, and Universal Conceptual Cognitive Annotation (UCCA). All of these capture some notion of predicate-argument structure as well as other kinds of semantic relations within sentences. They aim for broad coverage of within-sentence semantic relations in general-purpose text (as opposed to just annotating roles of verbs, for example).

I take the perspective that meaning is rich and multifaceted, and it is not practical to capture every detail of meaning in a single representation. Rather, each meaning representation reflects design tradeoffs and will be more or less appropriate for certain goals (ease of annotation and accurate parsing; different kinds of generalizations or applications; monolingual vs. multilingual, etc.). In my view it is important to consider what different representations bring to the table, how they relate to one another, and how they can be improved.

I have promoted AMR, FrameNet, and UCCA in the research community not only through specific research publications, but also with tutorials about meaning representation and parsing with these frameworks (**Schneider** et al., 2015; Baker et al., 2015; Abend et al., 2020a).

AMR. Beginning in 2012 I was part of the team that designed AMR for large-scale corpus annotation (Banarescu et al., 2013, 2015; Bonial et al., 2018). AMR combines several aspects of within-sentence meaning, including predicate-argument structure, named entity recognition, and coreference. While AMR makes some compromises relative to formal logic, it does so for practicality: it is simple enough for human annotators to apply it to a large volume of text—over 60,000 English sentences have been annotated from scratch.

This large resource has facilitated many parallel research threads, among them parsing (sentence-to-graph), generation (graph-to-sentence), applications such as summarization and human-robot dialogue, automata theory for graphs, and efforts to improve the sophistication of the representation, and to adapt it to other languages. To date, our AMR Bibliography documents 180+ scientific publications that use AMR, including several Ph.D. dissertations. Two of my own advisees completed AMR-based dissertations: Emma Manning on the evaluation of AMR-to-text generation systems (e.g., Manning et al., 2020; Manning and **Schneider**, 2021), and Austin Blodgett on linguistically-motivated characterizations of AMR compositionality and parsing (e.g., Blodgett and **Schneider**, 2019).

To highlight one technical challenge: The English AMR annotators provided an entire graph given an entire sentence, without specifying alignments between particular words and units of the graph. This is a source of difficulty for parsers and other algorithms that rely on some notion of a compositional derivation. I have worked on linguistically-motivated approaches to annotating and extracting alignments, first using dependency syntax (Szubert et al., 2018), and more recently with simpler structures that can be induced well

[5] Work in this vein can be found, for example, at the Designing Meaning Representations workshop series (Xue et al., 2019, 2020) and the Meaning Representation Parsing shared tasks (Oepen et al., 2019, 2020). Distributional approaches to text representation, such as in neural language models, have proved indispensible for accurate broad-coverage parsing. Using symbolic representations designed by linguists to better understand these "black-box" neural models is an important complementary research direction.

automatically (Blodgett and **Schneider**, 2021). We hope to develop new AMR parsers that take advantage of high-quality explicit alignments for accurate and interpretable performance.

Efforts are also ongoing to improve the amount of detail captured by AMR. For example, I was involved in a project led by former Ph.D. student Lucia Donatelli that proposed augmenting AMR with tense and aspect semantics, elements of grammatical meaning that were not included in the original annotations (Donatelli et al., 2018). In another recent collaboration, I contributed to the design of DocAMR, which represents aspects of meaning that cross sentence boundaries (Naseem et al., 2022). Finally, my current Ph.D. student Shira Wein is advancing AMR as a cross-lingual representation, for example as a tool for studying translation divergence (Wein and **Schneider**, 2021).

Frame-Semantic Parsing. FrameNet is a lexicographic resource instantiating Fillmore's theory of frame semantics for a large swath of the vocabulary of English (e.g., Fillmore and Baker, 2009). The analysis of frame-evoking terms and labeling their semantic roles was framed as a full-sentence structured analysis problem (Baker et al., 2007). I and others developed a body of work on probabilistic structured prediction models, including a system called SEMAFOR, for this challenging task (Das et al., 2010, 2014; Kshirsagar et al., 2015). This topic has seen continued research, and I think stands to benefit further from some of the recent findings for other structured linguistic representations.

UCCA. The UCCA framework (Abend and Rappoport, 2013) takes a typologically-grounded approach to semantic annotation, placing a premium on crosslinguistic stability: that is, parallel sentences convey similar meanings by virtue of being translations, and should therefore have similar semantic structures despite morphosyntactic divergences (e.g., Sulem et al., 2015). The basic UCCA annotations are coarser-grained than AMR or FrameNet, and do not require a predicate lexicon or explicit mapping to syntax. These properties make UCCA an attractive framework for comparing different languages' strategies for realizing semantic relations in grammar.

Lead UCCA creator Omri Abend (HUJI) and I obtained grant funding to collaborate on meaning representation design, annotation, and parsing. After teaching UCCA annotation in my semantic representation course, I helped with a significant revision of the annotation guidelines (Abend et al., 2020b). Using the improved annotations, we sought to determine to what extent the structures captured in UCCA are essentially notational variants of existing information in syntactic and lexical semantic annotations. I implemented a rule-based converter to transform the lexical and syntactic annotations in the STREUSLE corpus (§1) into UCCA graphs. This performed surprisingly well—on par with state-of-the-art statistical UCCA parsers—and examining the output revealed systematic similarities and differences between what is captured in the different representations (Hershcovich et al., 2020). Finally, with a view toward enhancing UCCA's representation of predicate-argument relations by adding general-purpose semantic roles, my student Jakob Prange led an effort establishing that SNACS supersenses (§1) can be integrated within UCCA structures. A comparison of several joint and multitask neural architectures showed that the structure-parsing task and role-labeling task can support one another (Prange et al., 2019).

In the future I anticipate delving into crosslinguistic questions using the UCCA framework. UCCA-annotated corpora exist for several languages, some of which also have SNACS resources. In addition, the UCCA team has explored applications to learner language, compatible with my lab's growing interest in nonnative language and crosslinguistic influences.

3 Syntactic Parsing: New Genres and Long Tails

Syntactic parsing can make the grammatical basis of semantic relationships and compositionality more explicit. One of the biggest problems in syntactic parsing is that statistical models tend to be overspecialized for the genres they are trained on: even a highly accurate parser trained on well-edited *Wall Street Journal* articles from 1989 will be flummoxed when presented with the styles of English found on social media. To adapt parsers to new domains, we need hand-annotated corpora for those domains. One line of work in this area has been the development of corpora and tools for part-of-speech tagging and dependency parsing of Twitter, a particularly challenging genre (Gimpel et al., 2011; Owoputi et al., 2013; **Schneider** et al., 2013; Kong et al., 2014; Liu et al., 2018). In another project (with collaborators at Edinburgh and HUJI; journal submission in preparation), I contributed to a corpus of child-directed utterances in the Universal Dependencies (UD) framework. This project demonstrates that the syntactic annotations can be automatically converted to semantic representations conducive to computational modeling of child language acquisition. In connection with this effort and the STREUSLE corpus (§1), I help maintain the UD English corpora and guidelines, frequently contributing corrections and clarifications to this important community resource—and in so doing, uncover inconsistencies requiring adjustments to the theory (**Schneider** and Zeldes, 2021).

Also in the realm of syntax, I am particularly excited about work concerning the *long tail* of syntactic behavior. My advisee Jakob Prange recently led a project that advanced the state of the art for constructive supertagging in the Combinatory Categorial Grammar (CCG) framework. CCG supertags[6] are complex lexicosyntactic labels, internally structured as trees. Many of the complex supertags are extremely rare, and a standard approach until recently was to give up on them and only model the ones seen a certain number of times in the training data. We show that explicitly tree-structured neural models are able to generalize better to rare and unseen supertags by predicting the component pieces (Prange et al., 2021). In ongoing work, we are looking at ways to improve the conditioning in the model and to additionally predict coindexation needed for semantic analysis.

Using Jakob's tagger, I worked with another one of my advisees, Michael Kranzlein, to study modeling of confidence calibration in the long tail (Kranzlein et al., 2021). Confidence calibration refers to the reliability of probability estimates produced by a model (i.e., whether the model is usually correct when assigns high probability to an output). In general, good calibration is desirable for a machine learning system because it means the probabilities can gives users insights into the system's behavior. However, previous studies of calibration have not considered tasks with large numbers of rare tags. We have developed a technique that can be used to evaluate calibration both for high-frequency and low-frequency tags, and to adjust the model's probabilities to be better calibrated across the frequency spectrum. Our experiments indicate the technique is successful for evaluations on in-domain data, and I am looking forward to evaluating out-of-domain calibration as well.

4 Outlook

Even as efforts in disambiguation, tagging, and parsing continue, I have in mind the following future directions:

Toward richer models of meaning. A well-worn narrative articulates the goal of computational semantics as freeing a pure form of *information* from its messy presentation in language. This can be a useful way to think about some practical NLP problems, but it fails to appreciate the richness of meaning conveyed through linguistic communication. In (Trott et al., 2020) we challenged the field to enagage with a richer notion of meaning that emerges from a dynamic process of **construal**, such that subtle differences in linguistic form may give rise to differences in understanding (even with equivalent information content). We summarized several dimensions by which linguistic choices help a speaker paint a meaning for a listener to conceptualize, above and beyond simply transferring information or characterizing conditions of truth. For example: as my coauthors and I argued in (Rohde et al., 2018), there are passages where multiple semantic relations simultaneously hold between two elements, but a connective word makes explicit only one of the relations.

[6] Not to be confused with supersense tags.

This may invite a construal in which the explicit relation is more salient than the implicit relation. I believe that this expanded view of meaning is urgently relevant in the current climate of NLU evaluation, and that it opens up a research program of empirically investigating the aforementioned construal dimensions.

Toward neuro-symbolic language models. There is further opportunity to leverage existing symbolic representations of structure and meaning (such as the syntactic and semantic graph frameworks discussed above). In recent work we showed that the linguistic graphs can be integrated into a neural language model alongside parameters from large-scale pretraining, with promising results (Prange et al., 2022).

Better NLP algorithms for corpus linguists. I am also eager to consider NLU application settings where the nature of the language itself—and not merely the information it encodes—is a matter of interest. One of these naturally arises in **corpus linguistics**: corpus studies of a non-superficial linguistic phenomenon, such as syntactic or semantic structure, could benefit from better algorithms. For example, well-calibrated confidence scores could help linguists to compensate for parser errors when retrieving examples from a corpus. As another example, I am worked with Ph.D. student Luke Gessler to investigate whether language models such as BERT can be used to retrieve, without supervision, rare senses of frequent words (e.g., the sense of "in" exhibited in "We have an excellent recruit **in** Beetle Bailey"—quite distinct from its vastly more frequent senses!). The results were strong, suggesting that we are opening the door to a new generation of semantic search tools for linguists (Gessler and **Schneider**, 2021).

NLP for legal language. Another area where the details of language matter is in the law, where linguistic interpretation is crucial, and court cases document judicial arguments about contested linguistic interpretations (of clauses in the U.S. Constitution, for example). Can we detect elements of judicial linguistic interpretation at scale? Can NLU technology be developed to assist humans drafting or interpreting legal language, e.g., by pointing out ambiguous wording or surfacing useful evidence from corpora? These are topics of ongoing work.

References

Omri Abend, Dotan Dvir, Daniel Hershcovich, Jakob Prange, and **Nathan Schneider**. 2020a. Cross-lingual semantic representation for NLP with UCCA. In *Proc. of the 28th International Conference on Computational Linguistics: Tutorial Abstracts*, pages 1–9, Barcelona, Spain (Online).

Omri Abend and Ari Rappoport. 2013. Universal Conceptual Cognitive Annotation (UCCA). In *Proc. of ACL*, pages 228–238, Sofia, Bulgaria.

Omri Abend, **Nathan Schneider**, Dotan Dvir, Jakob Prange, and Ari Rappoport. 2020b. UCCA's Foundational Layer: Annotation guidelines v2.1. *arXiv:2012.15810 [cs]*. arXiv: 2012.15810.

Aryaman Arora, Nitin Venkateswaran, and **Nathan Schneider**. 2022. MASALA: Modelling and analysing the semantics of adpositions in linguistic annotation of Hindi. In *Proc. of LREC*, Marseille, France.

Collin Baker, Michael Ellsworth, and Katrin Erk. 2007. SemEval-2007 Task 19: frame semantic structure extraction. In *Proc. of SemEval*, pages 99–104, Prague, Czech Republic.

Collin Baker, **Nathan Schneider**, Miriam R. L. Petruck, and Michael Ellsworth. 2015. Getting the roles right: using FrameNet in NLP. In *Proc. of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorial Abstracts*, pages 10–12, Denver, Colorado, USA.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and **Nathan Schneider**. 2013. Abstract Meaning Representation for sembanking. In *Proc. of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and **Nathan Schneider**. 2015. Abstract Meaning Representation (AMR) 1.2.2 Specification.

Austin Blodgett and **Nathan Schneider**. 2018. Semantic supersenses for English possessives. In *Proc. of LREC*, pages 1529–1534, Miyazaki, Japan.

Austin Blodgett and **Nathan Schneider**. 2019. An improved approach for semantic graph composition with CCG. In *Proc. of IWCS*, pages 55–70, Gothenburg, Sweden.

Austin Blodgett and **Nathan Schneider**. 2021. Probabilistic, structure-aware algorithms for improved variety, accuracy, and coverage of AMR alignments. In *Proc. of ACL-IJCNLP*, pages 3310–3321, Online.

Claire Bonial, Bianca Badarau, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Tim O'Gorman, Martha Palmer, and **Nathan Schneider**. 2018. Abstract Meaning Representation of constructions: The more we include, the better the representation. In *Proc. of LREC*, pages 1677–1684, Miyazaki, Japan.

Dipanjan Das, Desai Chen, André F. T. Martins, **Nathan Schneider**, and Noah A. Smith. 2014. Frame-semantic parsing. *Computational Linguistics*, 40(1):9–56.

Dipanjan Das, **Nathan Schneider**, Desai Chen, and Noah A. Smith. 2010. Probabilistic frame-semantic parsing. In *Proc. of NAACL-HLT*, pages 948–956, Los Angeles, California.

Lucia Donatelli, Michael Regan, William Croft, and **Nathan Schneider**. 2018. Annotation of tense and aspect semantics for sentential AMR. In *Proc. of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 96–108, Santa Fe, New Mexico, USA.

Charles J. Fillmore and Collin Baker. 2009. A frames approach to semantic analysis. In Bernd Heine and Heiko Narrog, editors, *The Oxford Handbook of Linguistic Analysis*, pages 791–816. Oxford University Press, Oxford, UK.

Luke Gessler, Austin Blodgett, Joseph Ledford, and **Nathan Schneider**. 2022. Xposition: An online multilingual database of adpositional semantics. In *Proc. of LREC*, Marseille, France.

Luke Gessler and **Nathan Schneider**. 2021. BERT has uncommon sense: similarity ranking for word sense BERTology. In *Proc. of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP*, pages 539–547, Online and Punta Cana, Dominican Republic.

Luke Gessler, Shira Wein, and **Nathan Schneider**. 2020. Supersense and sensibility: proxy tasks for semantic annotation of prepositions. In *Proc. of the 14th Linguistic Annotation Workshop*, pages 117–126, Barcelona, Spain (Online).

Kevin Gimpel, **Nathan Schneider**, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A. Smith. 2011. Part-of-speech tagging for Twitter: annotation, features, and experiments. In *Proc. of ACL-HLT*, pages 42–47, Portland, Oregon, USA.

Daniel Hershcovich, **Nathan Schneider**, Dotan Dvir, Jakob Prange, Miryam de Lhoneux, and Omri Abend. 2020. Comparison by conversion: reverse-engineering UCCA from syntax and lexical semantics. In *Proc. of COLING*, pages 2947–2966, Barcelona, Spain (Online).

Jena D. Hwang, Hanwool Choe, Na-Rae Han, and **Nathan Schneider**. 2020. K-SNACS: Annotating Korean adposition semantics. In *Proc. of the Second International Workshop on Designing Meaning Representations*, pages 53–66, Barcelona, Spain (Online).

Taelin Karidi, Yichu Zhou, **Nathan Schneider**, Omri Abend, and Vivek Srikumar. 2021. Putting words in BERT's mouth: navigating contextualized vector spaces with pseudowords. In *Proc. of EMNLP*, pages 10300–10313, Online and Punta Cana, Dominican Republic.

Lingpeng Kong, **Nathan Schneider**, Swabha Swayamdipta, Archna Bhatia, Chris Dyer, and Noah A. Smith. 2014. A dependency parser for tweets. In *Proc. of EMNLP*, pages 1001–1012, Doha, Qatar.

Michael Kranzlein, Nelson F. Liu, and **Nathan Schneider**. 2021. Making heads *and* tails of models with marginal calibration for sparse tagsets. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 4919–4928, Online and Punta Cana, Dominican Republic.

Michael Kranzlein, Emma Manning, Siyao Peng, Shira Wein, Aryaman Arora, and **Nathan Schneider**. 2020. PASTRIE: a corpus of prepositions annotated with supersense tags in Reddit international English. In *Proc. of the 14th Linguistic Annotation Workshop*, pages 105–116, Barcelona, Spain (Online).

Meghana Kshirsagar, Sam Thomson, **Nathan Schneider**, Jaime Carbonell, Noah A. Smith, and Chris Dyer. 2015. Frame-semantic role labeling with heterogeneous annotations. In *Proc. of ACL-IJCNLP*, pages 218–224, Beijing, China.

Nelson F. Liu, Daniel Hershcovich, Michael Kranzlein, and **Nathan Schneider**. 2021. Lexical semantic recognition. In *Proc. of the 17th Workshop on Multiword Expressions (MWE 2021)*, pages 49–56, Online.

Yijia Liu, Yi Zhu, Wanxiang Che, Bing Qin, **Nathan Schneider**, and Noah A. Smith. 2018. Parsing tweets into Universal Dependencies. In *Proc. of NAACL-HLT*, pages 965–975, New Orleans, Louisiana.

Emma Manning and **Nathan Schneider**. 2021. Referenceless parsing-based evaluation of AMR-to-English generation. In *Proc. of the 2nd Workshop on Evaluation and Comparison of NLP Systems*, pages 114–122, Online and Punta Cana, Dominican Republic.

Emma Manning, Shira Wein, and **Nathan Schneider**. 2020. A human evaluation of AMR-to-English generation systems. In *Proc. of COLING*, pages 4773–4786, Barcelona, Spain (Online).

Tahira Naseem, Austin Blodgett, Sadhana Kumaravel, Tim O'Gorman, Young-Suk Lee, Jeffrey Flanigan, Ramón Fernandez Astudillo, Radu Florian, Salim Roukos, and **Nathan Schneider**. 2022. DocAMR: Multi-sentence AMR representation and evaluation. In *Proc. of NAACL-HLT*, Seattle, Washington.

Stephan Oepen, Omri Abend, Lasha Abzianidze, Johan Bos, Jan Hajic, Daniel Hershcovich, Bin Li, Tim O'Gorman, Nianwen Xue, and Daniel Zeman. 2020. MRP 2020: The second shared task on cross-framework and cross-lingual meaning representation parsing. In *Proc. of CoNLL*, pages 1–22, Online.

Stephan Oepen, Omri Abend, Jan Hajic, Daniel Hershcovich, Marco Kuhlmann, Tim O'Gorman, Nianwen Xue, Jayeol Chun, Milan Straka, and Zdenka Uresova. 2019. MRP 2019: Cross-Framework Meaning Representation Parsing. In *Proc. of CoNLL*, pages 1–27, Hong Kong.

Olutobi Owoputi, Brendan O'Connor, Chris Dyer, Kevin Gimpel, **Nathan Schneider**, and Noah A. Smith. 2013. Improved part-of-speech tagging for online conversational text with word clusters. In *Proc. of NAACL-HLT*, pages 380–390, Atlanta, Georgia, USA.

Siyao Peng, Yang Liu, Yilun Zhu, Austin Blodgett, Yushi Zhao, and **Nathan Schneider**. 2020. A corpus of adpositional supersenses for Mandarin Chinese. In *Proc. of LREC*, pages 5988–5996, Marseille, France.

Jakob Prange and **Nathan Schneider**. 2021. Draw *mir* a sheep: a supersense-based analysis of German case and adposition semantics. *KI - Künstliche Intelligenz*, 35(3):291–306.

Jakob Prange, **Nathan Schneider**, and Omri Abend. 2019. Made for each other: Broad-coverage semantic structures meet preposition supersenses. In *Proc. of CoNLL*, pages 174–185, Hong Kong, China.

Jakob Prange, **Nathan Schneider**, and Lingpeng Kong. 2022. Linguistic frameworks go toe-to-toe at neuro-symbolic language modeling. In *Proc. of NAACL-HLT*, Seattle, Washington.

Jakob Prange, **Nathan Schneider**, and Vivek Srikumar. 2021. Supertagging the long tail with tree-structured decoding of complex categories. *Transactions of the Association for Computational Linguistics*, 9:243–260.

Lizhen Qu, Gabriela Ferraro, Liyuan Zhou, Weiwei Hou, **Nathan Schneider**, and Timothy Baldwin. 2015. Big data small data, in domain out-of domain, known word unknown word: the impact of word representations on sequence labelling tasks. In *Proc. of CoNLL*, pages 83–93, Beijing, China.

Hannah Rohde, Alexander Johnson, **Nathan Schneider**, and Bonnie Webber. 2018. Discourse coherence: concurrent explicit and implicit relations. In *Proc. of ACL*, pages 2257–2267, Melbourne, Australia.

Elior Sulem, Omri Abend, and Ari Rappoport. 2015. Conceptual annotations preserve structure across translations: a French-English case study. In *Proc. of the 1st Workshop on Semantics-Driven Statistical Machine Translation (S2MT 2015)*, pages 11–22, Beijing, China.

Ida Szubert, Adam Lopez, and **Nathan Schneider**. 2018. A structured syntax-semantics interface for English-AMR alignment. In *Proc. of NAACL-HLT*, pages 1169–1180, New Orleans, Louisiana.

Nathan Schneider, Emily Danchik, Chris Dyer, and Noah A. Smith. 2014a. Discriminative lexical semantic segmentation with gaps: running the MWE gamut. *Transactions of the Association for Computational Linguistics*, 2:193–206.

Nathan Schneider, Jeffrey Flanigan, and Tim O'Gorman. 2015. The logic of AMR: practical, unified, graph-based sentence semantics for NLP. In *Proc. of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorial Abstracts*, pages 4–5, Denver, Colorado, USA.

Nathan Schneider, Dirk Hovy, Anders Johannsen, and Marine Carpuat. 2016. SemEval-2016 Task 10: Detecting Minimal Semantic Units and their Meanings (DiMSUM). In *Proc. of SemEval*, pages 546–559, San Diego, California, USA.

Nathan Schneider, Jena D. Hwang, Archna Bhatia, Vivek Srikumar, Na-Rae Han, Tim O'Gorman, Sarah R. Moeller, Omri Abend, Adi Shalev, Austin Blodgett, and Jakob Prange. 2020. Adposition and Case Supersenses v2.5: Guidelines for English. *arXiv:1704.02134 [cs]*. arXiv: 1704.02134.

Nathan Schneider, Jena D. Hwang, Vivek Srikumar, Jakob Prange, Austin Blodgett, Sarah R. Moeller, Aviram Stern, Adi Bitan, and Omri Abend. 2018. Comprehensive supersense disambiguation of English prepositions and possessives. In *Proc. of ACL*, pages 185–196, Melbourne, Australia.

Nathan Schneider, Behrang Mohit, Kemal Oflazer, and Noah A. Smith. 2012. Coarse lexical semantic annotation with supersenses: an Arabic case study. In *Proc. of ACL*, pages 253–258, Jeju Island, Korea.

Nathan Schneider, Brendan O'Connor, Naomi Saphra, David Bamman, Manaal Faruqui, Noah A. Smith, Chris Dyer, and Jason Baldridge. 2013. A framework for (under)specifying dependency syntax without overloading annotators. In *Proc. of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 51–60, Sofia, Bulgaria.

Nathan Schneider, Spencer Onuffer, Nora Kazour, Emily Danchik, Michael T. Mordowanec, Henrietta Conrad, and Noah A. Smith. 2014b. Comprehensive annotation of multiword expressions in a social web corpus. In *Proc. of LREC*, pages 455–461, Reykjavík, Iceland.

Nathan Schneider and Noah A. Smith. 2015. A corpus and model integrating multiword expressions and supersenses. In *Proc. of NAACL-HLT*, pages 1537–1547, Denver, Colorado.

Nathan Schneider and Amir Zeldes. 2021. Mischievous nominal constructions in Universal Dependencies. In *Proc. of the Fifth Workshop on Universal Dependencies (UDW, SyntaxFest 2021)*, pages 160–172, Sofia, Bulgaria.

Sean Trott, Tiago Timponi Torrent, Nancy Chang, and **Nathan Schneider**. 2020. (Re)construing meaning in NLP. In *Proc. of ACL*, pages 5170–5184, Online.

Abigail Walsh, Claire Bonial, Kristina Geeraert, John P. McCrae, **Nathan Schneider**, and Clarissa Somers. 2018. Constructing an annotated corpus of verbal MWEs for English. In *Proc. of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 193–200, Santa Fe, New Mexico, USA.

Shira Wein and **Nathan Schneider**. 2021. Classifying divergences in cross-lingual AMR pairs. In *Proc. of The Joint 15th Linguistic Annotation Workshop (LAW) and 3rd Designing Meaning Representations (DMR) Workshop*, pages 56–65, Online and Punta Cana, Dominican Republic.

Nianwen Xue, Johan Bos, William Croft, Jan Hajič, Chu-Ren Huang, Stephan Oepen, Martha Palmer, and James Pustejovsky, editors. 2020. *Proceedings of the Second International Workshop on Designing Meaning Representations*. Association for Computational Linguistics, Barcelona, Spain (Online).

Nianwen Xue, William Croft, Jan Hajic, Chu-Ren Huang, Stephan Oepen, Martha Palmer, and James Pustejovksy, editors. 2019. *Proceedings of the First International Workshop on Designing Meaning Representations*. Association for Computational Linguistics, Florence, Italy.