Towards a Catalogue of Linguistic Graph Banks

Marco Kuhlmann*
Linköping University

Stephan Oepen**
University of Oslo

Graphs exceeding the formal complexity of rooted trees are of growing relevance to much NLP research. Although formally well understood in graph theory, there is substantial variation in the types of linguistic graphs, as well as in the interpretation of various structural properties. To provide a common terminology and transparent statistics across different collections of graphs in NLP, we propose to establish a shared community resource with an open-source reference implementation for common statistics.

1. Motivation

The predominant target representations in natural language parsing traditionally have been *trees*, in the formal sense that every node is reachable from a distinguished root node by exactly one directed path. With a gradual shift of emphasis from more surface-oriented, morpho-syntactic target representations in parsing towards "deeper," more semantic analyses, there is increasing interest in processing structures where characteristic properties of trees like the unique root, connectedness, or lack of reentrancies can be relaxed. Some recent parsing work targets graph-structured representations more general than trees (Sagae and Tsujii 2008, Das et al. 2010, Jones, Goldwater, and Johnson 2013, Flanigan et al. 2014, Martins and Almeida 2014; among others). This development is made possible by ongoing efforts to annotate deeper syntactico-semantic analyses at scale, and typically such annotations either directly take the form of directed graph structures, or can be interpreted as such under moderate transformations.

In computational linguistics and in particular in natural language parsing, however, there is less of an established tradition of using general graphs than in, say, theoretical computer science (although the central role of feature structures in unification-based grammar formalisms arguably marks an exception to this claim). Thus, we note a lack of consensus on which specific structural properties of graphs are most relevant

* Department of Computer and Information Science, Linköping University, 581 83 Linköping, Sweden. E-mail: marco.kuhlmann@liu.se.
** Department of Informatics, University of Oslo, Boks 1080 Blindern, 0316 Oslo, Norway. E-mail: oe@ifi.uio.no.

Submission received: September 29, 2015; revised submission received: March 7, 2016; accepted for publication: April 27, 2016.

doi:10.1162/COLI_a_00268

© 2017 Association for Computational Linguistics

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/COLI_a_00268 by guest on 29 March 2024

in terms of linguistic adequacy or formal effects on models and algorithms. As has been the case for various subclasses of mildly non-projective dependency trees, for example, we expect that the design of parsing algorithms for graph-structured target representations will benefit from the algebraic study of relevant graph subclasses. In this work, we seek to initiate a community process of systematizing the landscape of graph representations of linguistic structure, with particular emphasis on syntactico-semantic analysis. We present a "pilot" study over a selective sample of extant collections of linguistic graphs (Section 2), propose an initial inventory of formally well-defined properties (Section 3), and demonstrate how contrastive statistics over graph banks can contribute to improved understanding of different frameworks (Section 4). Finally we present a proposal for community follow-up action—which we hope may elicit more in-depth discussion of formal and linguistic differences across graph banks (Section 5).

2. A Menagerie of Graph Banks

For this study, we consider four larger graph banks that are generally available (through the Linguistic Data Consortium) and have already been applied in training and evaluation of data-driven parsers. To capture relevant variation, this selection represents different (and arguably increasing) levels of abstraction over the surface signal and its syntactic structure, viz. (a) Combinatory Categorial Grammar word–word dependencies (CCD); (b) Semantic Dependency Parsing targets from SemEval 2014 and 2015 (SDP); (c) the Elementary Dependency Structures (EDS) of Oepen and Lønning (2006); and (d) Abstract Meaning Representation (AMR; Banarescu et al. 2013). Additional candidate graph banks for inclusion in a community-maintained on-line catalogue are, for example, the Groningen Meaning Bank (GMB; Basile et al. 2012), Universal Conceptual Cognitive Annotation (UCCA; Abend and Rappoport 2013), as well as combinations of layers of annotations from the Penn Treebank (PTB; Marcus, Santorini, and Marcinkiewicz 1993) and OntoNotes (Hovy et al. 2006) ecosystems. Also, recent work on "deeper" syntax (Ballesteros et al. 2015) and the Universal Dependencies initiative (de Marneffe et al. 2014) push towards increasing use of non-tree structures.

There are multiple linguistic and formal differences between these resources. Most importantly, CCD and SDP represent **bilexical dependencies**, where graph nodes correspond to surface lexical units (words or tokens). In contrast, EDS and AMR take the form of **semantic networks** (or conceptual graphs), where nodes represent concepts and there need not be an explicit mapping to surface linguistic forms. In Section 3, we discuss some of the ramifications of this fundamental contrast for the analysis of semantically vacuous surface elements and other formal graph properties.

CCG Dependencies (CCD). Hockenmaier and Steedman (2007) construct CCGbank from a combination of careful interpretation of the syntactic annotations in the PTB with additional, manually curated lexical and constructional knowledge. In CCGbank (LDC2005T13), the strings of the venerable PTB *Wall Street Journal* (WSJ) corpus are annotated with pairs of (a) CCG syntactic derivations and (b) sets of semantic bilexical dependency triples, which we term CCD. The latter "include most semantically relevant non-anaphoric local and long-range dependencies" and are suggested by the CCGbank creators as a proxy for predicate–argument structure. Although CCD has mainly been used for contrastive parser evaluation (Clark and Curran 2007, Fowler and Penn 2010; among others), there is current work that views each set of triples

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/COLI_a_00268 by guest on 29 March 2024

as a directed graph and parses directly into these target representations (Du, Sun, and Wan 2015).

SDP 2014 and 2015: DM and PSD. For the SDP tasks at SemEval, Oepen et al. (2014, 2015) prepared aligned sets of semantic dependency graphs over the same WSJ text by reduction (i.e., lossy conversion) of independently developed syntactico-semantic treebanks into bilexical semantic dependencies. SDP (LDC2016T10) comprises multiple linguistic frameworks, but for our pilot comparison we focus on two sets of target representations that are *not derivative* of the PTB, viz. (a) DELPH-IN MRS-Derived Dependencies (DM; Oepen and Lønning 2006, Ivanova et al. 2012) and (b) Prague Semantic Dependencies (PSD; Hajič et al. 2012, Miyao, Oepen, and Zeman 2014). Both are rooted in general theories of grammar—Head-Driven Phrase Structure Grammar (Pollard and Sag 1994) and Prague Functional Generative Description (FGD; Sgall, Hajičová, and Panevová 1986), respectively—and there are numerous current reports on parsing into these target representations.

Elementary Dependency Structures (EDS). The DM bilexical dependencies originally derive from the underspecified logical forms of Copestake et al. (2005), which Oepen and Lønning (2006), by elimination of scope constraints, reduced to variable-free, unordered semantic dependency graphs called EDS (also included in LDC2016T10). These graphs are formally—if not linguistically—equivalent to AMR (see the next description). Nodes in EDS are independent of surface lexical units, but for each node there is an explicit, many-to-one mapping onto sub-strings of the underlying linguistic signal. Thus, we include EDS as a middle ground between the node-ordered lexicalized dependency graphs of CCD and SDP and the unordered AMR graphs, which provide no overt links to the surface signal.

Abstract Meaning Representation (AMR). Unlike the bilexical dependency graphs of CCD, DM, and PSD, AMR eschews explicit syntactic derivations and consideration of the syntax–semantics interface; it rather seeks to directly annotate "whole-sentence logical meanings" (Banarescu et al. 2013). Node labels in AMR name abstract concepts, which in large part draw on the ontology of OntoNotes predicate senses and corresponding semantic roles. Nodes are not overtly related to surface lexical units, and thus are unordered. Although AMR has its roots in semantic networks and earlier knowledge representation approaches (Langkilde and Knight 1998), larger-scale manual AMR annotation is a recent development only. We sample two variants of AMR, viz. (a) the graphs as annotated in AMRBank 1.0 (LDC2014T12), and (b) a normalized version that we call AMR^{-1}, where so-called "inverse roles" (like ARG0-of) are reversed. Such inverted edges are frequently used in AMR in order to render the graph as a single rooted structure, where the root is interpreted as the top-level focus.[1] In Section 3, we map this interpretation to our concept of *top* nodes for both AMR and AMR^{-1}. Flanigan et al. (2014) published the first parser targeting AMR, and the state of the art has been repeatedly updated since.

1 The graph bank is natively constructed and released with inverted edges, but for parser evaluation the AMR^{-1} normalization is typically assumed; our conversion builds on the code of Cai and Knight (2013).

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/COLI_a_00268 by guest on 29 March 2024

3. Graph Properties and Statistics

To help understand the similarities and differences in our sample of graph banks, in this section we propose an initial inventory of formally well-defined graph properties and calculate contrastive statistics; these are given in Table 1. For all resources, our statistics are computed for the designated training segments (e.g., Sections 02 through 21 for the PTB-derived CCGbank).

The structures in our graph banks can all be viewed as **directed graphs** or **digraphs**. A digraph is a pair $G = (V, E)$ where V is a set of **nodes** and $E \subseteq V \times V$ is a set of **edges**. The number of graphs and their average token counts (following PTB conventions) and node counts are given in rows (01) to (03) in the top part of Table 1. A higher proportion of nodes per token in EDS reflects its frequent use of lexical decomposition, for example, in nominalizations, compounding, and comparatives. In all representations, both nodes and edges are *labeled* with various data, such as lemmata, parts of speech, or predicates, and semantic roles, respectively. The number of labels varies greatly; counts for edge labels are given in row (04).

Singletons. CCD, DM, and PSD maintain technical compatibility with a strong tradition in syntactic dependency parsing: Tokens of the surface string correspond one-to-one to

Table 1
Sample statistics for each full graph bank (top) and for the common "control" subset (bottom); some metrics are average percentages over full graphs ($\%_g$), others over individual nodes ($\%_n$).

			CCD	DM	PSD	EDS	AMR	AMR^{-1}
COUNTS	(01)	number of graphs	39604	35656	35656	35656	10309	10309
	(02)	average number of tokens	23.47	22.51	22.51	22.51	20.62	20.62
	(03)	average number of nodes per token	0.88	0.77	0.64	0.99	0.67	0.67
	(04)	number of edge labels	6	59	90	10	135	100
TREENESS	(05)	$\%_g$ trees	1.45	2.31	42.26	0.98	52.48	18.60
	(06)	$\%_g$ treewidth one	29.27	69.82	43.08	65.37	52.72	52.72
	(07)	average treewidth	1.742	1.303	1.614	1.352	1.524	1.524
	(08)	maximal treewidth	5	3	7	3	4	4
	(09)	average edge density	1.070	1.019	1.073	1.047	1.065	1.065
	(10)	$\%_n$ reentrant	28.09	27.43	11.41	28.42	5.23	18.95
	(11)	$\%_g$ cyclic	1.28	0.00	0.00	0.04	3.15	0.71
	(12)	$\%_g$ not connected	12.53	6.57	0.70	1.49	0.00	0.00
	(13)	$\%_g$ multi-rooted	99.67	99.49	99.33	98.75	0.00	77.50
	(14)	percentage of non-top roots	47.78	44.94	4.34	41.15	0.00	19.39
ORDER	(15)	average edge length	2.582	2.684	3.320	–	–	–
	(16)	$\%_g$ noncrossing	48.23	69.21	64.61	–	–	–
	(17)	$\%_g$ pagenumber two	98.64	99.55	98.07	–	–	–
CONTROL	(01)	number of graphs	87	87	87	87	87	87
	(03)	average number of nodes per token	0.88	0.79	0.64	1.01	0.66	0.66
	(05)	$\%_g$ trees	1.15	1.15	45.98	1.15	60.92	3.45
	(06)	$\%_g$ treewidth one	37.93	81.61	47.13	81.61	60.92	60.92
	(07)	average treewidth	1.644	1.184	1.540	1.184	1.402	1.402
	(09)	average edge density	1.057	1.011	1.061	1.028	1.038	1.038
	(10)	$\%_n$ reentrant	28.92	27.73	10.28	27.77	2.88	21.09
	(11)	$\%_g$ cyclic	0.00	0.00	0.00	0.00	2.30	0.00
	(12)	$\%_g$ not connected	6.90	3.45	1.15	1.15	0.00	0.00
	(13)	$\%_g$ multi-rooted	100.00	100.00	100.00	98.85	0.00	93.10

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/COLI_a_00268 by guest on 29 March 2024

the nodes of the graph representing its syntactico-semantic analysis. For semantically vacuous surface elements, these graphs include nodes that are (a) isolated in the structure (with in- and out-degree zero) and (b) not designated as top nodes (see below).[2] Such nodes—called *singletons*—have no significance for meaning representation and are excluded from all graph statistics, for increased comparability, except in row (02).

Treeness. A digraph G is called a (rooted) **tree** if there exists a node r, the **root**, such that every node of G is reachable from r via a unique directed path. Although trees make up the minority of the structures in our sample of graph banks, their exact proportion varies greatly: from 0.98% in EDS to 52.48% in AMR (row 05). This percentage decreases to 18.60% for $\texttt{AMR}^{-1}$, where normalizing the inverted edges creates a significant number of reentrancies. The second-highest proportion of trees (42.26%) is observed in PSD, which here appears to show its origins in the underlying FGD tectogrammatical trees, where synthetic nodes and explicit identity edges serve to encode argument sharing across predicates (Miyao, Oepen, and Zeman 2014).

Treewidth. Intuitively, even a graph that is not a tree may be more or less "like" a tree. One well-known measure that can be used to quantify the "treeness" of a graph is its **treewidth** (Diestel 2005); trees are graphs with treewidth 1. Treewidth is relevant because it is a complexity parameter in some of the current AMR parsing algorithms (Chiang et al. 2013). Graphs with treewidth one cover between 29.27% (CCD) and 69.82% (DM) of the instances in the five data sets (row 06), and the average treewidth varies from 1.303 in the DM data to 1.742 in CCD (row 07). The relatively high treewidth in the PSD data (1.614) is interesting in light of the fact that this data set, at the same time, has the second-highest percentage of trees. PSD also has the highest maximal treewidth (row 08). Note that treewidth, as a measure defined on undirected graphs, is the same for the two AMR variants.

Edge Density. Another way to quantify the treeness of a (loop-free) digraph $G = (V, E)$ is to measure its **edge density**, the number of edges per node. More formally, we define the edge density as $|E|/(|V| - 1)$ if $|V| > 1$, and 1 otherwise. Because a tree on $|V|$ nodes has exactly $|V| - 1$ edges, trees have edge density 1. The average edge density of all five data sets is very close to this number (row 09): The smallest value (1.019) is observed for DM graphs, the highest (1.073) for PSD graphs.

Reentrancies. In a tree, every node except the root has in-degree 1. In our sample of graph banks, between 5% and 28% of the (non-singleton) nodes have in-degree 2 or greater (row 10). The lowest percentage is observed in the AMR data; the highest percentage in the EDS data.

Acyclicity. In contrast to trees, general digraphs may contain **cycles**. However, in the preparation of the SDP data, cycles have been explicitly ruled out (Oepen et al. [2015] report a proportion of 0.39% cyclic graphs in the raw data underlying the DM and PSD graphs). Cycles are relatively rare even in CCD (1.28%). Their percentage is highest for non-normalized AMR (3.15%), but decreases substantially (to 0.71%) with the reversal of edges in the normalized version $\texttt{AMR}^{-1}$ (row 11).

2 Particles, complementizers, and (most) punctuation marks, for example, are conventionally analyzed as not meaning-bearing, though the exact categorization varies substantially across linguistic frameworks.

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/COLI_a_00268 by guest on 29 March 2024

Connectedness. Another central property of trees is that they are **connected**, meaning that there exists an undirected path between any pair of nodes. This property is characteristic for AMR graphs; but graphs in the other collections are not generally connected (row 12), with proportions of non-connected graphs between 0.7% (PSD) and 12.5% (CCD).

Top Nodes. In contrast to the unique root node in trees, graphs can have multiple (structural) roots, which we define as nodes with in-degree zero; with the exception of unnormalized AMR, the majority of graphs are multi-rooted in all our samples (row 13). Thus, all our graph banks distinguish one or several nodes in each graph as *top* nodes; these correspond to the most central semantic entities in the graph, usually the main predicates.[3] For DM, CCD, and AMR, each graph has at most one top node. In PSD, top nodes are derived in a way that can lead to multiple top nodes per sentence in the case of conjunction. Root nodes that are not top occur in all data sets except AMR (row 14), although their proportion varies greatly, from 4% in PSD to 47% in CCD. High proportions of non-top roots in CCD and DM can in part be explained by the treatment of non-scopal modifiers (e.g., most attributive adjectives and adverbs) as semantic predicates.

Order-Related Properties. In the three surface-oriented data sets, the left-to-right order of the tokens in a sentence induces a natural linear order on the nodes. This makes it possible to quantify the **length** of an edge as the distance between the left and the right endpoint. Row (15) shows that the average edge lengths in CCD and DM are comparable (2.582 and 2.684), whereas edges in the PSD data are significantly longer (3.320). This is at least partially related to the analysis of coordinate structures in PSD, where dependencies from the predicate have been propagated to all conjuncts.

A natural way to visualize a bilexical dependency graph is to draw its edges as semicircles in the halfplane above the sentence. A graph is called **noncrossing** if in such a drawing, the semicircles intersect only at their endpoints. This property is a natural generalization of **projectivity** as it is known from dependency trees (Kuhlmann and Nivre 2006), and like projectivity can be exploited to obtain polynomial parsing algorithms (Kuhlmann and Jonsson 2015; Schluter 2015). However, the coverage of the noncrossing property (row 16) is lower than that of projectivity on syntactic data sets: The proportion is largest (69.21%) in the DM data but significantly smaller (48.23%) in CCD. At the same time, a natural generalization of the noncrossing property, where one is allowed to also use the halfplane *below* the sentence for drawing edges, covers more than 98% of all three data sets (row 17); in theoretical computer science, this extended class of graphs is characterized by a property called *pagenumber two*. The statistics suggest that the forms of crossings that are expressed in the data are severely limited.

4. A Control Experiment on Parallel Text

In comparing AMR to the other representations, a skeptic might argue that there are two separate dimensions at play, viz. (a) variation in text types and the phenomena they invoke and (b) actual linguistic differences in the semantic graphs. To tease these apart, we conduct a control experiment on the subset of graphs that all annotate the same basic text, 87 WSJ sentences from the PTB. A selection of our graph statistics over this parallel text is summarized in the bottom part of Table 1. Although it appears that this subset of graphs presents structurally mildly less complex and shorter (at an average length of 22 tokens) inputs, we find all general tendencies from Section 3 and relative ordering

3 In AMR, top nodes are called **roots**, a term that we reserve for the above structural interpretation.

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/COLI_a_00268 by guest on 29 March 2024

among representations confirmed. Thus, we conjecture that these general contrasts primarily reflect contentful linguistic differences. As additional supporting evidence for this assumption, we observe that the statistics are remarkably stable—often to the third decimal—when using only half the available training data.

5. Outlook: A Community Resource

We anticipate a bit of a cottage industry in linguistic graph banks and graph processing tasks over the next few years, which may make it difficult to keep track of contentful similarities and differences across frameworks and approaches. This pilot is intended to initiate the creation and maintenance of an on-line catalogue as a community resource.

To stimulate community engagement, we have (a) copied and expanded parts of this pilot into the ACL wiki, as well as (b) provided an open-source reference implementation of our toolkit for graph statistics.[4] We seek to enable the developers of additional linguistic graph banks to adapt the software to their resources and then contribute statistics and documentation to the catalogue wiki. Our mid- to long-term goal in this effort is three-fold, viz. (a) to contribute to enhanced comparability and replicability; (b) to help identify sub-classes of digraphs for which efficient algorithms can be designed; and (c) to aid the discovery and contrastive discussion of substantive linguistic variation across resources, of the kind indicated speculatively in the examples of Section 3.

Current reported "parsing success" measures range between graph similarity F_1 values in the mid sixties for AMR (Pust et al. 2015), high seventies for PSD (Martins and Almeida 2014), and low nineties for CCD and DM (Du, Sun, and Wan 2015; Miyao, Oepen, and Zeman 2014). Such variation may in principle be owed to diverging evaluation setups, to differences in linguistic "granularity" (i.e., the number and complexity of distinctions made), to the size, homogeneity, and consistency of training and test data, and of course to cumulative effort that has gone into advancing the state of the art on individual tasks. A shared understanding of these parameters in much greater depth will be a prerequisite to judging the relative suitability of different resources and approaches.

References

Abend, Omri and Ari Rappoport. 2013. UCCA. A semantics-based grammatical annotation scheme. In *Proceedings of the 10th International Conference on Computational Semantics*, pages 1–12, Potsdam.

Ballesteros, Miguel, Bernd Bohnet, Simon Mille, and Leo Wanner. 2015. Data-driven deep syntactic dependency parsing. *Natural Language Engineering*, 22:1–36.

Banarescu, Laura, 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 *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia.

Basile, Valerio, Johan Bos, Kilian Evang, and Noortje Venhuizen. 2012. Developing a large semantically annotated corpus. In *Proceedings of the 8th International Conference on Language Resources and Evaluation*, pages 3196–3200, Istanbul.

Cai, Shu and Kevin Knight. 2013. Smatch. An evaluation metric for semantic feature structures. In *Proceedings of the 51th Meeting of the Association for Computational Linguistics*, pages 748–752, Sofia.

Chiang, David, Jacob Andreas, Daniel Bauer, Karl Moritz Hermann, Bevan Keeley Jones, and Kevin Knight. 2013. Parsing graphs with hyperedge replacement grammars. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 924–932, Sofia.

Clark, Stephen and James R. Curran. 2007. Wide-coverage efficient statistical

4 http://www.aclweb.org/aclwiki/index.php?title=Graph_Parsing_(State_of_the_Art).

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/COLI_a_00268 by guest on 29 March 2024

parsing with CCG and log-linear models. *Computational Linguistics*, 33(4):493–552.

Copestake, Ann, Dan Flickinger, Carl Pollard, and Ivan A. Sag. 2005. Minimal Recursion Semantics. An introduction. *Research on Language and Computation*, 3(4):281–332.

Das, Dipanjan, Nathan Schneider, Desai Chen, and Noah A. Smith. 2010. Probabilistic frame-semantic parsing. In *Proceedings of Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 948–956, Los Angeles, CA.

de Marneffe, Marie-Catherine, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre, and Christopher D. Manning. 2014. Universal Stanford dependencies. A cross-linguistic typology. In *Proceedings of the 9th International Conference on Language Resources and Evaluation*, pages 4585–4592, Reykjavik.

Diestel, Reinhard. 2005. *Graph Theory*, volume 173 of *Graduate Texts in Mathematics*. Springer, 3rd edition.

Du, Yantao, Weiwei Sun, and Xiaojun Wan. 2015. A data-driven, factorization parser for CCG dependency structures. In *Proceedings of the 53rd Meeting of the Association for Computational Linguistics and of the 7th International Joint Conference on Natural Language Processing*, pages 1545–1555, Bejing.

Flanigan, Jeffrey, Sam Thomson, Jaime Carbonell, Chris Dyer, and Noah A. Smith. 2014. A discriminative graph-based parser for the Abstract Meaning Representation. In *Proceedings of the 52nd Meeting of the Association for Computational Linguistics*, pages 1426–1436, Baltimore, MD.

Fowler, Timothy A. D. and Gerald Penn. 2010. Accurate context-free parsing with Combinatory Categorial Grammar. In *Proceedings of the 48th Meeting of the Association for Computational Linguistics*, pages 335–344, Uppsala.

Hajič, Jan, Eva Hajičová, Jarmila Panevová, Petr Sgall, Ondřej Bojar, Silvie Cinková, Eva Fučíková, Marie Mikulová, Petr Pajas, Jan Popelka, Jiří Semecký, Jana Šindlerová, Jan Štěpánek, Josef Toman, Zdeňka Urešová, and Zdeněk Žabokrtský. 2012. Announcing Prague Czech-English Dependency Treebank 2.0. In *Proceedings of the 8th International Conference on Language Resources and Evaluation*, pages 3153–3160, Istanbul.

Hockenmaier, Julia and Mark Steedman. 2007. CCGbank. A corpus of CCG derivations and dependency structures extracted from the Penn Treebank. *Computational Linguistics*, 33:355–396.

Hovy, Eduard, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. 2006. OntoNotes. The 90% solution. In *Proceedings of Human Language Technologies: The 2006 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers*, pages 57–60, New York.

Ivanova, Angelina, Stephan Oepen, Lilja Øvrelid, and Dan Flickinger. 2012. Who did what to whom? A contrastive study of syntacto-semantic dependencies. In *Proceedings of the Sixth Linguistic Annotation Workshop*, pages 2–11, Jeju Island.

Jones, Bevan Keeley, Sharon Goldwater, and Mark Johnson. 2013. Modeling graph languages with grammars extracted via tree decompositions. In *Proceedings of the 11th International Conference on Finite State Methods and Natural Language Processing*, pages 54–62, St. Andrews.

Kuhlmann, Marco and Peter Jonsson. 2015. Parsing to noncrossing dependency graphs. *Transactions of the Association for Computational Linguistics*, 3:559–570.

Kuhlmann, Marco and Joakim Nivre. 2006. Mildly non-projective dependency structures. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Meeting of the Association for Computational Linguistics*, pages 507–514, Sydney.

Langkilde, Irene and Kevin Knight. 1998. Generation that exploits corpus-based statistical knowledge. In *Proceedings of the 17th International Conference on Computational Linguistics and the 36th Meeting of the Association for Computational Linguistics*, pages 704–710, Montréal.

Marcus, Mitchell, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpora of English. The Penn Treebank. *Computational Linguistics*, 19:313–330.

Martins, André F. T. and Mariana S. C. Almeida. 2014. Priberam. A turbo semantic parser with second order features. In *Proceedings of the 9th International Workshop on Semantic Evaluation*, pages 471–476, Dublin.

Miyao, Yusuke, Stephan Oepen, and Daniel Zeman. 2014. In-House. An ensemble of pre-existing off-the-shelf parsers. In *Proceedings of the 9th International Workshop on Semantic Evaluation*, pages 63–72, Dublin.

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/COLI_a_00268 by guest on 29 March 2024

Oepen, Stephan, Marco Kuhlmann, Yusuke Miyao, Daniel Zeman, Silvie Cinková, Dan Flickinger, Jan Hajič, and Zdeňka Urešová. 2015. SemEval 2015 Task 18. Broad-coverage semantic dependency parsing. In *Proceedings of the 10th International Workshop on Semantic Evaluation*, pages 915–926, Denver, CO.

Oepen, Stephan, Marco Kuhlmann, Yusuke Miyao, Daniel Zeman, Dan Flickinger, Jan Hajič, Angelina Ivanova, and Yi Zhang. 2014. SemEval 2014 Task 8. Broad-coverage semantic dependency parsing. In *Proceedings of the 9th International Workshop on Semantic Evaluation*, pages 63–72, Dublin.

Oepen, Stephan and Jan Tore Lønning. 2006. Discriminant-based MRS banking. In *Proceedings of the 5th International Conference on Language Resources and Evaluation*, pages 1250–1255, Genoa.

Pollard, Carl and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Studies in Contemporary Linguistics. The University of Chicago Press, Chicago, IL.

Pust, Michael, Ulf Hermjakob, Kevin Knight, Daniel Marcu, and Jonathan May. 2015. Parsing English into abstract meaning representation using syntax-based machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1143–1154, Lisbon.

Sagae, Kenji and Jun'ichi Tsujii. 2008. Shift-reduce dependency DAG parsing. In *Proceedings of the 22nd International Conference on Computational Linguistics*, pages 753–760, Manchester.

Schluter, Natalie. 2015. The complexity of finding the maximum spanning DAG and other restrictions for DAG parsing of natural language. In *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*, pages 259–268, Denver, CO.

Sgall, Petr, Eva Hajičová, and Jarmila Panevová. 1986. *The Meaning of the Sentence and Its Semantic and Pragmatic Aspects*. D. Reidel Publishing Company, Dordrecht, The Netherlands.

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/COLI_a_00268 by guest on 29 March 2024