The State of the Art in Semantic Representation

Omri Abend **Ari Rappoport**
Department of Computer Science, The Hebrew University of Jerusalem
{oabend|arir}@cs.huji.ac.il

Abstract

Semantic representation is receiving growing attention in NLP in the past few years, and many proposals for semantic schemes (e.g., AMR, UCCA, GMB, UDS) have been put forth. Yet, little has been done to assess the achievements and the shortcomings of these new contenders, compare them with syntactic schemes, and clarify the general goals of research on semantic representation. We address these gaps by critically surveying the state of the art in the field.

1 Introduction

Schemes for Semantic Representation of Text (SRT) aim to reflect the meaning of sentences and texts in a transparent way. There has recently been an influx of proposals for semantic representations and corpora, e.g. GMB (Basile et al., 2012), AMR (Banarescu et al., 2013), UCCA (Abend and Rappoport, 2013b) and Universal Decompositional Semantics (UDS; White et al., 2016). Nevertheless, no detailed assessment of the relative merits of the different schemes has been carried out, nor their comparison to previous sentential analysis schemes, notably syntactic ones. An understanding of the achievements and gaps of semantic analysis in NLP is crucial to its future prospects.

In this paper we begin to chart the various proposals for semantic schemes according to the **content** they support. As not many semantic queries on texts can at present be answered with near human-like reliability without using manual symbolic annotation, we will mostly focus on schemes that represent semantic distinctions explicitly.[1]

We begin by discussing the goals of SRT in Section 2. Section 3 surveys major represented meaning components, including predicate-argument relations, discourse relations and logical structure. Section 4 details the various concrete proposals for SRT schemes and annotated resources, while Sections 5 and 6 discuss criteria for their evaluation and their relation to syntax, respectively.

We find that despite the major differences in terms of formalism and interface with syntax, in terms of their content there is a great deal of convergence of SRT schemes. Principal differences between schemes are mostly related to their ability to abstract away from formal and syntactic variation, namely to assign similar structures to different constructions that have a similar meaning, and to assign different structures to constructions that have different meanings, despite their surface similarity. Other important differences are in the level of training they require from their annotators (e.g., expert annotators vs. crowd-sourcing) and in their cross-linguistic generality. We discuss the complementary strengths of different schemes, and suggest paths for future integration.

2 Defining Semantic Representation

The term *semantics* is used differently in different contexts. For the purposes of this paper we define a semantic representation as one that reflects the meaning of the text as it is understood by a language speaker. A semantic representation should thus be paired with a method for extracting information from it that can be directly evaluated by humans. The extraction process should be reliable and computationally efficient.

[1] Note that even a string representation of text can be regarded as semantic given a reliable enough parser.

Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, pages 77–89
Vancouver, Canada, July 30 - August 4, 2017.
https://doi.org/10.18653/v1/P17-1008

We stipulate that a fundamental component of the content conveyed by SRTs is argument structure – who did what to whom, where, when and why, i.e., events, their participants and the relations between them. Indeed, the fundamental status of argument structure has been recognized by essentially all approaches to semantics both in theoretical linguistics (Levin and Hovav, 2005) and in NLP, through approaches such as Semantic Role Labeling (SRL; Gildea and Jurafsky, 2002), formal semantic analysis (e.g., Bos, 2008), and Abstract Meaning Representation (AMR; Banarescu et al., 2013). Many other useful meaning components have been proposed, and are discussed at a greater depth in Section 3.

Another approach to defining an SRT is through external (extra-textual) criteria or applications. For instance, a semantic representation can be defined to support inference, as in textual entailment (Dagan et al., 2006) or natural logic (Angeli and Manning, 2014). Other examples include defining a semantic representation in terms of supporting knowledge base querying (Zelle and Mooney, 1996; Zettlemoyer and Collins, 2005), or defining semantics through a different modality, for instance interpreting text in terms of images that correspond to it (Kiros et al., 2014), or in terms of embodied motor and perceptual schemas (Feldman et al., 2010).

A different approach to SRT is taken by Vector Space Models (VSM), which eschew the use of symbolic structures, instead modeling all linguistic elements as vectors, from the level of words to phrases and sentences. Proponents of this approach generally invoke neural network methods, obtaining impressive results on a variety of tasks including lexical tasks such as cross-linguistic word similarity (Ammar et al., 2016), machine translation (Bahdanau et al., 2015), and dependency parsing (Andor et al., 2016). VSMs are also attractive in being flexible enough to model non-local and gradient phenomena (e.g., Socher et al., 2013). However, more research is needed to clarify the scope of semantic phenomena that such models are able to reliably capture. We therefore only lightly touch on VSMs in this survey.

Finally, a major consideration in semantic analysis, and one of its great potential advantages, is its cross-linguistic universality. While languages differ in terms of their form (e.g., in their phonology, lexicon, and syntax), they have often been assumed to be much closer in terms of their semantic content (Bar-Hillel, 1960; Fodor, 1975). See Section 5 for further discussion.

A terminological note: within formal linguistics, semantics is often the study of the relation between symbols (e.g., words, syntactic constructions) and what they signify. In this sense, semantics is the study of the aspects of meaning that are overtly expressed by the lexicon and grammar of a language, and is thus tightly associated with a theory of the syntax-semantics interface. We note that this definition of semantics is somewhat different from the one intended here, which defines semantic schemes as theories of meaning.

3 Semantic Content

We turn to discussing the main content types encoded by semantic representation schemes. Due to space limitations, we focus only on text semantics, which studies the meaning relationships between lexical items, rather than the meaning of the lexical items themselves.[2] We also defer discussion of more targeted semantic distinctions, such as sentiment, to future work.

We will use the following as a running example:

(1) Although Ann was leaving, she gave the present to John.

Events. Events (sometimes called frames, propositions or scenes) are the basic building blocks of argument structure representations. An event includes a predicate (main relation, frame-evoking element), which is the main determinant of what the event is about. It also includes arguments (participants, core elements) and secondary relations (modifiers, non-core elements). Example 1 is usually viewed as having two events, evoked by "leaving" and "gave".

Schemes commonly provide an ontology or a lexicon of event types (also a predicate lexicon), which categorizes semantically similar events evoked by different lexical items. For instance, FrameNet defines frames as schematized story fragments evoked by a set of conceptually similar predicates. In (1), the frames evoked by "leaving" and "gave" are DEPARTING and GIVING, but DEPARTING may also be evoked by "depart" and "exit", and GIVING by "donate" and "gift".

[2] We use the term "Text Semantics", rather than the commonly used "Sentence Semantics" to include inter-sentence semantic relations as well.

The events discussed here should not be confused with events as defined in Information Extraction and related tasks such as event coreference (Humphreys et al., 1997), which correspond more closely to the everyday notion of an event, such as a political or financial event, and generally consist of multiple events in the sense discussed here. The representation of such events is recently receiving considerable interest within NLP, e.g. the Richer Event Descriptions framework (RED; Ikuta et al., 2014).

Predicates and Arguments. While predicate-argument relations are universally recognized as fundamental to semantic representation, the interpretation of the terms varies across schemes. Most SRL schemes cover a wide variety of verbal predicates, but differ in which nominal and adjectival predicates are covered. For example, PropBank (Palmer et al., 2005), one of the major resources for SRL, covers verbs, and in its recent versions also eventive nouns and multi-argument adjectives. FrameNet (Ruppenhofer et al., 2016) covers all these, but also covers relational nouns that do not evoke an event, such as "president". Other lines of work address semantic arguments that appear outside sentence boundaries, or that do not explicitly appear anywhere in the text (Gerber and Chai, 2010; Roth and Frank, 2015).

Core and Non-core Arguments. Perhaps the most common distinction between argument types is between core and non-core arguments (Dowty, 2003). While it is possible to define the distinction distributionally as one between obligatory and optional arguments, here we focus on the semantic dimension, which distinguishes arguments whose meaning is predicate-specific and are necessary components of the described event (core), and those which are predicate-general (non-core). For example, FrameNet defines core arguments as conceptually necessary components of a frame, that make the frame unique and different from other frames, and peripheral arguments as those that introduce additional, independent or distinct relations from that of the frame such as time, place, manner, means and degree (Ruppenhofer et al., 2016, pp. 23-24).

Semantic Roles. Semantic roles are categories of arguments. Many different semantic role inventories have been proposed and used in NLP over the years, the most prominent being FrameNet (where roles are shared across predicates that evoke the same frame type, such as "leave" and "depart"), and PropBank (where roles are verb-specific). PropBank's role sets were extended by subsequent projects such as AMR. Another prominent semantic role inventory is VerbNet (Kipper et al., 2008) and subsequent projects (Bonial et al., 2011; Schneider et al., 2015), which define a closed set of abstract semantic roles (such as AGENT, PATIENT and INSTRUMENT) that apply to all predicate arguments.

Co-reference and Anaphora. Co-reference allows to abstract away from the different ways to refer to the same entity, and is commonly included in semantic resources. Coreference interacts with argument structure annotation, as in its absence each argument is arbitrarily linked to one of its textual instances. Most SRL schemes would mark "Ann" in (1) as an argument of "leaving" and "she" as an argument of "gave", although on semantic grounds "Ann" is an argument of both.

Some SRTs distinguish between the cases of argument sharing which is encoded by the syntax and is thus explicit (e.g., in "John went home and took a shower", "John" is both an argument of "went home" and of "took a shower"), and cases where the sharing of arguments is inferred (as in (1)). This distinction may be important for text understanding, as the inferred cases tend to be more ambiguous ("she" in (1) might not refer to "Ann"). Other schemes, such as AMR, eschew this distinction and use the same terms to represent all cases of coreference.

Temporal Relations. Most temporal semantic work in NLP has focused on temporal relations between events, either by timestamping them according to time expressions found in the text, or by predicting their relative order in time. Important resources include TimeML, a specification language for temporal relations (Pustejovsky et al., 2003), and the TempEval series of shared tasks and annotated corpora (Verhagen et al., 2009, 2010; UzZaman et al., 2013). A different line of work explores scripts: schematic, temporally ordered sequences of events associated with a certain scenario (Chambers and Jurafsky, 2008, 2009; Regneri et al., 2010). For instance, going to a restaurant includes sitting at a table, ordering, eating and paying, generally in this order.

Related to temporal relations, are causal relations between events, which are ubiquitous in language, and central for a variety of applications, including planning and entailment. See (Mirza et al., 2014) and (Dunietz et al., 2015) for recently proposed annotation schemes for causality and its sub-types. Mostafazadeh et al. (2016) integrated causal and TimeML-style temporal relations into a unified representation.

The internal temporal structure of events has been less frequently tackled. Moens and Steedman (1988) defined an ontology for the temporal components of an event, such as its preparatory process (e.g., "climbing a mountain"), or its culmination ("reaching its top"). Statistical work on this topic is unfortunately scarce, and mostly focuses on lexical categories such as aspectual classes (Siegel and McKeown, 2000; Palmer et al., 2007; Friedrich et al., 2016; White et al., 2016), and tense distinctions (Elson and McKeown, 2010). Still, casting events in terms of their temporal components, characterizing an annotation scheme for doing so and rooting it in theoretical foundations, is an open challenge for NLP.

Spatial Relations. The representation of spatial relations is pivotal in cognitive theories of meaning (e.g., Langacker, 2008), and in application domains such as geographical information systems or robotic navigation. Important tasks in this field include Spatial Role Labeling (Kordjamshidi et al., 2012) and the more recent SpaceEval (Pustejovsky et al., 2015). The tasks include the identification and classification of spatial elements and relations, such as places, paths, directions and motions, and their relative configuration.

Discourse Relations encompass any semantic relation between events or larger semantic units. For example, in (1) the leaving and the giving events are sometimes related through a discourse relation of type CONCESSION, evoked by "although". Such information is useful, often essential for a variety of NLP tasks such as summarization, machine translation and information extraction, but is commonly overlooked in the development of such systems (Webber and Joshi, 2012).

The Penn Discourse Treebank (PeDT; Miltsakaki et al., 2004) annotates discourse units, and classifies the relations between them into a hierarchical, closed category set, including high-level relation types like TEMPORAL, COMPARISON and CONTINGENCY and finer-grained ones such as JUSTIFICATION and EXCEPTION. Another commonly used resource is the RST Discourse Treebank (Carlson et al., 2003), which places more focus on higher-order discourse structures, resulting in deeper hierarchical structures than the PeDT's, which focuses on local discourse structure.

Another discourse information type explored in NLP is discourse segmentation, where texts are partitioned into shallow structures of discourse units categorized either according to their topic or according to their function within the text. An example is the segmentation of scientific papers into functional segments and their labeling with categories such as BACKGROUND and DISCUSSION (Liakata et al., 2010). See (Webber et al., 2011) for a survey of discourse structure in NLP.

Discourse relations beyond the scope of a single sentence are often represented by specialized semantic resources and not by general ones, despite the absence of a clear boundary line between them. This, however, is beginning to change with some schemes, e.g., GMB and UCCA, already supporting cross-sentence semantic relations.[3]

Logical Structure. Logical structure, including quantification, negation, coordination and their associated scope distinctions, is the cornerstone of semantic analysis in much of theoretical linguistics, and has attracted much attention in NLP as well. Common representations are often based on variants of predicate calculus, and are useful for applications that require mapping text into an external, often executable, formal language, such as a querying language (Zelle and Mooney, 1996; Zettlemoyer and Collins, 2005) or robot instructions (Artzi and Zettlemoyer, 2013). Logical structures are also useful for recognizing entailment relations between sentences, as some entailments can be computed from the text's logical structure by formal provers (Bos and Markert, 2005; Lewis and Steedman, 2013).

Inference and Entailment. A primary motivation for many semantic schemes is their ability to support inference and entailment. Indeed, means for predicting logical entailment are built into many forms of semantic representations. A different approach was taken in the tasks of Recognizing Textual Entailment (Dagan et al., 2013), and Natural Logic (van Eijck, 2005), which considers an inference valid if a reasonable annotator would find the hypothesis likely to hold given

[3] AMR will also support discourse structure in its future versions (N. Schneider; personal communication).

the premise, even if it cannot be deduced from it. See (Manning, 2006) for a discussion of this point. Such inference relations are usually not included in semantic treebanks, but annotated in specialized resources (e.g., Dagan et al., 2006; Bowman et al., 2015).

4 Semantic Schemes and Resources

This section briefly surveys the different schemes and resources for SRT. We focus on design principles rather than specific features, as the latter are likely to change as the schemes undergo continuous development. In general, schemes discussed in Section 3 are not repeated here.

Semantic Role Labeling. SRL schemes diverge in their event types, the type of predicates they cover, their granularity, their cross-linguistic applicability, their organizing principles and their relation with syntax. Most SRL schemes define their annotation relative to some syntactic structure, such as parse trees of the PTB in the case of PropBank, or specialized syntactic categories defined for SRL purposes in the case of FrameNet. Other than PropBank, FrameNet and VerbNet discussed above, other notable resources include Semlink (Loper et al., 2007) that links corresponding entries in different resources such as PropBank, FrameNet, VerbNet and WordNet, and the Preposition Supersenses project (Schneider et al., 2015), which focuses on roles evoked by prepositions. See (Palmer et al., 2010, 2013) for a review of SRL schemes and resources. SRL schemes are often termed "shallow semantic analysis" due to their focus on argument structure, leaving out other relations such as discourse events, or how predicates and arguments are internally structured.

AMR. AMR covers predicate-argument relations, including semantic roles (adapted from PropBank) that apply to a wide variety of predicates (including verbal, nominal and adjectival predicates), modifiers, co-reference, named entities and some time expressions.

AMR does not currently support relations above the sentence level, and is admittedly English-centric, which results in an occasional conflation of semantic phenomena that happen to be similarly realized in English, into a single semantic category. AMR thus faces difficulties when assessing the invariance of its structures across translations (Xue et al., 2014). As an example, consider the sentences "I happened to meet Jack in the office", and "I asked to meet Jack in the office". While the two have similar syntactic forms, the first describes a single "meeting" event, where "happened" is a modifier, while the second describes two distinct events: asking and meeting. AMR annotates both in similar terms, which may be suitable for English, where aspectual relations are predominantly expressed as subordinating verbs (e.g., "begin", "want"), and are syntactically similar to primary verbs that take an infinitival complement (such as "ask to meet" or "learn to swim"). However, this approach is less suitable cross-linguistically. For instance, when translating the sentences to German, the divergence between the semantics of the two sentences is clear: in the first "happened" is translated to an adverb: "Ich habe Jack im Büro zufällig getroffen" (lit. "I have Jack in-the office by-chance met"), and in the second "asked" is translated to a verb: "Ich habe gebeten, Jack im Büro zu treffen" (lit. "I have asked, Jack in-the office to meet").

UCCA. UCCA (Universal Conceptual Cognitive Annotation) (Abend and Rappoport, 2013a,b) is a cross-linguistically applicable scheme for semantic annotation, building on typological theory, primarily on Basic Linguistic Theory (Dixon, 2010). UCCA's foundational layer of categories focuses on argument structures of various types and relations between them. In its current state, UCCA is considerably more coarse-grained than the above mentioned schemes (e.g., it does not include semantic role information). However, its distinctions tend to generalize well across languages (Sulem et al., 2015). For example, unlike AMR, it distinguishes between primary and aspectual verbs, so cases such as "happened to meet" are annotated similarly to cases such as "met by chance", and differently from "asked to meet".

Another design principle UCCA evokes is support for annotation by non-experts. To do so the scheme reformulates some of the harder distinctions into more intuitive ones. For instance, the core/non-core distinction is replaced in UCCA with the distinction between pure relations (Adverbials) and those evoking an object (Participants), which has been found easier for annotators to apply.

UDS. Universal Decompositional Semantics (White et al., 2016) is a multi-layered scheme, which currently includes semantic role annotation, word senses and aspectual classes (e.g., realis/irrealis). UDS emphasizes accessible distinctions, which can be collected through crowd-sourcing. However, the skeletal structure of UDS representations is derived from syntactic dependencies, and only includes verbal argument structures that can be so extracted. Notably, many of the distinctions in UDS are defined using feature bundles, rather than mutually exclusive categories. For instance, a semantic role may be represented as having the features +VOLITION and +AWARENESS, rather than as having the category AGENT.

The Prague Dependency Treebank (PDT) Tectogrammatical Layer (PDT-TL) (Sgall, 1992; Böhmová et al., 2003) covers a rich variety of functional and semantic distinctions, such as argument structure (including semantic roles), tense, ellipsis, topic/focus, co-reference, word sense disambiguation and local discourse information. The PDT-TL results from an abstraction over PDT's syntactic layers, and its close relation with syntax is apparent. For instance, the PDT-TL encodes the distinction between a governing clause and a dependent clause, which is primarily syntactic in nature, so in the clauses "John came just as we were leaving" and "We were leaving just as John came" the governing and dependent clause are swapped, despite their semantic similarity.

CCG-based Schemes. CCG (Steedman, 2000) is a lexicalized grammar (i.e., nearly all semantic content is encoded in the lexicon), which defines a theory of how lexical information is composed to form the meaning of phrases and sentences (see Section 6.2), and has proven effective in a variety of semantic tasks (Zettlemoyer and Collins, 2005, 2007; Kwiatkowski et al., 2010; Artzi and Zettlemoyer, 2013, *inter alia)*. Several projects have constructed logical representations by associating CCG with semantic forms (by assigning logical forms to the leaves). For example, Boxer (Bos, 2008) and GMB, which builds on Boxer, use Discourse Representation Structures (Kamp and Reyle, 1993), while Lewis and Steedman (2013) used Davidsonian-style λ-expressions, accompanied by lexical categorization of the predicates. These schemes encode events with their argument structures, and include an elaborate logical structure, as well as lexical and discourse information.

HPSG-based Schemes. Related to CCG-based schemes are SRTs based on Head-driven Phrase Structure Grammar (HPSG; Pollard and Sag, 1994), where syntactic and semantic features are represented as feature bundles, which are iteratively composed through unification rules to form composite units. HPSG-based SRT schemes commonly use the Minimal Recursion Semantics (Copestake et al., 2005) formalism. Annotated corpora and manually crafted grammars exist for multiple languages (Flickinger, 2002; Oepen et al., 2004; Bender and Flickinger, 2005, *inter alia)*, and generally focus on argument structural and logical semantic phenomena. The Broad-coverage Semantic Dependency Parsing shared task and corpora (Oepen et al., 2014, 2015) include corpora annotated with the PDT-TL, and dependencies extracted from the HPSG grammars Enju (Miyao, 2006) and the LinGO English Reference Grammar (ERG; Flickinger, 2002).

Like the PDT-TL, projects based on CCG, HPSG, and other expressive grammars such as LTAG (Joshi and Vijay-Shanker, 1999) and LFG (Kaplan and Bresnan, 1982) (e.g., GlueTag (Frank and van Genabith, 2001)), yield semantic representations that are coupled with syntactic ones. While this approach provides powerful tools for inference, type checking, and mapping into external formal languages, it also often results in difficulties in abstracting away from some syntactic details. For instance, the dependencies derived from ERG in the SDP corpus use the same label for different senses of the English possessive construction, regardless of whether they correspond to ownership (e.g., "John's dog") or to a different meaning, such as marking an argument of a nominal predicate (e.g., "John's kick"). See Section 6.

OntoNotes is a useful resource with multiple inter-linked layers of annotation, borrowed from different schemes. The layers include syntactic, SRL, co-reference and word sense disambiguation content. Some properties of the predicate, such as which nouns are eventive, are encoded as well.

To summarize, while SRT schemes differ in the types of content they support, schemes evolve to continuously add new content types, making these differences less consequential. The fundamental difference between the schemes is the extent that they abstract away from syntax. For instance, AMR and UCCA abstract away from syntax as part of their design, while in most other schemes syntax and semantics are more tightly coupled.

Schemes also differ in other aspects discussed in Sections 5 and 6.

5 Evaluation

Human evaluation is the ultimate criterion for validating an SRT scheme given our definition of semantics as meaning as it is understood by a language speaker. Determining how well an SRT scheme corresponds to human interpretation of a text is ideally carried out by asking annotators to make some semantic prediction or annotation according to pre-specified guidelines, and to compare this to the information extracted from the SRT. Question Answering SRL (QASRL; He et al., 2015) is an SRL scheme which solicits non-experts to answer mostly wh-questions, converting their output to an SRL annotation. Hartshorne et al. (2013) and Reisinger et al. (2015) use crowdsourcing to elicit semantic role features, such as whether the argument was volitional in the described event, in order to evaluate proposals for semantic role sets.

Another evaluation approach is task-based evaluation. Many semantic representations in NLP are defined with an application in mind, making this type of evaluation natural. For instance, a major motivation for AMR is its applicability to machine translation, making MT a natural (albeit hitherto unexplored) testbed for AMR evaluation. Another example is using question answering to evaluate semantic parsing into knowledge-base queries.

Another common criterion for evaluating a semantic scheme is *invariance*, where semantic analysis should be similar across paraphrases or translation pairs (Xue et al., 2014; Sulem et al., 2015). For instance, most SRL schemes abstract away from the syntactic divergence between the sentences (1) "He gave a present to John" and (2) "It was John who was given a present" (although a complete analysis would reflect the difference of focus between them).

Importantly, these evaluation criteria also apply in cases where the representation is automatically induced, rather than manually defined. For instance, vector space representations are generally evaluated either through task-based evaluation, or in terms of semantic features computed from them, whose validity is established by human annotators (e.g., Agirre et al., 2013, 2014).

Finally, where semantic schemes are induced through manual annotation (and not through automated procedures), a common criterion for determining whether the guidelines are sufficiently clear, and whether the categories are well-defined is to measure agreement between annotators, by assigning them the same texts and measuring the similarity of the resulting structures. Measures include the SMATCH measure for AMR (Cai and Knight, 2013), and the PARSEVAL F-score (Black et al., 1991) adapted for DAGs for UCCA.

SRT schemes diverge in the background and training they require from their annotators. Some schemes require extensive training (e.g., AMR), while others can be (at least partially) collected by crowdsourcing (e.g., UDS). Other examples include FrameNet, which requires expert annotators for creating new frames, but employs less trained in-house annotators for applying existing frames to texts; QASRL, which employs non-expert annotators remotely; and UCCA, which uses in-house non-experts, demonstrating no advantage to expert over non-expert annotators after an initial training period. Another approach is taken by GMB, which uses online collaboration where expert collaborators participate in manually correcting automatically created representations. They further employ gamification strategies for collecting some aspects of the annotation.

Universality. One of the great promises of semantic analysis (over more surface forms of analysis) is its cross-linguistic potential. However, while the theoretical and applicative importance of universality in semantics has long been recognized (Goddard, 2011), the nature of universal semantics remains unknown. Recently, projects such as BabelNet (Ehrmann et al., 2014), UBY (Gurevych et al., 2012) and Open Multilingual Wordnet[4], constructed huge multi-lingual semantic nets, by linking resources such as Wikipedia and WordNet and processing them using modern NLP. However, such projects currently focus on lexical semantic and encyclopedic information rather than on text semantics.

Symbolic SRT schemes such as SRL schemes and AMR have also been studied for their cross-linguistic applicability (Padó and Lapata, 2009; Sun et al., 2010; Xue et al., 2014), indicating partial portability across languages. Translated versions of PropBank and FrameNet have been constructed for multiple languages (e.g., Akbik et al., 2016; Hartmann and Gurevych, 2013). How-

[4]http://compling.hss.ntu.edu.sg/omw/

ever, as both PropBank and FrameNet are lexicalized schemes, and as lexicons diverge wildly across languages, these schemes require considerable adaptation when ported across languages (Kozhevnikov and Titov, 2013). Ongoing research tackles the generalization of VerbNet's unlexicalized roles to a universally applicable set (e.g., Schneider et al., 2015). Few SRT schemes place cross-linguistically applicability as one of their main criteria, examples include UCCA, and the LinGO Grammar Matrix (Bender and Flickinger, 2005), both of which draw on typological theory.

Vector space models, which embed words and sentences in a vector space, have also been applied to induce a shared cross-linguistic space (Klementiev et al., 2012; Rajendran et al., 2015; Wu et al., 2016). However, further evaluation is required in order to determine what aspects of meaning these representations reflect reliably.

6 Syntax and Semantics

6.1 Syntactic and Semantic Generalization

Syntactic distinctions are generally guided by a combination of semantic and distributional considerations, where emphasis varies across schemes.

Consider phrase-based syntactic structures, common examples of which, such as the Penn Treebank for English (Marcus et al., 1993) and the Penn Chinese Treebank (Xue et al., 2005), are adaptations of X-bar theory. Constituents are commonly defined in terms of distributional criteria, such as whether they can serve as conjuncts, be passivized, elided or fronted (Carnie, 2002, pp. 50-53). Moreover, phrase categories are defined according to the POS category of their headword, such as Noun Phrase, Verb Phrase or Preposition Phrase, which are also at least partly distributional, motivated by their similar morphological and syntactic distribution. In contrast, SRT schemes tend to abstract away from these realizational differences and directly reflect the argument structure of the sentence using the same set of categories, irrespective of the POS of the predicate, or the case marking of its arguments.

Distributional considerations are also apparent with functional syntactic schemes (the most commonly used form of which in NLP are lexicalist dependency structures), albeit to a lesser extent. A prominent example is Universal Dependencies (UD; Nivre et al., 2016), which aims at producing a cross-linguistically consistent dependency-based annotation, and whose categories are motivated by a combination of distributional and semantic considerations. For example, UD would distinguish between the dependency type between "John" and "brother" in "John, my brother, arrived" and "John, who is my brother, arrived", despite their similar semantics. This is due to the former invoking an apposition, and the latter a relative clause, which are different in their distribution.

As an example of the different categorization employed by UD and by purely semantic schemes such as AMR and UCCA consider (1) "founding of the school", (2) "president of the United States" and (3) "United States president". UD is faithful to the syntactic structure and represents (1) and (2) similarly, while assigning a different structure to (3). In contrast, AMR and UCCA perform a semantic generalization and represents examples (2) and (3) similarly and differently from (1).

6.2 The Syntax-Semantics Interface

A common assumption on the interface between syntax and semantics is that semantics of phrases and sentences is compositional – it is determined recursively by the meaning of its immediate constituents and their syntactic relationships, which are generally assumed to form a closed set (Montague, 1970, and much subsequent work). Thus, the interpretation of a sentence can be computed bottom-up, by establishing the meaning of individual words, and recursively composing them, to obtain the full sentential semantics. The order and type of these compositions are determined by the syntactic structure.

Compositionality is employed by linguistically expressive grammars, such as those based on CCG and HPSG, and has proven to be a powerful method for various applications. See (Bender et al., 2015) for a recent discussion of the advantages of compositional SRTs. Nevertheless, a compositional account meets difficulties when faced with multi-word expressions and in accounting for cases like "he sneezed the napkin off the table", where it is difficult to determine whether "sneezed" or "off" account for the constructional meaning. Construction Grammar (Fillmore et al., 1988; Goldberg, 1995) answers these issues by using an open set of construction-specific compositional operators, and supporting lexical entries of varying lengths. Several ongoing projects address the implementation of the principles of Construction Grammar into explicit grammars, including Sign-based Construction Grammar (Fillmore et al., 2012), Embodied Construction Grammar (Feldman et al., 2010) and Fluid Construction Grammar (Steels and de Beules, 2006).

The achievements of machine learning methods in many areas, and optimism as to its prospects, have enabled the approaches to semantics discussed in this paper. Machine learning allows to define semantic structures on purely semantic grounds and to let algorithms identify how these distinctions are mapped to surface/distributional forms. Some of the schemes discussed in this paper take this approach in its pure form (e.g., AMR and UCCA).

7 Conclusion

Semantic representation in NLP is undergoing rapid changes. Traditional semantic work has either used shallow methods that focus on specific semantic phenomena, or adopted formal semantic theories which are coupled with a syntactic scheme through a theory of the syntax-semantics interface. Recent years have seen increasing interest in an alternative approach that defines semantic structures independently from any syntactic or distributional criteria, much due to the availability of semantic treebanks that implement this approach.

Semantic schemes diverge in whether they are anchored in the words and phrases of the text (e.g., all types of semantic dependencies and UCCA) or not (e.g., AMR and logic-based representations). We do not view this as a major difference, because most unanchored representations (including AMR) retain their close affinity with the words of the sentence, possibly because of the absence of a workable scheme for lexical decomposition, while dependency structures can be converted into logic-based representations (Reddy et al., 2016). In practice, anchoring facilitates parsing, while unanchored representations are more flexible to use where words and semantic components are not in a one-to-one correspondence.

Our survey concludes that the main distinguishing factors between schemes are their relation to syntax, their degree of universality, and the expertise and training they require from annotators, an important factor in addressing the annotation bottleneck. We hope this survey of the state of the art in semantic representation will promote discussion, expose more researchers to the most pressing questions in semantic representation, and lead to the wide adoption of the best components from each scheme.

Acknowledgements. We thank Nathan Schneider for his helpful comments. The work was support by the Intel Collaborative Research Institute for Computational Intelligence (ICRI-CI).

References

Omri Abend and Ari Rappoport. 2013a. UCCA: A semantic-based grammatical annotation scheme. In *Proc. of IWCS*. pages 1–12.

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

Eneko Agirre, Carmen Banea, Claire Cardie, Daniel Cer, Mona Diab, Aitor Gonzalez-Agirre, Weiwei Guo, Rada Mihalcea, German Rigau, and Janyce Wiebe. 2014. Semeval-2014 task 10: Multilingual semantic textual similarity. In *Proc. of SemEval*. pages 81–91.

Eneko Agirre, Daniel Cer, Mona Diab, Aitor Gonzalez-Agirre, and Weiwei Guo. 2013. *sem 2013 shared task: Semantic textual similarity. In *Proc. of SemEval*. pages 32–43.

Alan Akbik, vishwajeet kumar, and Yunyao Li. 2016. Towards semi-automatic generation of proposition banks for low-resource languages. In *Proc. of EMNLP*. pages 993–998.

Waleed Ammar, George Mulcaire, Yulia Tsvetkov, Guillaume Lample, Chris Dyer, and Noah A. Smith. 2016. Massively multilingual word embeddings. *CoRR* abs/1602.01925.

Daniel Andor, Chris Alberti, David Weiss, Aliaksei Severyn, Alessandro Presta, Kuzman Ganchev, Slav Petrov, and Michael Collins. 2016. Globally normalized transition-based neural networks. In *Proc. of ACL*. pages 2442–2452.

Gabor Angeli and Christopher D Manning. 2014. Naturalli: Natural logic inference for common sense reasoning. In *EMNLP*. pages 534–545.

Yoav Artzi and Luke Zettlemoyer. 2013. Weakly supervised learning of semantic parsers for mapping instructions to actions. *TACL* 1:49–62.

Dzmitry Bahdanau, KyungHyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *Proc. of ICLR*.

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 LAW*. pages 178–186.

Yehoshua Bar-Hillel. 1960. The present status of automatic translation of languages. In *Advances in computers*, Academic Press, New York, volume 1, pages 91–163.

Valerio Basile, Johan Bos, Kilian Evang, and Noortje Venhuizen. 2012. Developing a large semantically annotated corpus. In *Proc. of LREC*. pages 3196–3200.

Emily Bender and Dan Flickinger. 2005. Rapid prototyping of scalable grammars: Towards modularity in extensions to a language-independent core. In *Proc. of IJCNLP*. pages 203–208.

Emily M. Bender, Dan Flickinger, Stephan Oepen, Woodley Packard, and Ann Copestake. 2015. Layers of interpretation: On grammar and compositionality. In *Proc. of IWCS*. pages 239–249.

Ezra Black, Steve Abney, Dan Flickinger, C. Gdaniec, Ralph Grishman, P. Harrison, Donald Hindle, Robert Ingria, Frederick Jelinek, Judith Klavans, Mark Liberman, Mitch Marcus, Salim Roukos, Beatrice Santorini, and Thomas Strzalkowski. 1991. A procedure for quantitatively comparing the syntactic coverage of English grammars. In *Proc. of the DARPA Speech and Natural Language Workshop*. pages 204–210.

Alena Böhmová, Jan Hajič, Eva Hajičová, and Barbora Hladká. 2003. The Prague dependency treebank. In *Treebanks*, Springer, pages 103–127.

Claire Bonial, William Corvey, Martha Palmer, Volha V Petukhova, and Harry Bunt. 2011. A hierarchical unification of lirics and verbnet semantic roles. In *Semantic Computing (ICSC)*. pages 483–489.

Johan Bos. 2008. Wide-coverage semantic analysis with Boxer. In Johan Bos and Rodolfo Delmonte, editors, *Proc. of the Conference on Semantics in Text Processing (STEP)*. College Publications, Research in Computational Semantics, pages 277–286.

Johan Bos and Katja Markert. 2005. Recognising textual entailment with logical inference. In *Proc. of EMNLP*. pages 628–635.

Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. 2015. A large annotated corpus for learning natural language inference. In *Proc. of EMNLP*. pages 632–642.

Shu Cai and Kevin Knight. 2013. Smatch: an evaluation metric for semantic feature structures. In *Proc. of ACL*. pages 748–752.

Lynn Carlson, Daniel Marcu, and Mary Ellen Okurowski. 2003. Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *Current and new directions in discourse and dialogue*, Springer, pages 85–112.

Andrew Carnie. 2002. *Syntax: A Generative Introduction*. Wiley-Blackwell.

Nathanael Chambers and Dan Jurafsky. 2008. Unsupervised learning of narrative event chains. In *Proc. of ACL-HLT*. pages 789–797.

Nathanael Chambers and Dan Jurafsky. 2009. Unsupervised learning of narrative schemas and their participants. In *Proc. of ACL-IJCNLP*. pages 602–610.

Ann Copestake, Dan Flickinger, Carl Pollard, and Ivan A. Sag. 2005. Minimal recursion semantics: An introduction. *Research on Language and Computation* 3:281–332.

Ido Dagan, Oren Glickman, and Bernardo Magnini. 2006. The PASCAL recognising text entailment challenge. In Bernardo Magnini Joaquin Quiñonero Candela, Ido Dagan and Florence d'Alché Buc, editors, *Machine Learning Challenges*, Springer, Berlin, volume 3944 of *Lecture Notes in Computer Science*, pages 177–190.

Ido Dagan, Dan Roth, and Mark Sammons. 2013. *Recognizing textual entailment*. Morgan & Claypool Publishers.

Robert M.W. Dixon. 2010. *Basic Linguistic Theory: Methodology*, volume 1. Oxford University Press.

David Dowty. 2003. The dual analysis of adjuncts/complements in categorial grammar. In Ewald Lang, Claudia Maienborn, and Cathry Fabricius-Hansen, editors, *Modifying Adjuncts*, Mouton de Gruyter, Berlin, pages 33–66.

Jesse Dunietz, Lori Levin, and Jaime Carbonell. 2015. Annotating causal language using corpus lexicography of constructions. In *Proc. of LAW*. pages 188–196.

Maud Ehrmann, Francesco Cecconi, Daniele Vannella, John Philip McCrae, Philipp Cimiano, and Roberto Navigli. 2014. Representing multilingual data as linked data: the case of babelnet 2.0. In *Proc. of LREC*. pages 401–408.

David K Elson and Kathleen R McKeown. 2010. Tense and aspect assignment in narrative discourse. In *Proc. of the International Natural Language Generation Conference*. pages 47–56.

Jerome Feldman, Ellen Dodge, and John Bryant. 2010. Embodied construction grammar. In Bernd Heine and Heiko Narrog, editors, *The Oxford Handbook of Linguistic Analysis*, Oxford University Press, pages 111–158.

Charles Fillmore, Russell Lee-Goldman, and Russell Rhodes. 2012. The FrameNet Constructicon. In Hans Boas and Ivan Sag, editors, *Sign-based construction grammar*, CSLI Publications, pages 309–372.

Charles J Fillmore, Paul Kay, and Mary C O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language* 64(3):501–538.

Daniel Flickinger. 2002. On building a more efficient grammar by exploiting types. In Jun'ichi Tsujii, Stefan Oepen, Daniel Flickinger, and Hans Uszkoreit, editors, *Collaborative Language Engineering*, CLSI, Stanford, CA.

Jerry A Fodor. 1975. *The language of thought*, volume 5. Harvard University Press.

Anette Frank and Josef van Genabith. 2001. Gluetag: Linear logic based semantics construction for ltag and what it teaches us about the relation between LFG and LTAG. In *Proc. of LFG*.

Annemarie Friedrich, Alexis Palmer, and Manfred Pinkal. 2016. Situation entity types: automatic classification of clause-level aspect. In *Proceedings of ACL 2016*. pages 1757–1768.

Matthew Gerber and Joyce Y Chai. 2010. Beyond nombank: A study of implicit arguments for nominal predicates. In *Proc. of ACL*. pages 1583–1592.

Daniel Gildea and Dan Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics* 28(3):245–288.

Cliff Goddard. 2011. *Semantic analysis: A practical introduction*. Oxford University Press, 2nd edition.

Adèle Goldberg. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago University Press, Chicago.

Iryna Gurevych, Judith Eckle-Kohler, Silvana Hartmann, Michael Matuschek, Christian M. Meyer, and Christian Wirth. 2012. UBY - a large-scale unified lexical-semantic resource based on lmf. In *Proc. of EACL*. pages 580–590.

Silvana Hartmann and Iryna Gurevych. 2013. Framenet on the way to babel: Creating a bilingual framenet using wiktionary as interlingual connection. In *Proc. of ACL*. pages 1363–1373.

Joshua K. Hartshorne, Claire Bonial, and Martha Palmer. 2013. The VerbCorner project: Toward an empirically-based semantic decomposition of verbs. In *Proc. of EMNLP*. pages 1438–1442.

Luheng He, Mike Lewis, and Luke Zettlemoyer. 2015. Question-answer driven semantic role labeling: Using natural language to annotate natural language. In *Proc. of EMNLP*. pages 643–653.

Kevin Humphreys, Robert Gaizauskas, and Saliha Azzam. 1997. Event coreference for information extraction. In *Proc. of a Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts*. pages 75–81.

Rei Ikuta, Will Styler, Mariah Hamang, Tim O'Gorman, and Martha Palmer. 2014. Challenges of adding causation to richer event descriptions. In *Proc. of the Second Workshop on EVENTS: Definition, Detection, Coreference, and Representation*. pages 12–20.

Aravind Joshi and K. Vijay-Shanker. 1999. Compositional semantics with Lexicalized Tree-Adjoining Grammar (LTAG). In *Proc. of IWCS*. pages 131–146.

Hans Kamp and Uwe Reyle. 1993. *From Discourse to Logic*. Kluwer, Dordrecht.

Ronald M Kaplan and Joan Bresnan. 1982. Lexical-functional grammar: A formal system for grammatical representation. *Formal Issues in Lexical-Functional Grammar* pages 29–130.

Karen Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2008. A large-scale classification of English verbs. *Language Resources and Evaluation* 42:21–40.

Ryan Kiros, Ruslan Salakhutdinov, and Richard S. Zemel. 2014. Unifying visual-semantic embeddings with multimodal neural language models. *CoRR* abs/1411.2539.

Alexandre Klementiev, Ivan Titov, and Binod Bhattarai. 2012. Inducing crosslingual distributed representations of words. In *Proc. of COLING*. pages 1459–1474.

Parisa Kordjamshidi, Steven Bethard, and Marie-Francine Moens. 2012. Semeval-2012 task 3: Spatial role labeling. In *In Proc. of *SEM*. pages 365–373.

Mikhail Kozhevnikov and Ivan Titov. 2013. Cross-lingual transfer of semantic role labeling models. In *Proc. of ACL*. pages 1190–1200.

Tom Kwiatkowski, Luke Zettlemoyer, Sharon Goldwater, and Mark Steedman. 2010. Inducing probabilistic CCG grammars from logical form with higher-order unification. In *Proc. of EMNLP*. pages 1223–1233.

Ronald Langacker. 2008. *Cognitive Grammar: A Basic Introduction*. Oxford University Press, Oxford.

Beth Levin and Malka Rappaport Hovav. 2005. *Argument realization*. Cambridge University Press.

Michael Lewis and Mark Steedman. 2013. Combined distributional and logical semantics. *TACL* 1:179–192.

Maria Liakata, Simone Teufel, Advaith Siddharthan, and Colin Batchelor. 2010. Corpora for the conceptualisation and zoning of scientific papers. In *Proc. of LREC*. pages 2054–2061.

Edward Loper, Szu-Ting Yi, and Martha Palmer. 2007. Combining lexical resources: Mapping between PropBank and VerbNet. In *Proc. of the 7th International Workshop on Computational Linguistics*.

Christopher Manning. 2006. Local textual inference: It's hard to circumscribe, but you know it when you see it—and nlp needs it. unpublished ms.

Mitch Marcus, Beatrice Santorini, and M. Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics* 19:313–330.

Eleni Miltsakaki, Rashmi Prasad, Aravind K Joshi, and Bonnie L Webber. 2004. The penn discourse treebank. In *LREC*. pages 2237–2240.

Paramita Mirza, Rachele Sprugnoli, Sara Tonelli, and Manuela Speranza. 2014. Annotating causality in the tempeval-3 corpus. In *Proc. of the EACL Workshop on Computational Approaches to Causality in Language (CAtoCL)*. pages 10–19.

Yusuke Miyao. 2006. *Corpus-oriented grammar development and feature forest model*. Ph.D. thesis, University of Tokyo.

Marc Moens and Mark Steedman. 1988. Temporal ontology and temporal reference. *Computational Linguistics* 14:15–28. Reprinted in Inderjeet Mani, James Pustejovsky, and Robert Gaizauskas (eds.) *The Language of Time: A Reader*. Oxford University Press, 93-114.

Richard Montague. 1970. English as a formal language. In Bruno Visentini, editor, *Linguaggi nella Società e nella Technica*, Edizioni di Communità, Milan, pages 189–224. Reprinted as Thomason 1974:188-221.

Nasrin Mostafazadeh, Alyson Grealish, Nathanael Chambers, James Allen, and Lucy Vanderwende. 2016. Caters: Causal and temporal relation scheme for semantic annotation of event structures. In *Proc. of the Fourth Workshop on Events*. pages 51–61.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. 2016. Universal dependencies v1: A multilingual treebank collection. In *Proc. of LREC*. pages 1659–1666.

Stephan Oepen, Dan Flickinger, Kristina Toutanova, and Chris Manning. 2004. Lingo Redwoods. *Research on Language & Computation* 2:575–596.

Stephan Oepen, 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 *Proc. of SemEval*. pages 915–926.

Stephan Oepen, 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 *Proc. of SemEval*. pages 63–72.

Sebastian Padó and Mirella Lapata. 2009. Cross-lingual annotation projection of semantic roles. *Journal of Artificial Intelligence Research* 36:307–340.

Alexis Palmer, Elias Ponvert, Jason Baldridge, and Carlota Smith. 2007. A sequencing model for situation entity classification. In *Proc. of ACL*. pages 896–903.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics* 31(1):71–106.

Martha Palmer, Daniel Gildea, and Nianwen Xue. 2010. *Semantic Role Labeling*. Synthesis lectures on human language technologies. Morgan & Claypool Publishers.

Martha Palmer, Ivan Titov, and Shumin Wu. 2013. Semantic role labeling tutorial at naacl 2013. `http://ivan-titov.org/teaching/srl-tutorial-naacl13/`.

Carl Pollard and Ivan A Sag. 1994. *Head-driven phrase structure grammar*. University of Chicago Press.

James Pustejovsky, José Casteño, Robert Ingria, Roser Saurí, Robert Gaizauiuskas, Andrea Setzer, Graham Katz, and Dragomir Radev. 2003. Timeml: Robust specification of event and temporal expressions in text. In *Proc. of the 5th International Workshop on Computational Semantics*.

James Pustejovsky, Parisa Kordjamshidi, Marie-Francine Moens, Aaron Levine, Seth Dworman, and Zachary Yocum. 2015. Semeval-2015 task 8: Spaceeval. In *Proc. of SemEval*. pages 884–894.

Janarthanan Rajendran, Mitesh M. Khapra, Sarath Chandar, and Balaraman Ravindran. 2015. Bridge correlational neural networks for multilingual multimodal representation learning. *CoRR* abs/1510.03519.

Siva Reddy, Oscar Täckström, Michael Collins, Tom Kwiatkowski, Dipanjan Das, Mark Steedman, and Mirella Lapata. 2016. Transforming dependency structures to logical forms for semantic parsing. *TACL* 4:127–140.

Michaela Regneri, Alexander Koller, and Manfred Pinkal. 2010. Learning script knowledge with web experiments. In *Proc. of ACL*. pages 979–988.

Drew Reisinger, Rachel Rudinger, Francis Ferraro, Craig Harman, Kyle Rawlins, and Benjamin Van Durme. 2015. Semantic proto-roles. *TACL* 3:475–488.

Michael Roth and Anette Frank. 2015. Inducing implicit arguments from comparable texts: A framework and its applications. *Computational Linguistics* 41:625–664.

Josef Ruppenhofer, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson, Collin F. Baker, and Jan Scheffczyk. 2016. *FrameNet II: Extended Theory and Practice*. The Berkeley FrameNet Project.

Nathan Schneider, Vivek Srikumar, Jena D. Hwang, and Martha Palmer. 2015. A hierarchy with, of, and for preposition supersenses. In *Proc. of LAW*. pages 112–123.

Petr Sgall. 1992. Underlying Structure of Sentences and Its Relations to Semantics. In T. Reuthe, editor, *Wiener Slawistischer Almanach. Sonderband 33*, Wien: Gesellschaft zur Förderung slawistischer Studien, pages 273–282.

Eric Siegel and Kathy McKeown. 2000. Learning methods to combine linguistic indicators: Improving aspectual classification and revealing linguistic insights. *Computational Linguistics* 26:595–628.

Richard Socher, John Bauer, Christopher D. Manning, and Ng Andrew Y. 2013. Parsing with compositional vector grammars. In *Proc. of ACL*. pages 455–465.

Mark Steedman. 2000. *The Syntactic Process*. MIT Press, Cambridge, MA.

Luc Steels and Joachim de Beules. 2006. A (very) brief introduction to fluid construction grammar. In *Proc. of the 3rd Workshop on Scalable Natural Language Understanding*. pages 73–80.

Elior Sulem, Omri Abend, and Ari Rappoport. 2015. Conceptual annotations preserve structure across translations: A French-English case study. In *ACL 2015 Workshop on Semantics-Driven Statistical Machine Translation (S2MT)*. pages 11–22.

Lin Sun, Anna Korhonen, Thierry Poibeau, and Cédric Messiant. 2010. Investigating the cross-linguistic potential of verbnet: style classification. In *Proc. of COLING*. pages 1056–1064.

Richmond Thomason, editor. 1974. *Formal Philosophy: Papers of Richard Montague*. Yale University Press, New Haven, CT.

Naushad UzZaman, Hector Llorens, Leon Derczynski, James Allen, Marc Verhagen, and James Pustejovsky. 2013. Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations. In **SEM-SemEval '13*. pages 1–9.

Jan van Eijck. 2005. Natural logic for natural language. In Balder ten Cate and Henk Zeevat, editors, *Logic, Language, and Computation*. Springer, Berlin, Lecture Notes in Computer Science 4363, pages 216–230.

Marc Verhagen, Roser Sauri, Tomasso Caselli, and James Pustejovsky. 2010. Semeval-2010 task 13: Tempeval-2. In *Proc. of the 5th International Workshop on Semantic Evaluation*. ACL, pages 57–62.

Mark Verhagen, Robert Gaizauskas, Frank Schilder, Mark Hepple, Jessica Moszkowitcz, and James Pustejovsky. 2009. The tempeval challenge: Identifying temporal relations in text. *Language Resources and Evaluation* 43:161–179.

Bonnie Webber, Markus Egg, and Valia Kordoni. 2011. Discourse structure and language technology. *Natural Language Engineering* 18(4):437–490.

Bonnie Webber and Aravind Joshi. 2012. Discourse structure and computation: Past, present and future. In *Proceedings of the ACL-2012 Special Workshop on Rediscovering 50 Years of Discoveries*. pages 42–54.

Aaron Steven White, Drew Reisinger, Keisuke Sakaguchi, Tim Vieira, Sheng Zhang, Rachel Rudinger, Kyle Rawlins, and Benjamin Van Durme. 2016. Universal decompositional semantics on universal dependencies. In *Proc. of EMNLP*. pages 1713–1723.

Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Lukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. *CoRR* abs/1609.08144.

Naiwen Xue, Fei Xia, Fu-Dong Chiou, and Marta Palmer. 2005. The penn chinese treebank: Phrase structure annotation of a large corpus. *Natural language engineering* 11(02):207–238.

Nianwen Xue, Odrej Bojar, Jan Hajic, Martha Palmer, Zdenka Uresova, and Xiuhong Zhang. 2014. Not an intelingua, but close: comparison of English AMRs to Chinese and Czech. In *Proc. of LREC*. pages 1765–1772.

John Zelle and Ray Mooney. 1996. Learning to parse database queries using inductive logic programming. In *Proc. of the 14th National Conference on Artificial Intelligence*. pages 1050–1055.

Luke Zettlemoyer and Michael Collins. 2005. Learning to map sentences to logical form: Structured classification with Probabilistic Categorial Grammars. In *Proc. of UAI*. pages 658–666.

Luke Zettlemoyer and Michael Collins. 2007. Online learning of relaxed CCG grammars for parsing to logical form. In *Proc. of EMNLP-CoNLL*. pages 678–687.