

Semantic Role Labeling as Syntactic Dependency Parsing

Tianze Shi*

Cornell University
tianze@cs.cornell.edu

Igor Malioutov

Bloomberg L.P.
imalioutov@bloomberg.net

Ozan Irsoy

Bloomberg L.P.
oirsoy@bloomberg.net

Abstract

We reduce the task of (span-based) PropBank-style semantic role labeling (SRL) to syntactic dependency parsing. Our approach is motivated by our empirical analysis that shows three common syntactic patterns account for over 98% of the SRL annotations for both English and Chinese data. Based on this observation, we present a conversion scheme that packs SRL annotations into dependency tree representations through joint labels that permit highly accurate recovery back to the original format. This representation allows us to train statistical dependency parsers to tackle SRL and achieve competitive performance with the current state of the art. Our findings show the promise of syntactic dependency trees in encoding semantic role relations within their syntactic domain of locality, and point to potential further integration of syntactic methods into semantic role labeling in the future.

1 Introduction

Semantic role labeling (SRL; Palmer et al., 2010) analyzes texts with respect to predicate argument structures such as “*who did what to whom, and how, when and where*”. These generic surface semantic representations provide richer linguistic analysis than syntactic parsing alone and are useful in a wide range of downstream applications including question answering (Shen and Lapata, 2007; Khashabi et al., 2018), open-domain information extraction (Christensen et al., 2010), clinical narrative understanding (Albright et al., 2013), automatic summarization (Khan et al., 2015) and machine translation (Liu and Gildea, 2010; Xiong et al., 2012; Bazrafshan and Gildea, 2013), among others.

It is commonly acknowledged that syntax and semantics are tightly coupled with each other (Levin

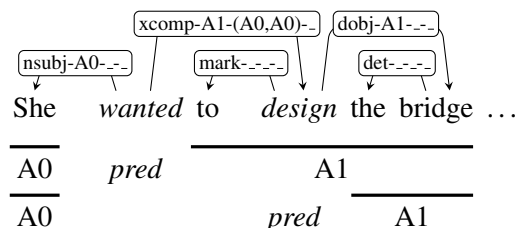


Figure 1: An example sentence with SRL annotations (below) and our joint syntacto-semantic dependency relations (above; described in §3). The two representations can be converted from one to the other. A0 and A1 are short for SRL relations ARG0 and ARG1.

and Hovav, 2005). In some forms of linguistic theories (Baker, 1996, 1997), semantic arguments are even hypothesized to be assigned under consistent and specific syntactic configurations. As a matter of practice, annotations of semantic roles (Palmer et al., 2005, *inter alia*) are typically based on existing syntactic treebanks as an additional annotation layer. Annotators are instructed (Babko-Malaya et al., 2006; Bonial et al., 2015) to identify semantic arguments within the predicates’ domain of locality,¹ respecting the strong connection between syntax and semantics.

Empirically, syntax has indeed been shown to be helpful to SRL in a variety of ways. Earlier SRL systems have successfully incorporated syntactic parse trees as features and pruning signals (Punyakanok et al., 2008). Recently, neural models with shared representations trained to predict both syntactic trees and predicate-argument structures in a multi-task learning setting achieve superior performance to syntax-agnostic models (Strubell et al., 2018; Swayamdipta et al., 2018), reinforcing the utility of syntax in SRL.

However, researchers are yet to fully leverage

¹The arguments can potentially be traces and null elements. If a trace is selected as an argument, it is automatically chained to its surface constituent after syntactic movement.

*Work done during an internship at Bloomberg L.P.

all the theoretical linguistic assumptions and the dataset annotation conventions surrounding the tight connections between syntax and SRL. To do so, ideally, one must perform deep syntactic processing to capture long-distance dependencies and argument sharing. One solution is to introduce traces into phrase-structure trees, which, unfortunately, is beyond the scope of most statistical constituency parsers partially due to their associated increased complexity (Kummerfeld and Klein, 2017). Another solution is to use richer grammar formalisms with feature structures such as combinatory categorial grammar (CCG; Steedman, 2000) and tree adjoining grammar (TAG; Joshi et al., 1975) that directly build syntactic relations within the predicates’ *extended* domain of locality. It is then possible to restrict the semantic argument candidates to only those “local” dependencies (Gildea and Hockenmaier, 2003; Liu, 2009; Liu and Sarkar, 2009; Konstas et al., 2014; Lewis et al., 2015). However, such treebank data are harder to obtain, and their parsing algorithms tend to be less efficient than parsing probabilistic context-free grammars (Kallmeyer, 2010).

On the other hand, syntactic dependency trees directly encode bilocal governor-dependent relations among the surface tokens, which implicitly extend the domain of locality (Schneider, 2008). Dependency parsing (Kübler et al., 2008) is empirically attractive for its simplicity, data availability, efficient and accurate parsing algorithms, and its tight connection to semantic analysis (Reddy et al., 2017). Despite ample research community interest in joint models for dependency parsing and SRL (Surdeanu et al., 2008; Hajič et al., 2009; Henderson et al., 2013), a precise characterization of the mapping between semantic arguments and syntactic configurations has been lacking.

In this paper, we provide a detailed empirical account of PropBank-style SRL annotations on both English and Chinese data. We show that a vast majority (over 98%) of the semantic relations are characterized by one of three basic dependency-based syntactic configurations: the semantic predicate 1) directly dominates, 2) is directly dominated by, or 3) shares a common syntactic governor with the semantic argument. The latter two cases are mostly represented by syntactic constructions including relativization, control, raising, and coordination.

Based on our observations, we design a back-and-forth conversion algorithm that embeds SRL

relations into dependency trees. The SRL relations are appended to the syntactic labels to form joint labels, while the syntactic governor for each token remains unaltered. The algorithms reach over 99% F1 score on English and over 97% on Chinese data in oracle back-and-forth conversion experiments. Further, we train statistical dependency parsing models that simultaneously predict SRL and dependency relations through these joint labels. Experiments show that our fused syntacto-semantic models achieve competitive performance with the state of the art.

Our findings show the promise of dependency trees in encoding PropBank-style semantic role relations: they have great potential in reducing the task of SRL to dependency parsing with an expanded label space. Such a task reduction facilitates future research into finding an empirically adequate granularity for representing SRL relations. It also opens up future possibilities for further integration of syntactic methods into SRL as well as adaptations of extensively-studied dependency parsing techniques to SRL, including linear-time decoding, efficiency-performance tradeoffs, multilingual knowledge transfer, and more. We hope our work can inspire future research into syntactic treatment of other shallow semantic representations such as FrameNet-style SRL (Baker et al., 1998; Fillmore et al., 2003). Our code is available at https://www.github.com/bloomberg/emnlp20_depsrl.

Contribution Our work (1) provides a detailed empirical analysis of the syntactic structures of semantic roles, (2) characterizes the tight connections between syntax and SRL with three repeating structural configurations, (3) proposes a back-and-forth conversion method that supports a fully-syntactic approach to SRL, and (4) shows through experiments that dependency parsers can reach competitive performance with the state of the art on span-based SRL. Additionally, (5) all our analysis, methods and results apply to two languages from distinctive language families, English and Chinese.

2 Syntactic Structures of Semantic Roles

It has been widely assumed in linguistic theories that the semantic representations of arguments are closely related to their syntactic positions with respect to the predicates (Gruber, 1965; Jackendoff, 1972, 1992; Fillmore, 1976; Baker, 1985; Levin,

1993).² This notion is articulated as linguistic hypotheses underlying many syntactic theories:

- (1) *Universal Alignment Hypothesis*: There exist principles of Universal Grammar which predict the initial [grammatical] relation borne by each nominal in a given clause from the meaning of the clause. (Perlmutter and Postal, 1984, p. 97)
- (2) *The Uniformity of Theta Assignment Hypothesis*: Identical thematic relationships between items are represented by identical structural relationships between those items at the level of D[EEP]-structure. (Baker, 1985, p. 57)

For theories that posit one-to-one correspondence between semantic roles and syntactic structures (Baker, 1996, 1997), SRL can be treated purely as a syntactic task. However, doing so would require deep structural analysis (Bowers, 2010) that hypothesizes more functional categories than what current syntactic annotations cover.

Nonetheless, the Proposition Bank (PropBank; Kingsbury and Palmer, 2002; Palmer et al., 2005) annotations do capture the domain of locality that is implicitly assumed by these linguistic theories. PropBank defines the domain of locality for verbal predicates to be indicated by “clausal boundary markers” and the annotators are instructed to limit their semantic role annotations to “the sisters of the verb relation (for example, the direct object) and the sisters of the verb phrase (for example, the subject)” (Bonial et al., 2017, p. 746). In cases of syntactically-displaced arguments, the annotators are asked to pick the empty elements that are within the domain of locality, and then syntactic coindexation chains are used to reconstruct the surface semantic role relations. Recognizing displaced arguments is crucial to SRL, so taking full advantage of locality constraints would also require modeling empty elements and movement, for which current NLP systems still lack accurate, efficient, and high-coverage solutions (Gabbard et al., 2006; Kummerfeld and Klein, 2017).

From an empirical perspective, most syntactic realizations for semantic arguments follow certain common patterns even when they are displaced. Indeed, this is partially why syntax-based features and candidate pruning heuristics have been suc-

cessful in SRL (Gildea and Palmer, 2002; Gildea and Jurafsky, 2002; Sun et al., 2008). Full parsing might not be necessary to account for the majority of cases in the annotations. Thus, knowing the empirical distributions of the arguments’ syntactic positions would be highly useful for deciding how detailed the syntactic analysis needs to be for the purpose of SRL. In this section, we provide such a characterization.

Our analysis is based on dependency syntax and complements prior constituent-based characterizations (Palmer et al., 2005). One advantage of syntactic dependencies over phrase-structure trees for the purposes of this paper is that the dependents are often more directly connected to the syntactic governors without intervening intermediate constituents. For example, when a verb has multiple adjunct modifiers, each would create an additional intermediate VP constituent in the argument structure, leading to further separation between the verb and the external argument (subject). In contrast, in a dependency representation, the subject is always directly dominated by the verbal predicate.

2.1 Material

We use the training splits of the CoNLL 2012 shared task data (Pradhan et al., 2012) on both English and Chinese; sentences are originally from OntoNotes 5.0 (Hovy et al., 2006). The SRL annotations are based on English and Chinese PropBank (Kingsbury and Palmer, 2002; Palmer et al., 2005; Xue and Palmer, 2003; Xue, 2008), which are extensively used in SRL research. We choose not to use the SRL-targeted CoNLL 2005 shared task (Carreras and Màrquez, 2005) data since earlier versions of PropBank (Babko-Malaya, 2005) contain many resolvable mismatches between syntactic and semantic annotations (Babko-Malaya et al., 2006). Updated annotation guidelines (Bonial et al., 2015) have fixed most of the identified issues. We convert the Penn TreeBank (PTB; Marcus et al., 1993) and the Penn Chinese TreeBank (CTB; Xue et al., 2005) phrase-structure trees into Stanford Dependencies (SD; de Marneffe et al., 2006) for English (de Marneffe and Manning, 2008; Silveira et al., 2014) and for Chinese (Chang et al., 2009) respectively.³ SD is semantically-friendly as noted by Schuster and Manning (2016, p. 2371), “Since its first version, SD representation has had the status of being both

²This is often termed *linking theory* in linguistics (See Levin and Hovav (2005) for a survey).

³During conversion, we set copular verbs to be heads, since PropBank marks some copular verbs as predicates.

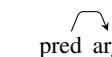
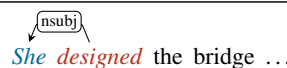

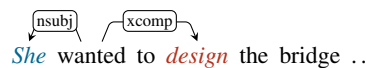
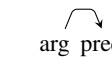
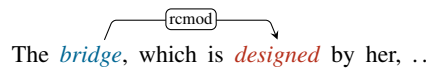
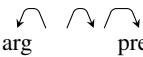
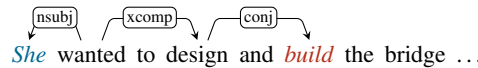
	Pattern	Example	Percentage	
			English	Chinese
(D)			87.5%	82.7%
(C)			6.1%	10.4%
(R)			4.7%	5.7%
			1.1%	1.0%
		Others	0.5%	0.2%

Table 1: The most common structural relations in the training data between the *predicates* (pred) and the *arguments* (arg). Appendix §C and §D include more examples as well as Chinese data.

a syntactic and a shallow semantic representation”, thus it is suited for the development of our joint modeling of syntactic and semantic structures. Indeed, Universal Dependencies (UD; Nivre et al., 2016), which builds upon SD, has been compared with and aligned to meaning representations including UCCA (Hershcovich et al., 2019) and AMR (Szubert et al., 2018).⁴

2.2 Observations

We categorize the syntactic configurations between predicates and arguments and present the results in Table 1. For both English and Chinese, the vast majority, more than 98%, of the predicate-argument relations fall into one of three major categories: the semantic argument is a syntactic child, sibling, or parent of the semantic predicate. Next, we give a brief account of our linguistic observations on the English data associated with each category. See Appendix §C and §D for more examples from both English and Chinese.

pred → arg (D) The predicate directly (D) dominates the semantic argument in the syntactic tree. Not surprisingly, this straightforward type of relation is the most prevalent in the PropBank data, accounting for more than 87% (82%) of all English (Chinese) predicate-argument relations.

arg ← → pred (C) The predicate and the argument share a common (C) syntactic parent. There are two major types of constructions resulting in this kind of configuration: 1) the common parent is a control or raising predicate, creating an open

⁴Our choice of SD instead of UD is motivated by the flexibility in conversion to set copular verbs as syntactic heads.

clausal complement (xcomp) relation and 2) there is a coordination structure between the predicate and the common parent and both predicates share a same argument in the semantic structure. Both cases are so common that they are converted to direct dependencies in the enhanced Stanford Dependencies (Schuster and Manning, 2016).

arg → pred (R) The dominance relation between the predicate and the argument is reversed (R). This type of relations is frequently realized through relative clauses (rcmod) and verb participles (e.g., *broken glass*).

Other constructions Many other constructions can be analyzed as combinations of the previously mentioned patterns.⁵ For example, a combination of (C)+(C) through control and coordination would derive the structural configuration of the fourth most frequent case in Table 1.

3 Reducing SRL to Dependency Parsing

3.1 Joint Labels

Building on the insights obtained from our analysis, we design a joint label space to encode both syntactic and SRL relations. The joint labels have four components: one syntactic relation and three semantic labels, each corresponding to one of the three most common structural patterns in Table 1.

⁵Combinations of (D), (C), and (R) can theoretically account for all possible predicate-argument configurations. However, for a lossless back-and-forth conversion with our proposed joint labels (§3), there are constraints on the argument structures of all the intermediate predicates along the shortest dependency path between the predicate and the argument. See Table 2 for an estimation of how many semantic relations may be decomposed as combinations of the three common structural patterns empirically given our conversion method.

Formally, for a length- n input sentence $w = w_1, \dots, w_n$, we denote the head of token w_i in the syntactic dependency tree t to be w_{h_i} , or h_i for short. The dependency tree also specifies a dependency relation labeled r_i between each (h_i, w_i) pair. To encode both syntactic and SRL information, we define a dependency tree t' , keeping all the h_i 's same as in t , but we modify relation r_i to be $r'_i := r_i^{\text{SYN}} - r_i^{(\text{D})} - r_i^{(\text{C})} - r_i^{(\text{R})}$, a concatenation of four labels: $r_i^{\text{SYN}} = r_i$ is the syntactic relation; $r_i^{(\text{D})}$ describes the SRL relation directly between the predicate h_i and the argument headed by w_i ; $r_i^{(\text{R})}$ specifies the reverse situation where w_i is the predicate and h_i the head of the argument; $r_i^{(\text{C})}$ encodes the parent-sharing pattern connecting the two predicates and is in the form of a tuple (a, b) , corresponding to the case where the SRL argument with label a for predicate h_i is an SRL argument labeled b with respect to predicate w_i .⁶ If there exist no such semantic relations, the component labels can be left unspecified, denoted as “_”.

In the example of Fig. 1, the joint label between *wanted* and *design* is `xcomp-ARG1-(ARG0,ARG0)-_`. We can break the joint label into four parts: “xcomp” describes the syntactic relation between the two tokens; “ARG1” indicates that the subtree *to design the bridge* is an argument labeled “ARG1” for predicate *wanted*; (ARG0,ARG0) establishes the argument sharing strategy that ARG0 *she* of *wanted* is an ARG0 for the predicate *design*; finally, “_” indicates there is no argument headed by *wanted* for the predicate *design*.

3.2 Back-and-Forth Conversion

The joint labels encode both syntactic and semantic relations, and it is straightforward to convert/recover the separate dependency and SRL annotations to/from the joint representations.

In the forward conversion (separate \rightarrow joint), we first extract the syntactic heads of all SRL arguments. Then we enumerate all predicate-argument pairs, and for each pair falling into one of the three most common patterns as listed in Table 1, we insert the SRL argument label in the corresponding slot in the joint label. For predicates sharing more than one argument, we observe that most cases are due to the two predicates sharing all their ARGM

⁶This assumes that the argument must also be an argument to the predicate h_i . In cases where there exists no such relation, we insert a dummy relation \emptyset that gets removed during post-processing between h_i and the argument, and the (C) label between h_i and w_i then becomes (\emptyset, b) .

relations, so we augment the (C) label with a binary indicator of whether or not to propagate all ARGM arguments. When the two predicates share more than one core argument, which occurs for around 2% of the argument-sharing predicates, we randomly select and record one of the shared arguments in $r_i^{(\text{C})}$. A more systematic assignment in such cases in future work may lead to further improvement.

As for the backward conversion (joint \rightarrow separate), the syntactic dependencies can be directly decoupled from the joint label, and we build the SRL relations in three steps: we first identify all the (D) and (R) dependency relations; then, with a top-down traversal of the tree, we identify the shared argument relations through (C) labels; finally, we rebuild the span boundaries using a rule-based approach. Top-down traversal is necessary to allow further propagation of arguments. It allows us to cover some of the less common cases through multiple argument sharings, e.g., the fourth example in Table 1. When a (C) label (a, b) is invalid⁷ in that the syntactic governor does not have an argument with label a , we simply ignore this (C) label. In reconstructing the span boundaries, we distinguish among different types of arguments. For (D)-type arguments, we directly take the entire subtrees dominated by the head words of the arguments. For (R)-type arguments, we adopt language-specific heuristics:⁸ in English, when the argument (syntactic head) is to the left of the predicate (syntactic child), as commonly happens in relative clause structures, we include all of the argument’s children subtrees to the left of the predicate; when the argument is to the right, which usually happens when the predicate is in participle form, we define the right subtree of the argument as its span. For (C)-type arguments, we reuse the span boundaries of the shared arguments.

Table 2 shows the oracle results of our back-and-forth conversion strategies on the training data. We take gold-standard syntactic and SRL annotations and convert them into joint-label representations. Then, we reconstruct the SRL relations through our backward conversion and measure span-based

⁷This should not happen in the oracle conversion but may occur in model predictions.

⁸The simple subtree approach does not apply to reconstructing (R)-type arguments since, by definition, the subtree of an (R)-type argument will contain its predicate, which contradicts data annotations. Our heuristics are designed to support a span-based evaluation, and span reconstruction can be omitted if one focuses on a dependency-based evaluation.

	P	R	F
English	99.7	98.3	99.0
Chinese	97.8	96.8	97.3

Table 2: Oracle back-and-forth conversion results on the training splits.

exact match metrics. Our procedures can faithfully reconstruct most of the SRL relations for both English and Chinese data.⁹ English sees a higher oracle score than Chinese. We attribute this result to the synchronization effort between the syntactic and SRL annotations during the evolution of English PropBank (Babko-Malaya et al., 2006; Bonial et al., 2017).

3.3 Models

Given that SRL can be reduced to a dependency parsing task with an extended label space, our model replicates and adapts that of a dependency parser. We follow the basic design of Dozat and Manning (2017), but instead of using LSTMs as input feature extractors, we opt for Transformer encoders (Vaswani et al., 2017), which have previously been shown to be successful in constituency parsing (Kitaev and Klein, 2018; Kitaev et al., 2019), dependency parsing (Kondratyuk and Straka, 2019), and SRL (Tan et al., 2018; Strubell et al., 2018). Next, we score all potential attachment pairs and dependency and SRL relations with the token-level representations through deep biaffine transformation (Dozat and Manning, 2017). After the dependency parsing decoding process, we retrieve the syntactic parse trees and SRL structures via our backward conversion algorithm.

Formally, we associate each token position with a context-sensitive representation by

$$[\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_n] = \text{Transformer}(w_0, w_1, \dots, w_n),$$

where w_0 denotes the root symbol for the dependency parse tree, and the inputs to the Transformer network are pretrained GloVe embeddings (Pennington et al., 2014). Alternatively, we can fine-tune a pre-trained contextualized feature extractor

⁹The English oracle F1 score is higher than the combined (D)+(C)+(R) occurrences of 98%. This is because (1) our method is precision-focused to minimize error propagation in prediction; recall loss of 1.7% is a direct reflection of the unaccounted less-frequent structures, and (2) many arguments, e.g., the fourth most frequent case in Table 1, can be reconstructed through the propagation of (C)-type labels.

such as BERT (Devlin et al., 2019):¹⁰

$$[\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_n] = \text{BERT}([\text{CLS}], w_1, \dots, w_n).$$

Next, the same representations \mathbf{x} serve as inputs to five different scoring modules, one for dependency attachment, one for syntactic labeling, and three modules for the newly-introduced SRL-related labels. All of the scoring modules use a deep biaffine (DBA) scoring function introduced by Dozat and Manning (2017) that is widely used in syntactic parsing (Dozat et al., 2017; Shi et al., 2017; Shi and Lee, 2018), semantic dependency parsing (Dozat and Manning, 2018) and SRL (Strubell et al., 2018). For an ordered pair of input vectors \mathbf{x}_i and \mathbf{x}_j , an r -dimensional DBA transforms each vector into a d -dimensional vector with multi-layer perceptrons and then outputs an r -dimensional vector $\mathbf{z}_{ij} = \text{DBA}(\mathbf{x}_i, \mathbf{x}_j)$, where

$$\mathbf{z}_{ijk} = [\text{MLP}^I(\mathbf{x}_i); 1]^\top \mathbf{U}_k [\text{MLP}^J(\mathbf{x}_j); 1],$$

$\mathbf{U} \in \mathbb{R}^{r \times (d+1) \times (d+1)}$, $[\cdot; 1]$ appends an element of 1 to the end of the vector, and MLP^I and MLP^J are two separate multi-layer perceptrons with non-linear activation functions. Following Dozat and Manning (2017), we model dependency attachment probabilities with a 1-dimensional DBA function:

$$P(h_j = i) \propto \exp(\text{DBA}^{\text{ATT}}(\mathbf{x}_i, \mathbf{x}_j)).$$

For syntactic labels from vocabulary V^{SYN} , we use a $|V^{\text{SYN}}|$ -dimensional DBA function:

$$P(r_j^{\text{SYN}} = V_t^{\text{SYN}}) \propto \exp(\text{DBA}_t^{\text{SYN}}(\mathbf{x}_{h_j}, \mathbf{x}_j)).$$

The three semantic label components $r^{(D)}$, $r^{(C)}$, and $r^{(R)}$ are modeled similarly to r^{SYN} .

All the above components are separately parameterized but they share the same feature extractor (Transformer or BERT). We train them with locally-normalized log-likelihood as objectives. During inference, we use a projective¹¹ maximum spanning tree algorithm (Eisner, 1996; Eisner and Satta, 1999) for unlabeled dependency parsing and then select the highest-scoring component label for each predicted attachment and each component.¹²

¹⁰In this case, we use the final-layer vector of the last subword unit for each word as its representation and the vector from the prepended [CLS] token for the root symbol.

¹¹The choice of a projective decoder is motivated by the empirical fact that both English and Chinese dependency trees are highly projective. One may consider a non-projective decoder when adapting to other languages.

¹²Structured and global inference that considers the interactions among all relation labels is a promising future direction.

English	P	R	F1
FitzGerald et al. (2015)	80.9	78.4	79.6
He et al. (2017)	81.7	81.6	81.7
He et al. (2018a)	–	–	82.1
Tan et al. (2018)	81.9	83.6	82.7
Ouchi et al. (2018)	84.4	81.7	83.0
Swayamdipta et al. (2018)	85.1	82.6	83.8
BIO-CRF baseline	83.4	83.6	83.5
Ours	83.3	83.0	83.2
<i>with pre-trained contextualized feature extractors</i>			
Peters et al. (2018)	–	–	84.6
He et al. (2018a)	–	–	85.5
Ouchi et al. (2018)	87.1	85.3	86.2
Li et al. (2019)	85.7	86.3	86.0
Li et al. (2020)	86.4	86.8	86.6
BIO-CRF baseline	86.4	87.1	86.7
Ours	85.9	85.6	85.8
Chinese	P	R	F1
BIO-CRF baseline	74.3	71.1	72.7
Ours	71.7	71.4	71.6
<i>with pre-trained contextualized feature extractors</i>			
BIO-CRF baseline	80.2	81.1	80.6
Ours	79.6	79.3	79.5

Table 3: Non-ensemble CoNLL 2012 test set results on both the English and the Chinese datasets.

4 Experiments

We evaluate on two datasets from OntoNotes 5.0 (Hovy et al., 2006) on English and Chinese. Similar to §2, we adopt the CoNLL 2012 dataset splits. To isolate the effects of predicate identification and following most existing work on SRL, we provide our models with pre-identified predicates. We report median performance across 5 runs of different random initialization for our models and our replicated reference models. Implementation details are provided in Appendix §A.

Main Results Table 3 reports the evaluation results. We compare our method with multiple state-of-the-art methods, including BIO-tagging (Tan et al., 2018; Peters et al., 2018), span-based (Ouchi et al., 2018; Li et al., 2019), semi-Markov CRF (Swayamdipta et al., 2018) and structured tuning (Li et al., 2020). We also implement a strong BIO-tagging model trained with a CRF loss as our baseline model (BIO-CRF), which has identical feature extractors as our proposed method.¹³ Results show that our models are competitive with the state-of-the-art models, even though our method reduces SRL to syntactic dependency parsing. Our models

¹³See, for example, He et al. (2017) for a BIO-tagging formulation of SRL.

Label	Count	BIO-CRF	Ours	+BERT
ARG0	11,444	90.8	90.4	91.8
ARG1	18,216	86.0	85.7	88.7
ARG2	6,429	80.1	78.4	83.7
ARGM-TMP	3,724	83.4	83.8	86.6
ARGM-ADV	2,089	65.0	63.9	66.5
ARGM-DIS	2,378	82.1	82.4	83.1
ARGM-MOD	1,844	97.8	98.0	97.8
Overall	53,906	83.5	83.0	85.9

Table 4: Per-type performance on the English dev set.

slightly underperform the BIO-CRF baseline models on English, and the gap is larger on Chinese.¹⁴ This can be attributed to the higher back-and-forth conversion loss on the Chinese data. We observe no significant difference in dependency parsing accuracy when training the Dozat and Manning’s (2017) parser alone versus jointly training with our SRL labels.

Additionally, our models make predictions for all predicates in a given sentence at the same time through $O(n)$ joint syntacto-semantic labels with identical features, while most other competitive methods either use different features extracted for different predicates (Tan et al., 2018; Ouchi et al., 2018; Swayamdipta et al., 2018), effectively requiring executing feature extraction multiple times, or require scoring for all $O(n^2)$ or $O(n^3)$ possible predicate-argument pairs¹⁵ (Strubell et al., 2018; Li et al., 2019). In our experiments, our models are 40% faster than the BIO-CRF baseline on average.

Results Broken Down by Argument Type Table 4 presents per-label F1 scores comparing our baseline model with our proposed method. Our method exhibits a similar overall performance to the baseline BIO-CRF model. Most of the difference is materialized on ARG2 and ARGM-ADV. Previous work in the literature finds that these labels are highly predicate-specific and known to be hard to predict (He et al., 2017). We further observe that pretrained feature extractors (BERT) tend to improve the most with respect to these two labels.

Effect of Different Components Table 5 summarizes the results when one or more components of our models are replaced by gold-standard labels.

¹⁴An anonymous reviewer hypothesizes that the accuracy loss may also be explained by the label sparsity introduced by our joint label scheme.

¹⁵ $O(n^3)$ results from considering all combinations of predicates, span start points and end points. In practice, however, Li et al. (2019) apply pruning to reduce number of candidates.

Setting	F1 Score	
	GloVe	BERT
All predicted	83.0	85.9
+ Gold syntax	88.9	90.0
+ Gold (D)	97.2	97.3
+ Gold (R) (C)	90.1	91.1
Upperbound (Oracle)	98.9	

Table 5: F1 performance on the English dev set if parts of the components are replaced by oracle, as an indicator of potential further gains from each component.

As expected, it is crucial to predict the syntactic trees correctly: failure to do so amounts to 35% or 29% of errors with or without pretrained feature extractors. Accuracy of (D)-type SRL relations has an even larger impact on the overall performance: it is responsible for half of the errors. This indicates that argument labeling is a harder sub-task than syntactic parsing. Further, we observe that the benefits of pretrained feature extractors mostly stem from improved accuracies of the syntactic component. Even with pretrained BERT features, semantic components remain challenging.

5 Related Work

SRL and syntax From the time the SRL task was first introduced (Gildea and Jurafsky, 2002; Gildea and Palmer, 2002; Màrquez et al., 2008; Palmer et al., 2010), syntax has been shown to be a critical factor in system performance. Most models use syntactically-derived features (Pradhan et al., 2005; Punyakanok et al., 2005; Swanson and Gordon, 2006; Johansson and Nugues, 2008; Toutanova et al., 2008; Xue, 2008; Zhao et al., 2009, *inter alia*) and syntax-based candidate pruning (Punyakanok et al., 2008). There have been many approaches for joint syntactic parsing and SRL models, including approximate search (Johansson, 2009) and dual decomposition (Lluís et al., 2013) to resolve feature dependencies, and synchronous parsing to simultaneously derive the (disjoint) syntactic and SRL structures (Henderson et al., 2008; Li et al., 2010; Henderson et al., 2013; Swayamdipta et al., 2016). In contrast, our work unifies the two representations into common structures.

Joint labels The idea of using joint labels for performing both syntactic and semantic tasks is similar to that of function parsing (Merlo and Musillo, 2005; Gabbard et al., 2006; Musillo and Merlo, 2006). Ge and Mooney (2005) use joint labels for

semantic parsing as well. Earlier approaches for SRL have considered joint syntactic and semantic labels. Due to lack of characterization of the common structures, most work either focuses on the subtask of argument identification (Yi and Palmer, 2005), predicts the set of all SRL labels for each argument and links them to predicates in a second stage (Merlo and Musillo, 2008), or models joint labels independently for each predicate (Samuelsson et al., 2008; Lluís and Màrquez, 2008; Morante et al., 2009; Rekaby Salama and Menzel, 2019). Instead, our work aims at extracting all predicate-argument structures from a sentence. Our joint label design is related to that of Qiu et al. (2016). They annotated a Chinese SRL corpus from scratch with a similar label scheme, while in this paper, we show that it is possible to extract such joint labels from existing data annotations.

Tree approximation In the task of semantic dependency parsing (Oepen et al., 2014), dependency structures are used to model more aspects of semantic phenomena than predicate-argument structures, and the representations are more general directed acyclic graphs. These graphs can be approximated by trees (Du et al., 2014; Schluter et al., 2014; Schluter, 2015) such that tree-based parsing algorithms become applicable. Unlike this line of research, we limit ourselves to the *given* syntactic trees, as opposed to finding the *optimal approximating* trees, and we focus on the close relations between syntax and SRL.

Dependency-based SRL Although predicate-argument structures are traditionally defined in constituency terms, dependency-based predicate-argument analysis (Hacioglu, 2004; Fundel et al., 2007) has been popularized through the CoNLL 2008 and 2009 shared tasks (Surdeanu et al., 2008; Hajič et al., 2009) and has been adopted by recent proposals of decompositional semantics (White et al., 2017). Choi and Palmer (2010) consider reconstructing constituency-based representations from dependency-based analysis. We confirm their findings that through a few heuristics, the reconstruction can be done faithfully.

Neural SRL The application of neural models to SRL motivates the question of whether modeling syntax is still necessary for the task (He et al., 2017). Similar to non-neural models, syntactic trees are used to construct features (Roth and Lapata, 2016; Kasai et al., 2019; Wang et al., 2019; Xia

et al., 2019b; Zhang et al., 2019) and to prune candidates (Täckström et al., 2015; He et al., 2018b, 2019). Alternatively, they are used to determine network structures (Marcheggiani and Titov, 2017; Li et al., 2018), including tree-LSTM, graph convolutional networks (Niepert et al., 2016) and syntax-aware LSTM (Qian et al., 2017). On the other hand, syntax-agnostic models (Collobert and Weston, 2007; Zhou and Xu, 2015; Cai et al., 2018; He et al., 2018a; Tan et al., 2018; Li et al., 2019) have shown competitive results. Our results contribute to the ongoing debate by adding further evidence that the two tasks are deeply-coupled. Future work may further explore how much syntactic knowledge has been implicitly obtained in the apparently syntax-agnostic models.

Multi-task learning Our models share neural representations across the syntactic and the SRL labelers. This is an instance of multi-task learning (MTL; Caruana, 1993, 1997). MTL has been successfully applied to SRL (Collobert and Weston, 2008; Collobert et al., 2011; Shi et al., 2016) in many state-of-the-art systems (Strubell et al., 2018; Swayamdipta et al., 2018; Cai and Lapata, 2019; Xia et al., 2019a). A potential future extension is to learn multiple syntactic (Søgaard and Goldberg, 2016) and semantic representations (Peng et al., 2017; Hershcovich et al., 2018) beyond dependency trees and PropBank-style SRL at the same time.

6 Conclusion

Linguistic theories assume a close relationship between the realization of semantic arguments and syntactic configurations. This work provides a detailed analysis of the syntactic structures of PropBank-style SRL and reveals that three common syntactic patterns account for 98% of annotated SRL relations for both English and Chinese data. Accordingly, we propose to reduce the task of SRL to syntactic dependency parsing through back-and-forth conversion to and from a joint label space. Experiments show that dependency parsers achieve competitive results on PropBank-style SRL with the state of the art.

This work shows promise of a syntactic treatment of SRL and opens up possibilities of applying existing dependency parsing techniques to SRL. We invite future research into further integration of syntactic methods into shallow semantic analysis in other languages and other formulations, such

as frame-semantic parsing, and other semantically-oriented tasks.

Acknowledgements

We thank the anonymous reviewers for their insightful reviews, and Sameer Bansal, Sam Brody, Prabhajan Kambadur, Lillian Lee, Daniel Preoțiuc-Pietro, Mats Rooth, Amanda Stent, and Chen-Tse Tsai for discussion and comments. Tianze Shi was supported by a Bloomberg Data Science Ph.D. fellowship.

References

- Daniel Albright, Arrick Lanfranchi, Anwen Fredriksen, William F. Styler, Colin Warner, Jena D. Hwang, Jinho D. Choi, Dmitriy Dligach, Rodney D. Nielsen, James Martin, Wayne Ward, Martha Palmer, and Guergana K Savova. 2013. [Towards comprehensive syntactic and semantic annotations of the clinical narrative](#). *Journal of the American Medical Informatics Association*, 20(5):922–930.
- Olga Babko-Malaya. 2005. [PropBank annotation guidelines](#). Technical report, University of Colorado at Boulder.
- Olga Babko-Malaya, Ann Bies, Ann Taylor, Szuting Yi, Martha Palmer, Mitch Marcus, Seth Kulick, and Libin Shen. 2006. [Issues in synchronizing the English treebank and PropBank](#). In *Proceedings of the Workshop on Frontiers in Linguistically Annotated Corpora 2006*, pages 70–77, Sydney, Australia.
- Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. [The Berkeley FrameNet project](#). In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1*, pages 86–90, Montreal, Canada.
- Mark C. Baker. 1985. [Incorporation, a Theory of Grammatical Function Changing](#). Thesis, Massachusetts Institute of Technology.
- Mark C. Baker. 1996. [On the structural positions of themes and goals](#). In Johan Rooryck and Laurie Zaring, editors, *Phrase Structure and the Lexicon*, Studies in Natural Language and Linguistic Theory, pages 7–34. Springer Netherlands, Dordrecht.
- Mark C. Baker. 1997. [Thematic roles and syntactic structure](#). In Liliane Haegeman, editor, *Elements of Grammar: Handbook in Generative Syntax*, Kluwer International Handbooks of Linguistics, pages 73–137. Springer Netherlands, Dordrecht.
- Marzieh Bazrafshan and Daniel Gildea. 2013. [Semantic roles for string to tree machine translation](#). In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 419–423, Sofia, Bulgaria.

- Claire Bonial, Julia Bonn, Kathryn Conger, Jena Hwang, Martha Palmer, and Nicholas Reese. 2015. [English PropBank annotation guidelines](#). Technical report, Center for Computational Language and Education Research Institute of Cognitive Science, University of Colorado at Boulder.
- Claire Bonial, Kathryn Conger, Jena D. Hwang, Aous Mansouri, Yahya Aseri, Julia Bonn, Timothy O’Gorman, and Martha Palmer. 2017. [Current directions in English and Arabic PropBank](#). In Nancy Ide and James Pustejovsky, editors, *Handbook of Linguistic Annotation*, pages 737–769. Springer Netherlands, Dordrecht.
- John Bowers. 2010. *Arguments as Relations*. The MIT Press, Cambridge, Massachusetts, USA.
- Jiaxun Cai, Shexia He, Zuchao Li, and Hai Zhao. 2018. [A full end-to-end semantic role labeler, syntactic-agnostic over syntactic-aware?](#) In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2753–2765, Santa Fe, New Mexico, USA.
- Rui Cai and Mirella Lapata. 2019. [Syntax-aware semantic role labeling without parsing](#). *Transactions of the Association for Computational Linguistics*, 7:343–356.
- Xavier Carreras and Lluís Màrquez. 2005. [Introduction to the CoNLL-2005 shared task: Semantic role labeling](#). In *Proceedings of the Ninth Conference on Computational Natural Language Learning*, pages 152–164, Ann Arbor, Michigan, USA.
- Rich Caruana. 1993. [Multitask learning: A knowledge-based source of inductive bias](#). In *Proceedings of the Tenth International Conference on International Conference on Machine Learning*, pages 41–48, Amherst, Massachusetts, USA.
- Rich Caruana. 1997. [Multitask learning](#). *Machine Learning*, 28(1):41–75.
- Pi-Chuan Chang, Huihsin Tseng, Dan Jurafsky, and Christopher D. Manning. 2009. [Discriminative reordering with Chinese grammatical relations features](#). In *Proceedings of the Third Workshop on Syntax and Structure in Statistical Translation*, pages 51–59, Boulder, Colorado, USA.
- Jinho Choi and Martha Palmer. 2010. [Retrieving correct semantic boundaries in dependency structure](#). In *Proceedings of the Fourth Linguistic Annotation Workshop*, pages 91–99, Uppsala, Sweden.
- Janara Christensen, Mausam, Stephen Soderland, and Oren Etzioni. 2010. [Semantic role labeling for open information extraction](#). In *Proceedings of the First International Workshop on Formalisms and Methodology for Learning by Reading*, pages 52–60, Los Angeles, California, USA.
- Ronan Collobert and Jason Weston. 2007. [Fast semantic extraction using a novel neural network architecture](#). In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 560–567, Prague, Czech Republic.
- Ronan Collobert and Jason Weston. 2008. [A unified architecture for natural language processing: Deep neural networks with multitask learning](#). In *Proceedings of the 25th International Conference on Machine Learning*, pages 160–167, Helsinki, Finland.
- Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. [Natural language processing \(almost\) from scratch](#). *The Journal of Machine Learning Research*, 12:2493–2537.
- Marie-Catherine de Marneffe, Bill MacCartney, and Christopher D. Manning. 2006. [Generating typed dependency parses from phrase structure parses](#). In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*, pages 449–454, Genoa, Italy.
- Marie-Catherine de Marneffe and Christopher D. Manning. 2008. [Stanford typed dependencies manual](#). Technical report, Stanford University.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional Transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota.
- Timothy Dozat and Christopher D. Manning. 2017. [Deep biaffine attention for neural dependency parsing](#). In *Proceedings of the 5th International Conference on Learning Representations*, Toulon, France.
- Timothy Dozat and Christopher D. Manning. 2018. [Simpler but more accurate semantic dependency parsing](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 484–490, Melbourne, Australia.
- Timothy Dozat, Peng Qi, and Christopher D. Manning. 2017. [Stanford’s graph-based neural dependency parser at the CoNLL 2017 shared task](#). In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 20–30, Vancouver, Canada.
- Yantao Du, Fan Zhang, Weiwei Sun, and Xiaojun Wan. 2014. [Peking: Profiling syntactic tree parsing techniques for semantic graph parsing](#). In *Proceedings of the 8th International Workshop on Semantic Evaluation*, pages 459–464, Dublin, Ireland.
- Jason Eisner and Giorgio Satta. 1999. [Efficient parsing for bilexical context-free grammars and head automaton grammars](#). In *Proceedings of the 37th Annual Meeting of the Association for Computational*

- Linguistics*, pages 457–464, College Park, Maryland, USA.
- Jason M. Eisner. 1996. [Three new probabilistic models for dependency parsing: An exploration](#). In *Proceedings of the 16th International Conference on Computational Linguistics*, pages 340–345, Copenhagen, Denmark.
- Charles J. Fillmore. 1976. [Frame semantics and the nature of language](#). *Annals of the New York Academy of Sciences*, 280(1):20–32.
- Charles J. Fillmore, Christopher R. Johnson, and Miriam R. L. Petruck. 2003. [Background to FrameNet](#). *International Journal of Lexicography*, 16(3):235–250.
- Nicholas FitzGerald, Oscar Täckström, Kuzman Ganchev, and Dipanjan Das. 2015. [Semantic role labeling with neural network factors](#). In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 960–970, Lisbon, Portugal.
- Katrin Fundel, Robert Küffner, and Ralf Zimmer. 2007. [RelEx—Relation extraction using dependency parse trees](#). *Bioinformatics*, 23(3):365–371.
- Ryan Gabbard, Seth Kulick, and Mitchell Marcus. 2006. [Fully parsing the Penn Treebank](#). In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pages 184–191, New York City, New York, USA.
- Ruifang Ge and Raymond Mooney. 2005. [A statistical semantic parser that integrates syntax and semantics](#). In *Proceedings of the Ninth Conference on Computational Natural Language Learning*, pages 9–16, Ann Arbor, Michigan, USA.
- Daniel Gildea and Julia Hockenmaier. 2003. [Identifying semantic roles using combinatory categorial grammar](#). In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, pages 57–64, Sapporo, Japan.
- Daniel Gildea and Daniel Jurafsky. 2002. [Automatic labeling of semantic roles](#). *Computational Linguistics*, 28(3):245–288.
- Daniel Gildea and Martha Palmer. 2002. [The necessity of parsing for predicate argument recognition](#). In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 239–246, Philadelphia, Pennsylvania, USA.
- Jeffrey Steven Gruber. 1965. *Studies in Lexical Relations*. Ph.D. thesis, Massachusetts Institute of Technology.
- Kadri Hacioglu. 2004. [Semantic role labeling using dependency trees](#). In *Proceedings of the 20th International Conference on Computational Linguistics*, pages 1273–1276, Geneva, Switzerland.
- Jan Hajič, Massimiliano Ciaramita, Richard Johansson, Daisuke Kawahara, Maria Antònia Martí, Lluís Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Štěpánek, Pavel Straňák, Mihai Surdeanu, Nianwen Xue, and Yi Zhang. 2009. [The CoNLL-2009 shared task: Syntactic and semantic dependencies in multiple languages](#). In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task*, pages 1–18, Boulder, Colorado, USA.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. [Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification](#). In *Proceedings of the 2015 IEEE International Conference on Computer Vision*, pages 1026–1034, Santiago, Chile.
- Luheng He, Kenton Lee, Omer Levy, and Luke Zettlemoyer. 2018a. [Jointly predicting predicates and arguments in neural semantic role labeling](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 364–369, Melbourne, Australia.
- Luheng He, Kenton Lee, Mike Lewis, and Luke Zettlemoyer. 2017. [Deep semantic role labeling: What works and what’s next](#). In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 473–483, Vancouver, Canada.
- Shexia He, Zuchao Li, and Hai Zhao. 2019. [Syntax-aware multilingual semantic role labeling](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 5353–5362, Hong Kong, China.
- Shexia He, Zuchao Li, Hai Zhao, and Hongxiao Bai. 2018b. [Syntax for semantic role labeling, to be, or not to be](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2061–2071, Melbourne, Australia.
- James Henderson, Paola Merlo, Gabriele Musillo, and Ivan Titov. 2008. [A latent variable model of synchronous parsing for syntactic and semantic dependencies](#). In *Proceedings of the Twelfth Conference on Computational Natural Language Learning*, pages 178–182, Manchester, UK.
- James Henderson, Paola Merlo, Ivan Titov, and Gabriele Musillo. 2013. [Multilingual joint parsing of syntactic and semantic dependencies with a latent variable model](#). *Computational Linguistics*, 39(4):949–998.
- Daniel Hershcovich, Omri Abend, and Ari Rappoport. 2018. [Multitask parsing across semantic representations](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 373–385, Melbourne, Australia.

- Daniel Hershcovich, Omri Abend, and Ari Rappoport. 2019. [Content differences in syntactic and semantic representation](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 478–488, Minneapolis, Minnesota.
- Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. 2006. [OntoNotes: The 90% solution](#). In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers*, pages 57–60, New York City, New York, USA.
- C.-T. James Huang, Y.-H. Audrey Li, and Yafei Li. 2008. *The Syntax of Chinese*. Cambridge Syntax Guides. Cambridge University Press, Cambridge, UK.
- Ray Jackendoff. 1972. *Semantic Interpretation in Generative Grammar*. Current Studies in Linguistics. The MIT Press, Cambridge, Massachusetts, USA.
- Ray Jackendoff. 1992. *Semantic Structures*. The MIT Press, Cambridge, Massachusetts, USA.
- Richard Johansson. 2009. [Statistical bistratal dependency parsing](#). In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 561–569, Singapore.
- Richard Johansson and Pierre Nugues. 2008. The effect of syntactic representation on semantic role labeling. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 393–400, Manchester, UK.
- Aravind K. Joshi, Leon S. Levy, and Masako Takahashi. 1975. [Tree adjunct grammars](#). *Journal of Computer and System Sciences*, 10(1):136–163.
- Laura Kallmeyer. 2010. *Parsing beyond Context-Free Grammars*. Cognitive Technologies. Springer-Verlag, Berlin Heidelberg.
- Jungo Kasai, Dan Friedman, Robert Frank, Dragomir Radev, and Owen Rambow. 2019. [Syntax-aware neural semantic role labeling with supertags](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 701–709, Minneapolis, Minnesota, USA.
- Atif Khan, Naomie Salim, and Yogan Jaya Kumar. 2015. [A framework for multi-document abstractive summarization based on semantic role labelling](#). *Applied Soft Computing*, 30:737–747.
- Daniel Khashabi, Tushar Khot, Ashish Sabharwal, and Dan Roth. 2018. [Question answering as global reasoning over semantic abstractions](#). In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, pages 1905–1914, New Orleans, Louisiana, USA.
- Diederik P. Kingma and Jimmy Ba. 2015. [Adam: A method for stochastic optimization](#). In *Proceedings of the 3rd International Conference on Learning Representations*, San Diego, California, USA.
- Paul Kingsbury and Martha Palmer. 2002. [From Tree-Bank to PropBank](#). In *Proceedings of the Third International Conference on Language Resources and Evaluation*, pages 1989–1993, Las Palmas, Canary Islands, Spain.
- Nikita Kitaev, Steven Cao, and Dan Klein. 2019. [Multilingual constituency parsing with self-attention and pre-training](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3499–3505, Florence, Italy.
- Nikita Kitaev and Dan Klein. 2018. [Constituency parsing with a self-attentive encoder](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2676–2686, Melbourne, Australia.
- Dan Kondratyuk and Milan Straka. 2019. [75 languages, 1 model: Parsing Universal Dependencies universally](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 2779–2795, Hong Kong, China.
- Ioannis Konstas, Frank Keller, Vera Demberg, and Mirella Lapata. 2014. [Incremental semantic role labeling with tree adjoining grammar](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 301–312, Doha, Qatar.
- Sandra Kübler, Ryan McDonald, and Joakim Nivre. 2008. *Dependency Parsing*, volume 2 of *Synthesis Lectures on Human Language Technologies*. Morgan and Claypool.
- Jonathan K. Kummerfeld and Dan Klein. 2017. [Parsing with traces: An \$O\(n^4\)\$ algorithm and a structural representation](#). *Transactions of the Association for Computational Linguistics*, 5:441–454.
- Beth Levin. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press, Chicago, Illinois, USA.
- Beth Levin and Malka Rappaport Hovav. 2005. *Argument Realization*. Cambridge University Press, Cambridge, UK.
- Mike Lewis, Luheng He, and Luke Zettlemoyer. 2015. [Joint A* CCG parsing and semantic role labelling](#). In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1444–1454, Lisbon, Portugal.
- Charles N. Li and Sandra A. Thompson. 1989. *Mandarin Chinese: A functional reference grammar*. University of California Press, Berkeley, California, USA.

- Junhui Li, Guodong Zhou, and Hwee Tou Ng. 2010. [Joint syntactic and semantic parsing of Chinese](#). In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1108–1117, Uppsala, Sweden.
- Tao Li, Parth Anand Jawale, Martha Palmer, and Vivek Srikumar. 2020. [Structured tuning for semantic role labeling](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8402–8412, Online.
- Zuchao Li, Shexia He, Jiaxun Cai, Zhuosheng Zhang, Hai Zhao, Gongshen Liu, Linlin Li, and Luo Si. 2018. [A unified syntax-aware framework for semantic role labeling](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2401–2411, Brussels, Belgium.
- Zuchao Li, Shexia He, Hai Zhao, Yiqing Zhang, Zhuosheng Zhang, Xi Zhou, and Xiang Zhou. 2019. [Dependency or span, end-to-end uniform semantic role labeling](#). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6730–6737.
- Ding Liu and Daniel Gildea. 2010. [Semantic role features for machine translation](#). In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 716–724, Beijing, China.
- Yudong Liu. 2009. *Semantic Role Labeling Using Lexicalized Tree Adjoining Grammars*. Thesis, Simon Fraser University.
- Yudong Liu and Anoop Sarkar. 2009. [Exploration of the LTAG-spinal formalism and treebank for semantic role labeling](#). In *Proceedings of the 2009 Workshop on Grammar Engineering Across Frameworks*, pages 1–9, Suntec, Singapore.
- Xavier Lluís, Xavier Carreras, and Lluís Màrquez. 2013. [Joint arc-factored parsing of syntactic and semantic dependencies](#). *Transactions of the Association for Computational Linguistics*, 1:219–230.
- Xavier Lluís and Lluís Màrquez. 2008. [A joint model for parsing syntactic and semantic dependencies](#). In *Proceedings of the Twelfth Conference on Computational Natural Language Learning*, pages 188–192, Manchester, UK.
- Diego Marcheggiani and Ivan Titov. 2017. [Encoding sentences with graph convolutional networks for semantic role labeling](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1506–1515, Copenhagen, Denmark.
- Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. [Building a large annotated corpus of English: The Penn Treebank](#). *Computational Linguistics*, 19(2):313–330.
- Lluís Màrquez, Xavier Carreras, Kenneth C. Litkowski, and Suzanne Stevenson. 2008. [Semantic role labeling: An introduction to the special issue](#). *Computational Linguistics*, 34(2):145–159.
- Paola Merlo and Gabriele Musillo. 2005. [Accurate function parsing](#). In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 620–627, Vancouver, Canada.
- Paola Merlo and Gabriele Musillo. 2008. [Semantic parsing for high-precision semantic role labelling](#). In *Proceedings of the Twelfth Conference on Computational Natural Language Learning*, pages 1–8, Manchester, UK.
- Roser Morante, Vincent Van Asch, and Antal van den Bosch. 2009. [Dependency parsing and semantic role labeling as a single task](#). In *Proceedings of the International Conference Recent Advances in Natural Language Processing*, pages 275–280, Borovets, Bulgaria.
- Gabriele Musillo and Paola Merlo. 2006. [Accurate parsing of the Proposition Bank](#). In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers*, pages 101–104, New York City, New York, USA.
- Mathias Niepert, Mohamed Ahmed, and Konstantin Kutzkov. 2016. [Learning convolutional neural networks for graphs](#). In *Proceedings of the 33rd International Conference on International Conference on Machine Learning*, pages 2014–2023, New York City, New York, USA.
- Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, 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 *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, pages 1659–1666, Portorož, Slovenia.
- 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 *Proceedings of the 8th International Workshop on Semantic Evaluation*, pages 63–72, Dublin, Ireland.
- Hiroki Ouchi, Hiroyuki Shindo, and Yuji Matsumoto. 2018. [A span selection model for semantic role labeling](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1630–1642, Brussels, Belgium.
- 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*, volume 3 of *Synthesis Lectures on Human Language Technologies*. Morgan and Claypool.

- Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. 2013. [On the difficulty of training recurrent neural networks](#). In *Proceedings of the 30th International Conference on Machine Learning - Volume 28*, pages 1310–1318, Atlanta, Georgia, USA.
- Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. [Automatic differentiation in PyTorch](#). In *Proceedings of the Workshop on the Future of Gradient-based Machine Learning Software*, Long Beach, California, USA.
- Hao Peng, Sam Thomson, and Noah A. Smith. 2017. [Deep multitask learning for semantic dependency parsing](#). In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2037–2048, Vancouver, Canada.
- Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. [GloVe: Global vectors for word representation](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1532–1543, Doha, Qatar.
- David M. Perlmutter and Paul M. Postal. 1984. The 1-advancement exclusiveness law. In David M. Perlmutter and Carol G. Rosen, editors, *Studies in Relational Grammar*, volume 2, pages 81–125. The University of Chicago Press, Chicago, Illinois, USA.
- Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. [Deep contextualized word representations](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana, USA.
- Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Olga Uryupina, and Yuchen Zhang. 2012. [CoNLL-2012 shared task: Modeling multilingual unrestricted coreference in OntoNotes](#). In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning - Shared Task*, pages 1–40, Jeju Island, Korea.
- Sameer Pradhan, Wayne Ward, Kadri Hacioglu, James Martin, and Daniel Jurafsky. 2005. [Semantic role labeling using different syntactic views](#). In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pages 581–588, Ann Arbor, Michigan, USA.
- Vasin Punyakanok, Dan Roth, and Wen-tau Yih. 2005. [The necessity of syntactic parsing for semantic role labeling](#). In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, pages 1117–1123, Edinburgh, UK.
- Vasin Punyakanok, Dan Roth, and Wen-tau Yih. 2008. [The importance of syntactic parsing and inference in semantic role labeling](#). *Computational Linguistics*, 34(2):257–287.
- Feng Qian, Lei Sha, Baobao Chang, Lu-chen Liu, and Ming Zhang. 2017. [Syntax aware LSTM model for semantic role labeling](#). In *Proceedings of the 2nd Workshop on Structured Prediction for Natural Language Processing*, pages 27–32, Copenhagen, Denmark.
- Likun Qiu, Yue Zhang, and Meishan Zhang. 2016. [Dependency tree representations of predicate-argument structures](#). In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 2645–2651, Phoenix, Arizona, USA.
- Siva Reddy, Oscar Täckström, Slav Petrov, Mark Steedman, and Mirella Lapata. 2017. [Universal semantic parsing](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 89–101, Copenhagen, Denmark.
- Amr Rekaby Salama and Wolfgang Menzel. 2019. [Joint labeling of syntactic function and semantic role using probabilistic finite state automata](#). In *Proceedings of the 2018 Intelligent Systems Conference: Intelligent Systems and Applications*, pages 588–605, London, UK.
- Michael Roth and Mirella Lapata. 2016. [Neural semantic role labeling with dependency path embeddings](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1192–1202, Berlin, Germany.
- Yvonne Samuelsson, Oscar Täckström, Sumithra Velupillai, Johan Eklund, Mark Fishel, and Markus Saers. 2008. [Mixing and blending syntactic and semantic dependencies](#). In *Proceedings of the Twelfth Conference on Computational Natural Language Learning*, pages 248–252, Manchester, UK.
- Natalie Schluter. 2015. [Maximising spanning subtree scores for parsing tree approximations of semantic dependency digraphs](#). In *Proceedings of the 14th International Conference on Parsing Technologies*, pages 92–96, Bilbao, Spain.
- Natalie Schluter, Anders Søgaard, Jakob Elming, Dirk Hovy, Barbara Plank, Héctor Martínez Alonso, Anders Johanssen, and Sigrid Klerke. 2014. [Copenhagen-Malmö: Tree approximations of semantic parsing problems](#). In *Proceedings of the 8th International Workshop on Semantic Evaluation*, pages 213–217, Dublin, Ireland.
- Gerold Schneider. 2008. *Hybrid Long-Distance Functional Dependency Parsing*. Ph.D. thesis, University of Zurich.
- Sebastian Schuster and Christopher D. Manning. 2016. [Enhanced English Universal Dependencies: An improved representation for natural language understanding tasks](#). In *Proceedings of the Tenth Interna-*

- tional Conference on Language Resources and Evaluation, pages 2371–2378, Portorož, Slovenia.
- Dan Shen and Mirella Lapata. 2007. [Using semantic roles to improve question answering](#). In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 12–21, Prague, Czech Republic.
- Peng Shi, Zhiyang Teng, and Yue Zhang. 2016. [Exploiting mutual benefits between syntax and semantic roles using neural network](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 968–974, Austin, Texas, USA.
- Tianze Shi, Liang Huang, and Lillian Lee. 2017. [Fast\(er\) exact decoding and global training for transition-based dependency parsing via a minimal feature set](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 12–23, Copenhagen, Denmark.
- Tianze Shi and Lillian Lee. 2018. [Valency-augmented dependency parsing](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1277–1291, Brussels, Belgium.
- Natalia Silveira, Timothy Dozat, Marie-Catherine de Marneffe, Samuel Bowman, Miriam Connor, John Bauer, and Chris Manning. 2014. [A gold standard dependency corpus for English](#). In *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, pages 2897–2904, Reykjavik, Iceland.
- Anders Søgaard and Yoav Goldberg. 2016. [Deep multi-task learning with low level tasks supervised at lower layers](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 231–235, Berlin, Germany.
- Mark Steedman. 2000. *The Syntactic Process*. The MIT Press, Cambridge, Massachusetts, USA.
- Emma Strubell, Patrick Verga, Daniel Andor, David Weiss, and Andrew McCallum. 2018. [Linguistically-informed self-attention for semantic role labeling](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 5027–5038, Brussels, Belgium.
- Weiwei Sun, Zhifang Sui, and Haifeng Wang. 2008. [Prediction of maximal projection for semantic role labeling](#). In *Proceedings of the 22nd International Conference on Computational Linguistics*, pages 833–840, Manchester, UK.
- Mihai Surdeanu, Richard Johansson, Adam Meyers, Lluís Màrquez, and Joakim Nivre. 2008. [The CoNLL 2008 shared task on joint parsing of syntactic and semantic dependencies](#). In *Proceedings of the Twelfth Conference on Computational Natural Language Learning*, pages 159–177, Manchester, UK.
- Reid Swanson and Andrew S. Gordon. 2006. [A comparison of alternative parse tree paths for labeling semantic roles](#). In *Proceedings of the Joint Conference of the International Committee on Computational Linguistics and the Association for Computational Linguistics*, pages 811–818, Sydney, Australia.
- Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer, and Noah A. Smith. 2016. [Greedy, joint syntactic-semantic parsing with stack LSTMs](#). In *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning*, pages 187–197, Berlin, Germany.
- Swabha Swayamdipta, Sam Thomson, Kenton Lee, Luke Zettlemoyer, Chris Dyer, and Noah A. Smith. 2018. [Syntactic scaffolds for semantic structures](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3772–3782, Brussels, Belgium.
- Ida Szubert, Adam Lopez, and Nathan Schneider. 2018. [A structured syntax-semantics interface for English-AMR alignment](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1169–1180, New Orleans, Louisiana, USA.
- Oscar Täckström, Kuzman Ganchev, and Dipanjan Das. 2015. [Efficient inference and structured learning for semantic role labeling](#). *Transactions of the Association for Computational Linguistics*, 3:29–41.
- Zhixing Tan, Mingxuan Wang, Jun Xie, Yidong Chen, and Xiaodong Shi. 2018. [Deep semantic role labeling with self-attention](#). In *Proceedings of The Thirty-Second AAAI Conference on Artificial Intelligence*, pages 4926–4936, New Orleans, Louisiana.
- Kristina Toutanova, Aria Haghighi, and Christopher D. Manning. 2008. [A global joint model for semantic role labeling](#). *Computational Linguistics*, 34(2):161–191.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. [Attention is all you need](#). In *Advances in Neural Information Processing Systems 30*, pages 5998–6008, Long Beach, CA, USA.
- Yufei Wang, Mark Johnson, Stephen Wan, Yifang Sun, and Wei Wang. 2019. [How to best use syntax in semantic role labelling](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5338–5343, Florence, Italy.
- Aaron Steven White, Kyle Rawlins, and Benjamin Van Durme. 2017. [The semantic proto-role linking model](#). In *Proceedings of the 15th Conference of the*

- European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 92–98, Valencia, Spain.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, and Jamie Brew. 2019. [HuggingFace’s Transformers: State-of-the-art natural language processing](#). *arXiv e-prints*. 1910.03771v1.
- Qingrong Xia, Zhenghua Li, and Min Zhang. 2019a. [A syntax-aware multi-task learning framework for Chinese semantic role labeling](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 5385–5395, Hong Kong, China.
- Qingrong Xia, Zhenghua Li, Min Zhang, Meishan Zhang, Guohong Fu, Rui Wang, and Luo Si. 2019b. [Syntax-aware neural semantic role labeling](#). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 7305–7313, Honolulu, Hawaii, USA.
- Deyi Xiong, Min Zhang, and Haizhou Li. 2012. [Modeling the translation of predicate-argument structure for SMT](#). In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 902–911, Jeju Island, Korea.
- 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(2):207–238.
- Nianwen Xue. 2008. [Labeling Chinese predicates with semantic roles](#). *Computational Linguistics*, 34(2):225–255.
- Nianwen Xue and Martha Palmer. 2003. [Annotating the propositions in the Penn Chinese Treebank](#). In *Proceedings of the Second SIGHAN Workshop on Chinese Language Processing*, pages 47–54, Sapporo, Japan.
- Jie Yang, Shuailong Liang, and Yue Zhang. 2018. [Design challenges and misconceptions in neural sequence labeling](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3879–3889, Santa Fe, New Mexico, USA.
- Szu-ting Yi and Martha Palmer. 2005. [The integration of syntactic parsing and semantic role labeling](#). In *Proceedings of the Ninth Conference on Computational Natural Language Learning*, pages 237–240, Ann Arbor, Michigan, USA.
- Yue Zhang, Rui Wang, and Luo Si. 2019. [Syntax-enhanced self-attention-based semantic role labeling](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 616–626, Hong Kong, China.
- Hai Zhao, Wenliang Chen, Jun’ichi Kazama, Kiyotaka Uchimoto, and Kentaro Torisawa. 2009. [Multilingual dependency learning: Exploiting rich features for tagging syntactic and semantic dependencies](#). In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task*, pages 61–66, Boulder, Colorado, USA.
- Jie Zhou and Wei Xu. 2015. [End-to-end learning of semantic role labeling using recurrent neural networks](#). In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1127–1137, Beijing, China.

A Implementation Details

We use two types of encoders in our models: randomly-initialized Transformer (Vaswani et al., 2017) networks and pre-trained BERT (Devlin et al., 2019). For Transformer networks, the input representations are concatenations of 100-dimensional randomly-initialized word embeddings, 100-dimensional pre-trained GloVe (Pennington et al., 2014) embeddings, 16-dimensional predicate indicator embeddings and 128-dimensional positional embeddings. The Transformer networks have 8 self-attention and feed-forward layers. Each self-attention layer has 8 attention heads, and each feed-forward layer has a dimensionality of 2048. For BERT models, we fine-tune the pretrained BASE model by Devlin et al. (2019) and Wolf et al. (2019).

The decoders consist of an unlabeled attachment scorer and several labeling components for the syntactic dependencies and SRL relations. The design and hyperparameters follow that of Dozat and Manning (2017). The biaffine scoring function for the unlabeled attachment scorer has a dimensionality of 400, while each labeling component has 100 dimensions. For our baseline, we build on top of Tan et al.’s (2018) BIO tagging model and further add a CRF-based decoding layer following Yang et al. (2018). Contextualized representation at each token’s position is passed through a multi-layer perceptron with one hidden layer consisting of 256 hidden units and PReLU (He et al., 2015) activation function to obtain the scores for each tag.

64 training instances (16 when using BERT) are grouped into a minibatch, and the gradients are clipped (Pascanu et al., 2013) at 5.0. We use Adam (Kingma and Ba, 2015) optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 1 \times 10^{-8}$. When using GloVe embeddings and Transformers, we set the learning rate to be 1×10^{-4} ; when fine-tuning BERT, the learning rate is lowered to 1×10^{-5} . Learning rates are multiplied by 0.1 once the development performance stops increasing for 5 epochs. All the models are trained until the learning rates are lowered three times and the performance plateaus on the development sets. Our implementation is based on PyTorch (Paszke et al., 2017).

On a single V100 GPU, the baseline BIO-CRF model parses 96.4 sentences/sec and our proposed model processes at 159.1 sentences/sec on average.

Throughout our experiments, all the hyperparameters are taken directly from relevant suggestions

	Trained with	P	R	F
Oracle	Gold	99.7	98.2	98.9
	Predicted	99.6	93.2	96.3
GloVe	Gold	83.2	82.9	83.0
	Predicted	84.7	80.6	82.6
BERT	Gold	86.2	85.5	85.8
	Predicted	86.9	83.2	85.1

Table 6: English back-and-forth oracle and dev set results using *gold-standard* dependency trees versus *predicted* trees as *training* data.

in previous literature (Dozat and Manning, 2017; Tan et al., 2018; Kitaev et al., 2019) without tuning. An extensive hyperparameter search may lead to further accuracy improvements.

B Additional Model Analysis

B.1 Training without Gold Syntactic Trees

Our method leverages the gold-standard dependency trees in the *training* data to design high-fidelity back-and-forth conversion algorithms. Table 6 considers a scenario where we do not have access to such gold trees during *training*: we jack-knife the data into 8 folds, train parsers using 7 folds and predict trees on the remaining fold. Our models show similar F1 scores under this condition as that of using gold trees, while the recall is traded for precision since our conversion method is precision-focused.

This is not a realistic scenario given that existing PropBank-style SRL annotations are all based on syntax, so as a matter of practice we always have access to gold trees during *training*. Nonetheless, these experiments point to the viability of using predicted trees in practice without incurring a significant loss in F1 scores.

B.2 Accuracies by SRL Relation Types

In Table 7, we break down the accuracies by the syntactic patterns of the SRL relations. Compared with our baseline, a replication of Tan et al. (2018), our models achieves higher or competitive results on (D)-type and (R)-type SRL relations. These two types establish a direct or reverse semantic relation with respect to the syntactic structure. In contrast, the (C)-type relations require accurate predictions of sibling relations as well as at least two SRL-related labels and are thus more prone to error propagation. We hypothesize that global scoring of the dependency structures can alleviate

Type	English		Chinese	
	Baseline	Ours	Baseline	Ours
(D)	88.1	88.1	82.8	83.4
(C)	80.3	76.6	53.3	50.0
(R)	78.7	79.3	46.3	46.8
Overall	83.5	83.0	74.1	72.9

Table 7: Per-pattern F1 scores on the dev sets.

	GloVe			BERT		
	Base.	Ours	Δ	Base.	Ours	Δ
10%	71.3	72.4	+1.1	82.3	81.5	-0.8
3%	59.9	62.0	+2.1	78.2	77.6	-0.6
1%	47.5	48.7	+1.2	69.3	73.2	+3.9

Table 8: English dev F1 scores, when trained with different percentages of the training data.

this issue, and we leave that to future work.

B.3 Learning Curve

In Table 8, we train the models with varying amounts of training data. With GloVe embeddings, our models exhibit higher performance when training data is limited, as compared with the corresponding baselines. When the pre-trained BERT feature extractor is used, both the baseline and our model require far less data to reach similar levels of performance. Our model shows significant improvement when the amount of training data is extremely limited (1%), and the baseline edges out for the other two settings (3% and 10%).

C Additional English Data Analysis

Among the three common patterns, (D)-type SRL relations are the most frequent and easiest to understand. In this section, we provide additional examples to shed light on (C)-type and (R)-type relations. We also show some sentences with more complex syntactic phenomena than what can be handled by our joint-label scheme. In all the examples, we boldface the predicates, underline the head words of the arguments, and highlight only the shortest dependency paths connecting them.

C.1 (C)-Type Relations

The (C)-type relations are most frequently used in ARG0 (55%) and ARG1 (19%) relations, in contrast to (D)-type relations, where the percentages are much lower (34% and 17% respectively). This

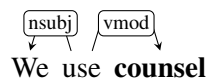
can be explained by the fact that a lot of (C)-type relations are used in control and raising verb constructions. A second major construction associated with (C)-type relations is conjunction, which shares either core or peripheral arguments among the conjuncts. The most common dependency relation labels connecting the common parents and the predicates are: “xcomp” (39%), “conj” (37%), “vmod” (9%), and “dep” (6%).

“xcomp” signifies control/raising structures. Popular common parent words (the control/raising verbs) include “want”, “expect”, “continue”, “begin”, etc.

“conj” represents a coordination structure. Since the first conjunct is a syntactic head of other conjuncts in Stanford Dependencies, any shared argument will result in a (C)-type relation.

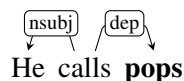
“vmod” denotes non-finite verbal modifiers whose missing subjects can often be found in the main clauses. For example:

- (3) We use all wisdom to **counsel** every person.



A lot of problematic instances of “dep” can be attributed to failures of constituency-to-dependency conversion, where it should have been recognized as a relation corresponding to another construction. For example:

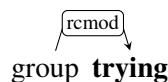
- (4) He calls ... and **pops** in every once in a while.



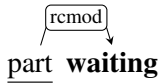
C.2 (R)-Type Relations

(R)-type relations are frequently used in relative clauses, as “rcmod” accounts for 47% of the syntactic relations connecting the predicates and the arguments. For examples:

- (5) ... another group that is always **trying** to ...

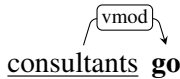


- (6) ... outer part of the nursery where we were **waiting** ...

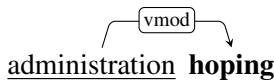


The second most common construction involves “vmod” (28%). Different from the “vmod” relations involved in (C)-type relations, the non-finite clauses usually modify noun phrases in (R)-type relations. For examples:

- (7) ... developed ... management consultants to go out ...

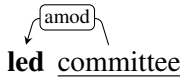


- (8) The administration, **hoping** to de-escalate the violence, is appealing to both sides.



The third most common type of cases involves participial adjectives, using “amod” syntactic relation (17%). Since the verb is modifying the noun as an adjective, the syntactic dependency and the semantic relation are reversed. For example:

- (9) ... a fact finding American **led** committee ...

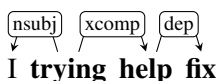


C.3 Others

The other constructions besides the three most common patterns are a mixture of data annotation errors, constituency-to-dependency failures, and combinations of the frequent patterns.

If the argument is shared with other predicates along the dependency path, then our conversion algorithm can recover the SRL relation through multiple (C)-type labels. For example, in the following sentence, the argument “I” is shared across three predicates “trying”, “help” and “fix” as ARG0’s.

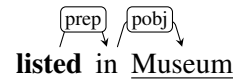
- (10) I’ve been **trying** to **help** him **fix** ...



Annotation inconsistencies can result in rare patterns beyond the scope of the current design of our joint label. For example, in the following sentence, the SRL annotation decides that “the Museum of Modern Art” is ARGM-LOC of “listed”, making

the predicate a grandparent of the argument. A simple fix that simply includes the preposition “in” as part of the argument span (as is annotated in most other examples) will change this case into a (D)-type relation.

- (11) Now your name is **listed** in the Museum of Modern Art.



D Chinese Data Analysis

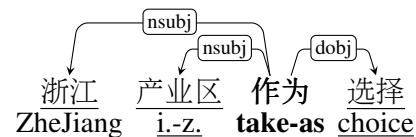
Despite the fact that Chinese and English are very different languages from two distinctive language families, they exhibit similar distributions of patterns when it comes to the syntactic patterns of SRL relations. The three most common types, (D)-, (C)- and (R)-type relations, account for over 98% of all annotated predicate-argument relations. In the following examples, BA denotes a *ba* construction, DE refers to a *de* particle, and CLASSIFIER represents Chinese measure words for quantity expressions (Huang et al., 2008).

D.1 (D)-Type Relations

Similarly to English, most Chinese SRL relations parallel the syntactic counterparts. For example, in the following sentence, each of the three arguments of the predicate corresponds to a direct dependent in the dependency structure.

- (12) 浙江 把 特色 产业区
ZheJiang BA featured industrial-zones
作为 经济 发展 的战略
take-as economic development DE strategic
选择
choice

“Zhejiang uses its featured industrial zones as a strategic choice for economic development.”



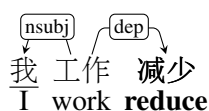
“i.-z.” is short for “industrial-zones”.

D.2 (C)-Type Relations

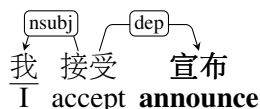
In a (C)-type relation, the most frequent syntactic labels between the common parent and the predicate are “dep” (37.5%), “conj” (34.4%) and “mmod” (9.6%). “conj” denotes coordinations, as

discussed in the English data analysis section. Unlike English data, the Chinese annotations use a large percentage of “dep” relations. A closer inspection reveals that most of the instances correspond to open clausal complements (“xcomp”) and coordinations (“conj”). (See English data analysis section for analysis.)

- (13) 我 真的 非常 努力 地 工作 ... 以
 I really very diligently DE work ... to
 减少 ...
 reduce ...
 “I work very diligently to reduce ...”

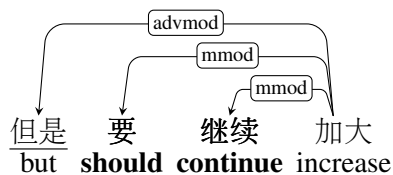


- (14) 我 能 接受 这个 , 并且 能 宣布 它
 I can accept this , and can announce it
 “I can accept this and announce it”



“mmod” is a dependency relation specific to Chinese. This label represents modal verb modifiers. In the Chinese SRL data, many of the modal verbs are annotated as predicates, resulting in (C)-type patterns. Additionally, “mmod” is frequently overloaded with conversions from some multi-verb constructions. The following sentence shows a common argument of two predicates. The first one (“should”) is a modal verb, while the second one (“continue”) is often treated as a standalone verb in a multi-verb construction (Li and Thompson, 1989).

- (15) 但是 要 继续 加大 改革
 but should continue increase reform
 力度
 strength
 “but we should continue to strengthen the reforms”

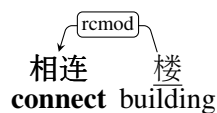


D.3 (R)-Type Relations

Similar to English, (R)-type relations are frequently used in relative clauses in Chinese as well. “rcmod”

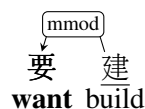
accounts for 64% of the syntactic relations connecting the predicates and the arguments. For example:

- (16) 一栋 众多 商户 相连 的
 a CLASSIFIER many merchants connect DE
 商业 楼
 commercial building
 “a commercial building that connects many merchants”



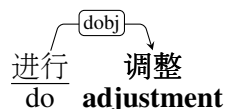
Other common constructions include “mmod” (12.9%), “dep” (6.6%) and “dobj” (5.9%). Complements of modal verbs can result in (R)-type patterns, illustrated as follows:

- (17) 上海 要 建 四个 中心
 Shanghai want build four CLASSIFIER center
 “Shanghai wants to build itself as four centers”



In (R)-type patterns, “dobj” commonly corresponds to light verb constructions where the object nouns are nominalized predicates while the syntactic heads are light verbs without much semantic meaning. Here we show an example:

- (18) 进行 适当 调整
 do adequate adjustment
 “adjust adequately”



Finally, cases involving “dep” relations include miscellaneous data annotation errors and constituency-to-dependency conversion errors.

D.4 Others

Similar to English, many of the unaccounted Chinese syntactic patterns of SRL relations are combinations of the three basic patterns. The following sentence illustrates propagation of an argument through multiple (C)-type structures.

- (19) 医院 扩大 药品 和 医疗
 hospital expand medicine and medical
 仪器 采购 规模 从而 压缩
 equipment purchase scale in-order-to reduce
 单位 成本 、 扩大 服务
 unit cost , expand service

“Hospitals expand the scale of purchasing medicines and medical equipments in order to reduce unit costs and to expand their service”

