InVeRo-XL: Making Cross-Lingual Semantic Role Labeling Accessible with <u>In</u>telligible <u>V</u>erbs and <u>Ro</u>les

Simone Conia
Sapienza University of Rome
conia@di.uniroma1.it

Riccardo Orlando
Babelscape, Italy
orlando@babelscape.com

Fabrizio Brignone
Babelscape, Italy
brignone@babelscape.com

Francesco Cecconi
Babelscape, Italy
cecconi@babelscape.com

Roberto Navigli
Sapienza University of Rome
navigli@diag.uniroma1.it

Abstract

Notwithstanding the growing interest in cross-lingual techniques for Natural Language Processing, there has been a surprisingly small number of efforts aimed at the development of easy-to-use tools for cross-lingual Semantic Role Labeling. In this paper, we fill this gap and present InVeRo-XL, an off-the-shelf state-of-the-art system capable of annotating text with predicate sense and semantic role labels from 7 predicate-argument structure inventories in more than 40 languages. We hope that our system – with its easy-to-use RESTful API and Web interface – will become a valuable tool for the research community, encouraging the integration of sentence-level semantics into cross-lingual downstream tasks. InVeRo-XL is available online at http://nlp.uniroma1.it/invero.

1 Introduction

Informally, Semantic Role Labeling (SRL) is often defined as the task of automatically answering the question *"Who did What, to Whom, Where, When, and How?"* (Màrquez et al., 2008). More precisely, SRL aims at recovering the predicate-argument structures within a sentence, providing an explicit overlay that uncovers the underlying semantics of text. For this reason, SRL is thought to be key in enabling Natural Language Understanding (Navigli, 2018). Today SRL is still an open problem, with several research papers being published each year at top-tier conferences, revealing novel insights and proposing better approaches. Over the years, thanks to this active development, SRL has been successfully exploited in a wide array of downstream tasks that span across different areas of Artificial Intelligence, from Natural Language Processing (NLP) with Information Retrieval (Christensen et al., 2010), Question Answering (He et al., 2015), Machine Translation (Marcheggiani et al., 2018) and Semantic Parsing (Banarescu et al., 2013), to Computer Vision with Visual Semantic Role Labeling (Gupta and Malik, 2015) and Situation Recognition (Yatskar et al., 2016).

Recently, the growing interest in cross-lingual NLP, supported by the increasingly wide availability of pretrained multilingual language models such as BERT (Devlin et al., 2019) and XLM-RoBERTa (Conneau et al., 2020), has sparked renewed interest in multilingual and cross-lingual SRL. In just a few years, researchers have found ways to design fully-neural end-to-end systems for SRL (Cai et al., 2018), to take advantage of contextual word representations (Peters et al., 2018; Li et al., 2019), to achieve high performance on multiple languages (He et al., 2019a; Conia and Navigli, 2020), to generate sense and role labels with sequence-to-sequence models (Blloshmi et al., 2021) and to perform SRL jointly across heterogeneous inventories (Conia et al., 2021).

Since SRL is a task that involves complex linguistic theories, inventories and techniques, there have been efforts to develop easy-to-use tools that offer automatic predicate sense and semantic role annotations to users interested in the integration of sentence-level semantics into downstream tasks. Some notable examples include SENNA[1] (Collobert et al., 2011), which uses an ensemble of feature-based classifiers (Koomen et al., 2005), AllenNLP's SRL demo[2], which provides a reimplementation of a BERT-based model (Shi and Lin, 2019), and InVeRo (Conia et al., 2020), which offers annotations according to two different linguistic inventories, PropBank (Palmer et al., 2005) and

---

[1] https://ronan.collobert.com/senna
[2] https://demo.allennlp.org/semantic-role-labeling

VerbAtlas (Di Fabio et al., 2019). However, one important drawback of the above-mentioned tools is that they are able to perform SRL only in English, which hinders the exploitation of their annotations in multilingual and cross-lingual NLP.

In order to fill this gap, we build upon InVeRo and propose its next major release, InVeRo-XL, with the objective of making SRL accessible in multiple languages. We rebuild InVeRo-XL from the ground up to offer:

- The first end-to-end system to tackle the whole SRL pipeline in over 40 languages;

- The first off-the-shelf system to provide SRL annotations for 7 linguistic inventories;

- A RESTful API service that can be queried either online, so as not to install any software, or offline, to maximize throughput;

- A Web interface that provides a visualization of the system output which can be useful for teaching purposes, comparing linguistic theories, and prototyping new ideas.

We believe that InVeRo-XL can provide a stepping stone for the integration of explicit sentence-level semantics into cross-lingual tasks, attracting new researchers to the field of SRL and its applications.[3]

1.1 What's New in InVeRo-XL

As previously mentioned, InVeRo-XL is the successor of InVeRo. Although its main new feature is the ability to provide predicate sense and semantic role annotations in over 40 languages with 7 different inventories, InVeRo-XL has been overhauled to also improve several other important aspects. In particular:

- **Preprocessing:** while its predecessor used a very limited set of rules to preprocess English text, InVeRo-XL features a multilingual preprocessing module based on spaCy and Stanza;

- **SRL model:** the English-only model has been replaced by a cross-lingual model that is able to perform not only span-based SRL, but also dependency-based SRL;

- **API:** the service is now able to handle batched requests and documents of arbitrary length;

- **Offline usage:** InVeRo-XL is now available for download, free for research purposes, allowing users to host their own instance locally.

2 System Overview

In this Section, we provide an overview of the main components of InVeRo-XL and how they interact, describing in detail the preprocessing module (Section 2.1) and the SRL model (Section 2.2).

2.1 Preprocessing

The previous version of InVeRo-XL preprocessed an English sentence using a very limited and simple set of rules. In order to correctly support more languages, InVeRo-XL now relies on both spaCy (Honnibal et al., 2020) and Stanza (Qi et al., 2020) to deal transparently with document splitting and tokenization. An automatic language detector based on fastText[4] (Joulin et al., 2017) is used to dynamically choose between the two preprocessing tools, depending on the language detected: spaCy is faster for high-resource languages, e.g. English, but also less reliable on lower-resource languages, e.g. Catalan, for which our system falls back to Stanza.

2.2 Model Architecture

In line with its predecessor, InVeRo-XL encapsulates an SRL model that falls within the broad category of end-to-end systems, tackling the whole SRL pipeline – predicate identification, predicate sense disambiguation, argument identification and argument classification – in a single forward pass. However, the design of the SRL model itself has been completely revamped and now follows the architecture recently proposed by Conia et al. (2021), which is capable of performing cross-lingual SRL with heterogeneous linguistic inventories. In the following, we describe the main components of the SRL model architecture provided by InVeRo-XL.

Multilingual word encoder. The first component of our model is a multilingual word encoder that takes advantage of a pretrained language model, XLM-RoBERTa (Conneau et al., 2020), to provide rich contextualized word representations.

---

[3]InVeRo-XL can be downloaded upon request at http://nlp.uniroma1.it/resources. InVeRo-XL is licensed under Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International.

[4]https://fasttext.cc

More formally, for each word w_i in an input sentence $\mathbf{w} = \langle w_1, w_2, \ldots, w_n \rangle$ of length n, it computes an encoding $\mathbf{e}_i = \text{Swish}(\mathbf{W}^w \mathbf{h}_i + \mathbf{b}^w)$ as a non-linear projection of the concatenation $\mathbf{h}_i$ of the corresponding hidden states of the four topmost layers of the language model.

Universal word encoder. The resulting sequence of multilingual word encodings $\mathbf{E} = \langle \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n \rangle$ is then given to a "universal" word encoder that computes a sequence of task-specific timestep encodings $\mathbf{T} = \langle \mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_n \rangle$ as follows:

$$\mathbf{t}_i^j = \begin{cases} \mathbf{e}_i & \text{if} \quad j = 0 \\ \mathbf{t}_i^{j-1} \oplus \text{BiLSTM}_i^j(\mathbf{t}^{j-1}) & \text{otherwise} \end{cases}$$
$$\mathbf{T} = \langle \mathbf{t}_1^K, \mathbf{t}_2^K, \ldots, \mathbf{t}_n^K \rangle$$

where $\text{BiLSTM}_i^j(\cdot)$ is the i-th timestep of the j-th BiLSTM layer and K is the total number of BiLSTM layers. The purpose of this encoder is to create representations that are shared across languages and inventories and are, therefore, "universal".

Universal predicate-argument encoder. Similarly to the encoder above, the objective of the universal predicate-argument encoder is to build predicate-specific argument representations that lie in a vector space shared across languages and inventories. Assuming that w_p is a predicate in the input sentence $\mathbf{w}$, this encoder builds a sequence $\mathbf{A}$ of predicate-specific argument encodings as follows:

$$\mathbf{a}_i^j = \begin{cases} \mathbf{t}_p \oplus \mathbf{t}_i & \text{if} \quad j = 0 \\ \mathbf{a}_i^{j-1} \oplus \text{BiLSTM}_i^j(\mathbf{a}^{j-1}) & \text{otherwise} \end{cases}$$
$$\mathbf{A} = \langle \mathbf{a}_1^{K'}, \mathbf{a}_2^{K'}, \ldots, \mathbf{a}_n^{K'} \rangle$$

where $\mathbf{t}_i$ is the i-th timestep encoding from the universal sentence encoder, $\mathbf{a}_i^j$ is the argument encoding for w_i with respect to w_p produced after the j-th BiLSTM layer, and K' is the total number of BiLSTM layers.

Inventory-specific decoders. Finally, the universal encodings are given to a set of classifiers in order to obtain the desired output labels. More specifically, for each inventory, we need three types of output: i) whether a word w_i is a predicate w_p; ii) the most appropriate sense s for a predicate w_p; iii) which semantic role r, possibly the null role, exists between a word w_i and a predicate w_p. More formally, our model features three classifiers for each inventory I as follows:

$$\sigma^p(w_i|I) = \mathbf{W}^{p|I} \text{Swish}(\mathbf{W}^p \mathbf{t}_i + \mathbf{b}^p) + \mathbf{b}^{p|I}$$
$$\sigma^s(w_p|I) = \mathbf{W}^{s|I} \text{Swish}(\mathbf{W}^s \mathbf{t}_p + \mathbf{b}^s) + \mathbf{b}^{s|I}$$
$$\sigma^r(w_r|w_p, I) = \mathbf{W}^{r|I} \text{Swish}(\mathbf{W}^r \mathbf{a}_i + \mathbf{b}^r) + \mathbf{b}^{r|I}$$

where each $\sigma^{\cdot}(\cdot)$ provides a score distribution over the possible output classes, i.e. two (true or false) for predicate identification, the number of senses of an inventory for predicate sense disambiguation, and the number of semantic roles (including the null role) of an inventory for argument labeling.

Miscellanea. While we follow the architecture proposed by Conia et al. (2021), the SRL model of InVeRo-XL also comes with a small but significant number of enhancements. One such enhancement is that, while Conia et al. (2021) propose a model for dependency-based SRL, our model is also able to perform span-based SRL by treating spans as sequences of BIO tags. In order to correctly decode valid spans at inference time, InVeRo-XL makes use of a Viterbi decoder. Other improvements include training the model with the RAdam optimizer (Liu et al., 2020), ensuring that each training batch features a balanced number of instances for each language in the training set, and searching randomly for better hyperparameter values.

2.3 Evaluation

Datasets. We report the performance of InVeRo-XL on two gold standard benchmarks for SRL: CoNLL-2009 (Hajič et al., 2009) for dependency-based SRL and CoNLL-2012 (Pradhan et al., 2012) for span-based SRL. To the best of our knowledge, CoNLL-2009 is the largest benchmark for multilingual SRL as it comprises six languages, namely, Catalan, Chinese, Czech, English, German and Spanish.[5] The main challenge of this benchmark is that each language was annotated with a different predicate-argument structure inventory, e.g. the English PropBank (Palmer et al., 2005) for English, AnCora (Taulé et al., 2008) for Spanish/Catalan and PDT-Vallex (Hajic et al., 2003) for Czech. While CoNLL-2009 is an ideal test bed for evaluating the multilingual capabilities of an SRL system, dependency-based annotations may look unfamiliar to end users who are not used to the

---

[5]Japanese is not available anymore from LDC due to licensing issues.

		Catalan	Czech	German	English	Spanish	Chinese
Span	AllenNLP's SRL demo	–	–	–	86.5	–	–
	InVeRo	–	–	–	86.2	–	–
	InVeRo-XL_{span-based}	**83.3**	**85.9**	**87.0**	**86.8**	**81.8**	**84.9**
Dependency	Marcheggiani et al. (2017)	—	86.0	—	87.7	80.3	81.2
	Chen et al. (2019)	81.7	88.1	76.4	91.1	81.3	81.7
	Cai and Lapata (2019b)	—	—	82.7	90.0	81.8	83.6
	Cai and Lapata (2019a)	—	—	83.8	91.2	82.9	85.0
	Lyu et al. (2019)	80.9	87.5	75.8	90.1	80.5	83.3
	He et al. (2019b)	86.0	89.7	81.1	90.9	85.2	86.9
	Conia and Navigli (2020)	88.3	**92.1**	89.1	**92.4**	86.9	**89.1**
	Conia et al. (2021)	88.0	91.5	88.0	91.8	86.3	87.7
	InVeRo-XL_{dependency-based}	**88.7**	**92.1**	**89.9**	92.1	**87.2**	**89.1**

Table 1: Comparison between InVeRo-XL and other recent systems for SRL. **Top:** F_1 scores on argument labeling with pre-identified predicates using the official CoNLL-2005 scoring script on the CoNLL-2012 English test set for span-based SRL and the CoNLL-2009 test sets converted from dependency-based to span-based as described in Section 2.3. **Bottom:** F_1 scores on argument labeling and sense disambiguation with pre-identified predicates using the official CoNLL-2009 scoring script on the test sets of the CoNLL-2009 shared task for dependency-based multilingual SRL.

notion of syntactic/semantic heads. Therefore, differently from Conia et al. (2021), we also adapt the system to perform span-based SRL and evaluate its effectiveness on the standard English datasets of CoNLL-2012 and on CoNLL-2009, converting dependency-based annotations to span-based annotations. We convert an argument head to an argument span by considering all those words that fall in the syntactic subtree whose root is the argument head and discarding all those predicates for which this conversion produces overlapping spans.

Experimental setup. We train the dependency-based SRL model on the standard training splits of CoNLL-2009, making the model learn from all six languages jointly. Instead, we train the span-based SRL model on the union of the English training split of CoNLL-2012 and the Catalan, Chinese, Czech, German and Spanish training sets converted from dependency-based to span-based, as explained above. Each model configuration is trained for 30 epochs using the RAdam optimizer with learning rates of 10^{-5} for the weights of XLM-RoBERTa and 10^{-3} for the other weights. Following standard practice, we select the model checkpoint with highest F_1 score on the development set.

Results. Table 1 (top) shows how InVeRo-XL performs on the English test set of CoNLL-2012 for span-based SRL compared to its previous release

(InVeRo) and the previously best-performing online system for SRL (AllenNLP's SRL demo). Not only does InVeRo-XL achieve better results, but it is also the only system that is capable of performing span-based cross-lingual SRL, showing strong results on each of the non-English test sets of CoNLL-2009 converted from dependency-based to span-based as described in Section 2.3. Furthermore, Table 1 (bottom) shows that InVeRo-XL achieves results that are comparable to or better than those of current state-of-the-art models on 5 of the 6 languages of CoNLL-2009 for dependency-based SRL, the key advantages being that our model is part of a prepackaged tool with additional user-friendly features (see Sections 3 and 4).

3 The InVeRo-XL API

In order to facilitate the integration of predicate-argument structure information into downstream tasks, InVeRo-XL exposes its fully self-contained end-to-end multilingual SRL pipeline through an easy-to-use RESTful API. In the following, we provide an overview of the main functionalities of the InVeRo-XL API, from its Resource API (Section 3.1) to its Model API (Section 3.2) and how to host InVeRo-XL locally on a user's own hardware (Section 3.3). We refer users to the online documentation for the complete list of supported languages

and inventories, together with other details.[6]

3.1 Resource API

The Resource API is a simple way for obtaining semantic information about predicates using the intelligible verb senses and semantic roles defined by VerbAtlas, a large-scale predicate-argument structure inventory which clusters WordNet synsets (Miller, 1992) that share similar semantic behavior. The Resource API of InVeRo-XL builds upon the functionalities provided by its predecessor with the key difference that it now supports multiple languages thanks to BabelNet 5.0[7] (Navigli and Ponzetto, 2012; Navigli et al., 2021), a multilingual encyclopedic dictionary that provides unified access to several knowledge bases including WordNet.

More specifically, the Resource API defines two endpoints:

- `/api/verbatlas/predicate`: given a predicate p, this endpoint retrieves the set of VerbAtlas frames which include at least one sense of p.

- `/api/verbatlas/frame`: given a VerbAtlas frame f, this endpoint retrieves its predicate-argument structure, i.e., the semantic roles, and the WordNet/BabelNet synsets that belong to f.

3.2 Model API

The Model API of InVeRo-XL has been updated to not only take advantage of the new multilingual SRL system but also to provide quality-of-life improvements. The Model API now accepts requests in over 40 languages and returns semantic annotations according to 7 linguistic inventories. On top of this, the Model API is now able to process documents of arbitrary length and to handle batches of documents in a single request.

More specifically, the Model API exposes an endpoint named `/api/model`. This endpoint accepts `POST` requests with a `JSON` body containing a list of input objects, one for each document the user wishes to annotate. Each input object shall specify the following fields:

- `text`: a mandatory field that contains the text of the document.

```
[{
    "tokens": [
        ...
        { "index": 2, "text": "volpe" },
        { "index": 3, "text": "salta" },
        { "index": 4, "text": "sul" },
        ...
    ],
    "annotations": [
        {
            "tokenIndex": 3,
            "verbatlas": {
                "sense": "GO-FORWARD",
                "arguments": [
                    {
                        "role": "Agent",
                        "score": 1.0,
                        "span": [0, 3]
                    },{
                        "role": "Destination",
                        "score": 1.0,
                        "span": [4, 6]
                    },
                    ...
                ]
            },
            "englishPropbank": {...},
            "chinesePropbank": {...},
            "germanPropbank": {...},
            "pdtVallex": {...},
            "catalanAncora": {...},
            "spanishAncora": {...}
        }
        ...
    ]
}]
```

Figure 1: An example of a response from the Model API for an input Italian sentence. The response contains the tokenized input sentence and the automatic SRL annotations according to 7 different linguistic inventories.

- `lang`: an optional field that indicates the language of the document. If omitted, InVeRo-XL will use an automatic language detector (see Section 2.1).

Each request to the Model API returns a `JSON` response containing a list of output objects, one for each input document, containing the automatic annotations according to each of the 7 linguistic inventories, as shown in Figure 1.

---

[6] http://nlp.uniroma1.it/invero/api-documentation
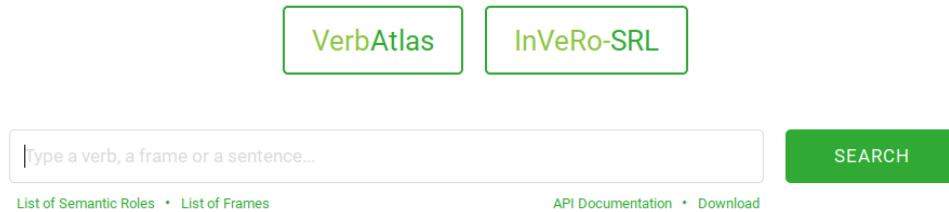[7] https://babelnet.org

Figure 2: The home page of the Web interface of InVeRo-XL. Users can search for predicate information (e.g. the VerbAtlas frames a verb belongs to) and tag sentences in multiple languages with different linguistic inventories (see Figure 3).

3.3 Offline Usage

One of the most requested features that is currently missing from InVeRo is the possibility of running an offline instance of the service so as to annotate large quantities of text in a shorter time, independently of the latency of the network and the volume of requests being processed by our Web server. To address this issue, InVeRo-XL is also distributed as a Docker[8] image that can be deployed locally on a user's own hardware.[9] While network latency is often a bottleneck for processing a request, an offline instance of InVeRo-XL does not suffer from such a constraint and can therefore benefit greatly from running on better hardware, e.g. on GPU. We distribute InVeRo-XL in two configurations:

- `invero-xl-span` is the configuration that performs multilingual span-based SRL and is the one used by InVeRo-XL's Web server;

- `invero-xl-dependency` is an alternative configuration built to perform multilingual dependency-based SRL.

Running a local instance of InVeRo-XL is also simple. First, users are required to perform a one-time setup to load one of the available images:

```
#!/bin/bash
docker load -i invero-xl-span_2.0.0.tar
```

After that, InVeRo-XL can be started with:

```
#!/bin/bash
PORT=12345
LANGUAGES="EN IT FR ZH"

docker run \
  --name invero-xl-span-en \
  -p $PORT:80 \
  -e LANGUAGES=$LANGUAGES
  invero-xl-span:2.0.0
```

Once started, users can forward their requests locally. We refer the reader to the online documentation for further details.[10]

4 Web Interface

Similarly to its predecessor, InVeRo-XL includes a public-facing Web interface (Figures 2 and 3) that provides a visual environment for both the Resource API and the Model API, allowing users to explore the main functionalities while also providing an intuitive overview of how an SRL system annotates a sentence or a short document. Most importantly, the Web interface of InVeRo-XL has been updated to reflect the changes in the Model API and the underlying SRL model; now users can annotate text in 12 languages[11] and visualize predicate senses and semantic roles in 7 linguistic inventories on the fly, without having to write code.

Figure 3 shows an example sentence in Italian with its corresponding predicate senses and seman-

---

[8] https://www.docker.com
[9] Docker images for InVeRo-XL are freely available for research purposes at http://nlp.uniroma1.it/resources.

[10] http://nlp.uniroma1.it/invero/api-documentation
[11] We limit the number of languages available on the Web interface due to hardware constraints as Stanza and spaCy use one preprocessing model for each language.

Figure 3: The beginning of the Divine Comedy by Dante Alighieri in Italian as tagged by InVeRo-XL with three predicate-argument structure inventories – VerbAtlas, the Chinese PropBank and the Spanish AnCora. "Nel mezzo del cammin di nostra vista mi ritrovai per una selva oscura, ché la diritta via era smarrita" translates into "Midway upon the journey of our life I found myself within a forest dark, for the straightforward pathway had been lost".

tic roles as provided by InVeRo-XL. Thanks to a dropdown menu, users can immediately switch from the labels of one inventory to those of another, independently of the input language, without reloading the Web page. We argue that this Web interface should help teachers explain SRL to their students, allow linguists to compare linguistic inventories on particular case studies, attract new researchers to the field, and inspire others to exploit SRL in downstream tasks or even real-world scenarios.

5 Conclusion and Future Work

Over the years, the research community has greatly advanced the field of SRL, proposing ever more complex approaches to tackle the task more effectively. However, despite the growing interest in cross-lingual NLP, there have been very few efforts to develop automatic tools to perform SRL in multiple languages. Our objective with InVeRo-XL is to fill this gap and equip researchers with an easy-to-use, high-performing system capable of providing predicate sense and semantic role annotations in over 40 languages with 7 linguistic inventories. Users can take advantage of our state-of-the-art sys-

tem for cross-lingual SRL through a RESTful API that relieves them from the need to reimplement complex neural models and/or to build an efficient preprocessing/postprocessing pipeline.

Although InVeRo-XL is a major step forward compared to its predecessor, we intend to further improve our system by adopting future and more advanced SRL model architectures and by including new training datasets, such as UniteD-SRL (Tripodi et al., 2021). We strongly believe that InVeRo-XL will facilitate the integration of SRL into downstream cross-lingual tasks, hopefully aiding further advancements in cross-lingual Natural Language Understanding.

Acknowledgments

References

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

Rexhina Blloshmi, Simone Conia, Rocco Tripodi, and Roberto Navigli. 2021. Generating Senses and RoLes: An end-to-end model for dependency- and span-based Semantic Role Labeling. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 3786–3793. Main Track.

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. 2019a. Semi-supervised Semantic Role Labeling with cross-view training. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1018–1027, Hong Kong, China.

Rui Cai and Mirella Lapata. 2019b. Syntax-aware Semantic Role Labeling without parsing. *Trans. Assoc. Comput. Linguistics*, 7:343–356.

Xinchi Chen, Chunchuan Lyu, and Ivan Titov. 2019. Capturing argument interaction in Semantic Role Labeling with capsule networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5415–5425, Hong Kong, China.

Janara Christensen, Mausam, Stephen Soderland, and Oren Etzioni. 2010. Semantic Role Labeling for open information extraction. In *Proceedings of the NAACL HLT 2010 First International Workshop on Formalisms and Methodology for Learning by Reading*, pages 52–60, Los Angeles, California.

Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel P. Kuksa. 2011. Natural language processing (almost) from scratch. *J. Mach. Learn. Res.*, 12:2493–2537.

Simone Conia, Andrea Bacciu, and Roberto Navigli. 2021. Unifying cross-lingual Semantic Role Labeling with heterogeneous linguistic resources. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 338–351, Online.

Simone Conia, Fabrizio Brignone, Davide Zanfardino, and Roberto Navigli. 2020. InVeRo: Making Semantic Role Labeling accessible with intelligible verbs and roles. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 77–84, Online.

Simone Conia and Roberto Navigli. 2020. Bridging the gap in multilingual Semantic Role Labeling: A language-agnostic approach. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1396–1410, Barcelona, Spain (Online).

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online.

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.

Andrea Di Fabio, Simone Conia, and Roberto Navigli. 2019. VerbAtlas: A novel large-scale verbal semantic resource and its application to 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 (EMNLP-IJCNLP)*, pages 627–637, Hong Kong, China.

Saurabh Gupta and Jitendra Malik. 2015. Visual Semantic Role Labeling. *arXiv*, abs/1505.04474.

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 (CoNLL 2009): Shared Task*, pages 1–18, Boulder, Colorado.

Jan Hajic, Jarmila Panevová, Zdenka Urešová, Alevtina Bémová, Veronika Kolárová, and Petr Pajas. 2003. PDT-VALLEX: Creating a large-coverage valency lexicon for treebank annotation. In *Proceedings of the second workshop on treebanks and linguistic theories*, volume 9, pages 57–68.

Luheng He, Mike Lewis, and Luke Zettlemoyer. 2015. Question-answer driven Semantic Role Labeling:

Using natural language to annotate natural language. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 643–653, Lisbon, Portugal.

Shexia He, Zuchao Li, and Hai Zhao. 2019a. 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 (EMNLP-IJCNLP)*, pages 5350–5359, Hong Kong, China.

Shexia He, Zuchao Li, and Hai Zhao. 2019b. 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, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 5349–5358.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python.

Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. Bag of tricks for efficient text classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 427–431, Valencia, Spain.

Peter Koomen, Vasin Punyakanok, Dan Roth, and Wen-tau Yih. 2005. Generalized inference with multiple Semantic Role Labeling systems. In *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)*, pages 181–184, Ann Arbor, Michigan.

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 *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 6730–6737.

Liyuan Liu, Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao, and Jiawei Han. 2020. On the variance of the adaptive learning rate and beyond. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*.

Chunchuan Lyu, Shay B. Cohen, and Ivan Titov. 2019. Semantic Role Labeling with iterative structure refinement. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1071–1082, Hong Kong, China.

Diego Marcheggiani, Jasmijn Bastings, and Ivan Titov. 2018. Exploiting semantics in Neural Machine Translation with graph convolutional networks. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 486–492, New Orleans, Louisiana.

Diego Marcheggiani, Anton Frolov, and Ivan Titov. 2017. A simple and accurate syntax-agnostic neural model for dependency-based Semantic Role Labeling. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pages 411–420, Vancouver, Canada.

Lluís Màrquez, Xavier Carreras, Kenneth C. Litkowski, and Suzanne Stevenson. 2008. Special issue introduction: Semantic Role Labeling: An introduction to the special issue. *Computational Linguistics*, 34(2):145–159.

George A. Miller. 1992. WordNet: A lexical database for English. In *Speech and Natural Language: Proceedings of a Workshop Held at Harriman, New York, February 23-26, 1992*.

Roberto Navigli. 2018. Natural Language Understanding: Instructions for (present and future) use. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, pages 5697–5702.

Roberto Navigli, Michele Bevilacqua, Simone Conia, Dario Montagnini, and Francesco Cecconi. 2021. Ten years of BabelNet: A survey. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 4559–4567. Survey Track.

Roberto Navigli and Simone Paolo Ponzetto. 2012. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artif. Intell.*, 193:217–250.

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

Matthew E. 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.

Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Olga Uryupina, and Yuchen Zhang. 2012. CoNLL-2012 shared task: Modeling multilingual unrestricted coreference in OntoNotes. In *Joint Conference on EMNLP and CoNLL - Shared Task*, pages 1–40, Jeju Island, Korea.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A Python Natural Language Processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 101–108, Online.

Peng Shi and Jimmy Lin. 2019. Simple BERT models for relation extraction and semantic role labeling. *CoRR*, abs/1904.05255.

Mariona Taulé, M. Antònia Martí, and Marta Recasens. 2008. AnCora: Multilevel annotated corpora for Catalan and Spanish. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco.

Rocco Tripodi, Simone Conia, and Roberto Navigli. 2021. UniteD-SRL: A unified dataset for span- and dependency-based multilingual and cross-lingual Semantic Role Labeling. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Mark Yatskar, Luke S. Zettlemoyer, and Ali Farhadi. 2016. Situation Recognition: Visual Semantic Role Labeling for image understanding. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 5534–5542.