european lexicographic infrastructure

D3.4 Multilingual Semantic Parsing – Initial Report

Authors: Rexhina Blloshmi, Roberto Navigli, Paolo Spadoni, Paola Velardi
Date: 31.10.2020

european lexicographic infrastructure

H2020-INFRAIA-2016-2017

Grant Agreement No. 731015

ELEXIS - European Lexicographic Infrastructure

3.4 Multilingual Semantic Parsing – initial report

Deliverable Number:	3.4
Dissemination Level:	PUBLIC
Delivery Date:	31.10.2020
Version:	1.0
Authors:	Rexhina Blloshmi Roberto Navigli Paolo Spadoni Paola Velardi

european lexicographic infrastructure

Project Acronym:	ELEXIS
Project Full Title:	European Lexicographic Infrastructure
Grant Agreement No.:	731015

Deliverable/Document Information

Project Acronym:	ELEXIS
Project Full Title:	European Lexicographic Infrastructure
Grant Agreement No.:	731015

Document History

Version Date	Changes/Approval	Approved by
31.10.2020	Initial version	Roberto Navigli

Table of Contents

Glossary

AMR – Abstract Meaning Representation

NLP – Natural Language Processing

NLU – Natural Language Understanding

1 Introduction

Natural Language Understanding (NLU) is the research area cutting across Natural Language Processing (NLP), information retrieval and human computer interaction. Semantic parsing is a central task of NLU and it would seem to hold the potential to achieve the ambitious objective of machine reading, one of the long-standing goals of Artificial Intelligence. Semantic parsing aims at representing the meaning of the natural language sentences as formal structural representations. High performance semantic parsing solutions are desirable as they provide symbolic representations (which can be manipulated by both humans and computer programs) and moreover these structured forms can be easily executed on a knowledge base and on semantic networks. During years, different formalisms have been proposed such as Elementary Dependency Structures (Oepen and Lønning, 2006, EDS), Prague Tectogrammatical Graphs (Hajič et al., 2012, PTG), Abstract Meaning Representation (Banarescu et al., 2013, AMR), Universal Conceptual Cognitive Annotation (Abend and Rappoport, 2013, UCCA), Universal Decompositional Semantics (White et al., 2016, UDS).

AMR is one of the most popular formalisms for natural language among all the semantic representations that have been proposed. It has gained a lot of interest in literature for representing the meaning of a sentence and is actively integrated in different applications of NLP, such as Machine Translation (Song et al, 2019b), Text Summarization (Hardy and Vlachos, 2018; Liao et al., 2018) and Information Extraction (Rao et al., 2017), achieving promising performances. For this reason, in this report, we focus on AMR parsing and work towards the goal of multilingual semantic parsing exploiting this formalism.

Semantic parsing formalisms, including AMR, are however mostly focused towards English and little or no progress has been done across languages. Therefore, expanding across languages is challenging especially because the annotations needed for producing sufficiently large datasets to train supervised systems that are able to learn semantic representations, is a costly and slow process. In fact, semantic parsing performance can be greatly improved with the availability of an interconnected network of resources, especially for wide coverage across languages. In addition, AMR makes extensive use of PropBank framesets, whose availability across languages is limited. This dependency could eventually be an obstacle for the goal of wide-coverage multilingual semantic parsing and due to this the requirement for a non language-specific verbal resource emerges.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

The aim of Task 3.2, Multilingual Semantic Parsing, is both to develop innovative algorithms that achieve state-of-the-art semantic parsing across languages and to exploit multilingual resources and without relying on manually curated data for performing it. In this report, we provide details about our solution to cross-lingual AMR parsing (Damonte and Cohen, 2018), a task that represents sentences across languages with the AMR graph of their English translation. We enable semantic parsing across four languages without relying on manually annotated data and by exploiting the available bilingual and multilingual data. Moreover we introduce VerbAtlas (Di Fabio et al., 2019), a hand-crafted lexical-semantic resource whose goal is to bring together all verbal synsets from BabelNet into semantically-coherent frames. Within VerbAtlas, different lexicalizations of the same verb are clustered together within the same frame and have the same argument structure. Since VerbAtlas is connected to BabelNet (Navigli and Ponzetto, 2010), it further clusters together different lexicalizations of verbs across languages. This resource could be highly beneficial for this task for which multilingual coverage is the central goal.

Provided that this is an initial report of this task, and that the work aimed at creating the ELEXIS dictionary matrix which will contain parallel and/or comparable meanings is still ongoing, in the first part of the project we focused on solutions geared towards enabling multilingual semantic parsing by exploiting existing resources and achieving satisfying performance cross-lingually. In the second part of the project, instead, the additional, multilingually-interconnected resources made available from other tasks (mostly WP2) will play a key role to further improving our proposed approach and techniques towards wider coverage multilingual semantic parsing.

This deliverable is organized as follows. We first introduce the new VerbAtlas resource. Secondly, we define the AMR formalism and describe the task of AMR and cross-lingual AMR parsing. Third, we pose the main challenges of parsing sentences across languages using the AMR formalism. Then we explain our approach to these problems and detail the algorithm employed to achieve the objectives and goals of Task 2.3. Finally, we report a

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

quantitative and qualitative evaluation on a cross-lingual AMR evaluation benchmark on Chinese, German, Italian and Spanish.

VerbAtlas and experiments on Semantic Role Labeling improved by the joint use of PropBank and VerbAtlas are described in the following publication:

Andrea Di Fabio, Simone Conia, Roberto Navigli. VerbAtlas: a Novel Large-Scale Verbal Semantic Resource and Its Application to Semantic Role Labeling. *In Proceedings of the 2019 Empirical Methods for Natural Language Processing (EMNLP/IJCNLP 2019)*, pp. 627-637

The cross-lingual AMR parsing approach included in this report has been recently published in the following paper*:*

Rexhina Blloshmi, Rocco Tripodi, and Roberto Navigli. 2020. XL-AMR: Enabling Cross-Lingual AMR Parsing with Transfer Learning Techniques. *In Proceedings of the 2020 Empirical Methods for Natural Language Processing (EMNLP2020)*.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

2 VerbAtlas

The semantics of verbs plays an important role in understanding the whole meaning of a sentence as they define the main argument structure of "*who did what to whom*" which in turn is at the core of what is expressed by the semantic formalism in semantic parsing. For this reason devising proper large-scale verbal resources is of high interest in NLU. Several existing verb inventories in the literature, such as PropBank (Palmer et al., 2005) FrameNet (Baker et al., 1998) and VerbNet (Kipper-Schuler, 2005) - with PropBank being extensively used in AMR semantic parsing - have several limitations such as language specificity, low verbal coverage and lack of interoperability with existing knowledge bases. In addition, the included semantic roles range from underspecified as in PropBank to overspecified as in FrameNet. VerbAtlas is a manually-crafted inventory of verbs and argument structures which addresses the limitations of the foregoing verbal resources and provides not only full coverage of the English verb lexicon associated with explicit and informative argument structure roles, but also links to the BabelNet (Navigli and Ponzetto, 2010) and Open Multilingual Wordnet (Bond and Foster, 2013) knowledge bases which make it scale across languages, in line with the goal of Task 3.2. In addition, a novelty of VerbAtlas is the specification of refined semantic information and selectional preferences for the argument structure of frames.

In summary, VerbAtlas organizes the English verb lexicon in semantically clustered frames which in turn consist of verbs that share the same argument structure expressed by explicit semantic roles, often associated with selectional preferences. Table 1 summarizes the size of the existing resources in terms of cluster types, distinct argument roles and meaning units.

	Cluster types	**#**	**Argument roles**	**#**	**Meaning units**	**#**
FrameNet	Frames	1,224	Frame elements	10,542	Lexical units	5,200
VerbNet	Levin's classes	329	Thematic roles	39	Senses	6,791
PropBank	Verbs	5,649	Proto-roles	6	Framesets	10,687
WordNet	–	–	–	–	Synsets	13,767
VerbAtlas	Frames	466	Semantic roles	25	Synsets	13,767

Table 1: Quantitative analysis of verbal resources

The data and software is available at http://verbatlas.org.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

2.1 Frames

VerbAtlas frames expand upon the frame notion of FrameNet. A frame in VerbAtlas is defined as a cluster of WordNet synsets which express similar shades of meaning and that express a certain scenario, which in turn defines their argument structure. As shown in Table 1, VerbAtlas frames are organized into 466 full WordNet synset coverage and semantically coherent clusters which provide cross-frame argument structures. This overcomes some of the issues of other existing resources such as the sparsity of frames in FrameNet, the independent argument structure of verbs in PropBank and the clustering according to syntactic similarity of VerbNet.

An example of a VerbAtlas frame is the frame of EAT, which comprises all the synsets that express an *"eating"* scenario and includes synsets for *eat, devour, guttle, raven, pig*, etc.

2.2 Semantic Roles and Prototypical Argument Structure

VerbAtlas semantic roles come with two advantages: they are cross-frame, i.e., verbs within the same frame share the same argument structure, and they are explicit and human readable. VerbAtlas inventory of 25 semantic roles is inspired by VerbNet, whose 39 labels (like AGENT, PATIENT, LOCATION, etc.) are explicit, cross-frame and domain-general, but instead VerbAtlas merges together some of the VerbNet roles which can be seen as complementary. In Table 2 we provide the full list of roles in VerbAtlas.

Agent	Material
Attribute	Patient
Beneficiary	Product
Cause	Purpose
Co-Agent	Recipient
Co-Patient	Result
Co-Theme	Source
Destination	Stimulus
Experiencer	Theme
Extent	Time
Goal	Topic
Instrument	Value
Location	

Table 2: VerbAtlas semantic roles

Each VerbAtlas frame expresses a Prototypical Argument Structure (PAS) that generalizes over all the synsets in a particular frame which defines the frame's overall meaning.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

VerbAtlas does not distinguish between core roles and adjuncts. Therefore, in order to be fully inclusive of the possible scenarios in which the verbs within a frame can be expressed, VerbAtlas includes in PAS roles which might be projected optionally by argument structures that are nonetheless present in the scenario evoked by the frame.

Interestingly, VerbAtlas contains selectional preferences for the semantic roles of the PAS labeled with a set of 116 macro-concepts, defined by WordNet synsets whose hyponyms are expected to be likely candidates to the corresponding argument slot. This is done to narrow down the number of candidates for a particular argument slot thus providing further semantic structure.

2.3 Synset-level Semantic Information

VerbAtlas makes use of synsets' glosses and examples in WordNet to enrich the semantic representation of predicate synsets with semantic and pragmatic information that includes implicit, shadow and default arguments - inspired by Pustejovsky (1995) - making it the only large-scale verbal resource providing this kind of semantics. This information, if properly exploited, could be beneficial to better representing the meaning.

An **implicit argument** in the argument structure of a verb is not always expressed syntactically but can be inferred from the synset gloss and used to imply a selectional preference on the role's synset. For example, the gloss of the synset {overleap, vault} in the JUMP frame is *"Jump across or leap over (an obstacle)"* implies that the PATIENT of this verb can be a hyponym of {obstacle}.

A **shadow argument** is incorporated in the meaning of a verb but is, likewise, not syntactically expressed. An example from the EAT frame is {eat in, dine in} ("Eat at home"). This synset is tagged with the shadow argument LOCATION = {home}.

A **default argument** is logically implied but not syntactically expressed. These are also tagged as shadow arguments. For instance, the synset {deliver} (as in "Our local supermarket delivers") has the label PATIENT = {grocery} to provide the commonsense information that what a supermarket usually delivers is groceries.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

3 Abstract Meaning Representation

After providing a resource which is as independent of the language as possible (compared to existing resources like PropBank and FrameNet), we started work on semantic parsing using AMR, which - as will be seen later in the report - is prodromal to future work on the integration of VerbAtlas and further lexical-semantic knowledge from the ELEXIS dictionary matrix, so as to enable multilingual semantic parsing.

3.1 AMR

AMR (Banarescu et al., 2013) is a popular formalism for natural language, which represents a sentence into a rooted, directed and acyclic graph. AMR unifies, in a single structure, a rich set of information coming from different tasks such as Named Entity Recognition (NER), Semantic Role Labeling (SRL), Word Sense Disambiguation (WSD), and coreference resolution.

The nodes on the graph are concepts drawn from PropBank framesets (Kingsbury and Palmer, 2002), English vocabulary and special AMR keywords. The edges in the graph are semantic relations between the concepts which represent the core semantic structure of predicates in the sentence taken from PropBank and additional semantic relations defined within AMR annotation guidelines. The graph structure is necessary because the same concept can be part of multiple relations, i.e., namely reentrancy.

AMR captures "*who is doing what to whom*" in a sentence. It aims to abstract away from the syntactic idiosyncrasies and does not represent every individual word in the sentence. Given this abstract nature of AMR, it does not put any constraints about how it needs to be processed and it can be used to represent any number of sentences with close meanings. Table 3 shows an example of the AMR graph representing several sentences of the same meaning.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

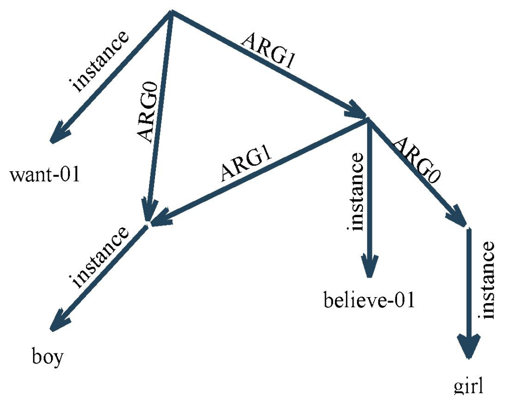
• The boy wants the girl to believe him. • The boy wants to be believed by the girl. • The boy desires the girl to believe him. • The boy desires to be believed by the girl. • The boy has a desire to be believed by the girl. • The boy’s desire is for the girl to believe him.	

Table 3: AMR example

3.2 Cross-Lingual AMR

One of the main aims of AMR is to abstract away from syntax and for this reason AMR is *unanchored*, i.e., the linkage between the words in the sentence and nodes in the graph is not explicitly annotated. This makes AMR adequate to be explored in representing the meaning across languages. Damonte and Cohen (2018) presented the task of cross-lingual AMR parsing which aims at representing sentences in any language with the AMR graph of their English translation, as a universal representation. To this end, the AMR graph contains concepts drawn from PropBank framesets (Kingsbury and Palmer, 2002), English vocabulary and special AMR keywords in the same way as the AMR parsing task.

Referring to the example in Table 3, the same AMR graph is used to represent the meaning of the sentences in other languages as follows:

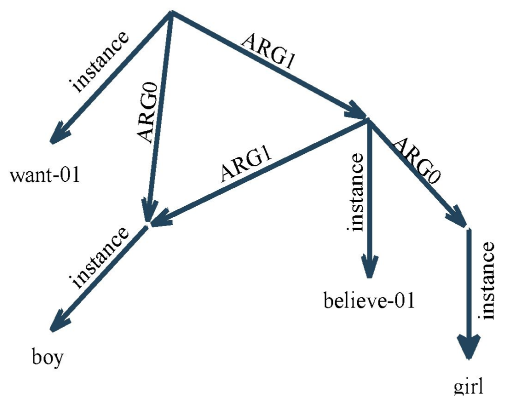
• The boy wants the girl to believe him. • El niño quiere que la niña le crea. **(ES)** • Il ragazzo vuole che la ragazza gli creda. **(IT)** • Der Junge möchte, dass das Mädchen ihm glaubt. **(DE)** • 男孩要女孩相信他。**(ZH)**	

Table 4: Cross-lingual AMR example

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

However, since AMR is designed to be biased towards English and not an interlingua, research in AMR is concentrated mainly in English. Therefore, the existing resources and tools to process text into AMR graphs and vice versa are available mostly in English. To this end, the task of cross-lingual AMR parsing is more challenging as it requires to develop novel cross-lingual parsers and large enough training data (preferably without manual effort) to train a supervised system.

4 Challenges of Cross-Lingual AMR Parsing

4.1 AMR Alignments

Since AMR is not anchored, there exists no clear linkage between the words in the sentence and nodes in the graph. For this reason, it is necessary to identify the concepts in the graph given a sequence of words. English AMR parsers in the literature often rely on word-to-node AMR alignments to identify the concepts in the graph (Flanigan et al., 2014, Damonte et al., 2017, Lyu and Titov, 2018). These alignments are automatically created using heuristics (Flanigan et al., 2014) or pretrained aligners (Pourdamghani et al., 2014; Liu et al., 2018), and are referred to as explicit AMR alignments. Others instead, consider concept identification as a generation task where the generator is allowed to copy words in the sentence and place them as concepts of the graph using attention mechanisms (Zhang et al., 2019), referred to as implicit AMR alignments.

However, both explicit and implicit AMR alignments take advantage from the similarity between the AMR concepts and the English vocabulary (Pourdamghani et al., 2014). For example, in AMR 2.0, a standard manually annotated AMR dataset for English, roughly 60% of the nodes are English words. In addition, PropBank predicates are often similar to English words, e.g., one can heuristically align *publish-01* to *publish*. This similarity does not hold at large, and AMR alignments are hard to be projected across languages through English without introducing a lot of noise (Damonte and Cohen, 2018).

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

4.2 Cross-Lingual AMR Annotations

Cross-lingual properties of AMR have been mainly studied on the scope of annotation analysis for which researchers manually annotated a limited number of sentences which are not enough for training high performance parsers. In this direction, Damonte and Cohen (2018) produce cross-lingual silver AMR annotations by exploiting parallel sentences selected from the Europarl corpus (Koehn, 2005): English sentences are parsed using an English parser (Damonte et al., 2017, AMREager) and the resulting graphs are associated with the corresponding parallel sentences. However, the data on which AMREager was trained is very different from those used to produce the silver annotations, thus the parsing errors highly affect the quality and reliability of the produced AMR graphs. Using the automatically created data only to train a cross-lingual parser leads to low performance as they introduce noise in the training process (Damonte and Cohen, 2018).

4.3 Translation divergences

Some preliminary studies showed the limits of AMR as an interlingua, categorizing them as due to distinctions in the underlying ontologies or structural divergences among languages (Xue et al., 2014; Hajič et al., 2014). Language-specific ontologies might express similar meanings with different lexicalizations of the predicates. The semantics of a verb plays a crucial role in understanding the meaning of a sentence as it defines the argument structure of "*who did what to whom*", therefore verbal resource distinctions could lead to substantially non-parallel structures for parallel meanings across languages. Translation divergences arise when source and target languages have different lexical and syntactic ordering properties. Moreover, some aspects of meaning across languages are lacking in English, therefore it is necessary to observe if an English-centric AMR is able to abstract away from these distinctions and most importantly, to devise neural algorithms that properly handle translation divergences and achieve high performance despite them.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

5 Enabling Cross-Lingual AMR Parsing

We tackle the problems listed above with XL-AMR (Blloshmi et al., 2020), a cross-lingual AMR parser that disposes of AMR alignments, and by creating silver data of diverse quality to train it. The software is available at https://github.com/SapienzaNLP/xl-amr.

5.1 XL-AMR Model

Similar to state-of-the-art models in literature, XL-AMR employs a two-stage approach: concept identification and relation identification. The first step consists of obtaining a sequence of AMR concepts for a given sentence in any language. To remove the need for AMR alignments, we view the task as a sequence generation task and employ an encoder-decoder model (see Sec. 4.1.1). For the second step, that of relation identification, on top of the concept identification module, which determines whether there is a relation between the previously generated concept nodes, and assigns a semantic label to each identified relation (see Sec. 4.1.2). The model is trained to jointly minimize the loss of reference nodes and edges.

5.1.1 Concept Identification

Given a sentence $S_L = \{w_1, w_2, \ldots., w_n\}$ in language L, the goal is to generate a sequence of concepts $C = \{c_1, c_2, \ldots, c_m\}$ where $c \in$ PropBank framesets $\cup$ English vocabulary $\cup$ AMR keywords. The concept identification step is modeled as a seq2seq problem. It employs a BiLSTM encoder-decoder architecture that, in addition to the standard procedure of generating an output token from the output vocabulary, it can copy from the previously predicted nodes sampling from a decoder attention distribution. We use this node-copy mechanism as it was shown to be beneficial in English AMR parsing for handling coreference resolution, i.e., reentrancies (Zhang et al., 2019). To achieve this, at each decoder timestep, we compute an attentional vector which combines encoder context and current decoder representation. Then it is passed through a softmax switch to compute the probability of copying or generating a new token from the output vocabulary. These probabilities are then used to compute the final distribution from which is sampled the next node to be predicted.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

At training time we obtain the list of nodes by first converting the graph into a tree, duplicating the nodes occurring in multiple relations, i.e., reentrant nodes, and then using a pre-order traversal over the tree. To account for reentrancies we assign a unique index to each node during traversal.

5.1.2 Relation Identification

This step is inspired by the arc-factored approaches employed in dependency parsing (Kiperwasser and Goldberg, 2016), i.e., searching for the maximum-scoring connected subgraph over the concepts identified in the previous step. A score for each possible edge is learned through a deep biaffine attention classifier (Dozat and Manning, 2017) which takes as input the decoder states and factorizes the edge prediction in two components; predicting i) whether there is an edge between a pair of nodes, and ii) the edge label for each possible edge, respectively. Then, given the list of predicted nodes $C = \{c_1, c_2, ..., c_m\}$ and a score for candidate edges, we search for the highest-scoring spanning tree using the Chu-Liu-Edmonds algorithm. We then merge the duplicate nodes based on the node indices to restore the final AMR graph.

5.2 Silver Data Creation

To train XL-AMR we exploit transfer learning and create silver data of diverse quality via annotation projections: i) through parallel sentences, in which case the quality of the sentences is gold, since the sentences are human translated into other languages while the quality AMR graphs is silver as they are automatically created, and ii) through machine translation, in which case the quality of the AMR graphs is gold drawn from a manually annotated English AMR dataset and the quality of the sentences is silver as the sentences of the annotated corpus are automatically translated into other languages.

5.2.1 Through Parallel Sentences (ParSentsSilverAMR)

For this approach we use a sample of sentences from Europarl corpus which contains parallel sentences for an English (EN) sentence in several languages, including our languages of interest, German (DE), Italian (IT) and Spanish (ES). To obtain the AMR graph of the

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

English sentence we use a pretrained English AMR parser from the literature (Zhang et al., 2019). Different from Damonte and Cohen (2018), which also use the same set of sentences from Europarl, we do not need to align nodes of the graph with words in the sentence as XL-AMR does not need this linkage.

In Figure 1, we illustrate the process of creating silver data through parallel sentences. Given an English sentence and its translations in other languages, we first parse the English using the AMR parser and then associate the produced graph with all the parallel sentences in other languages.

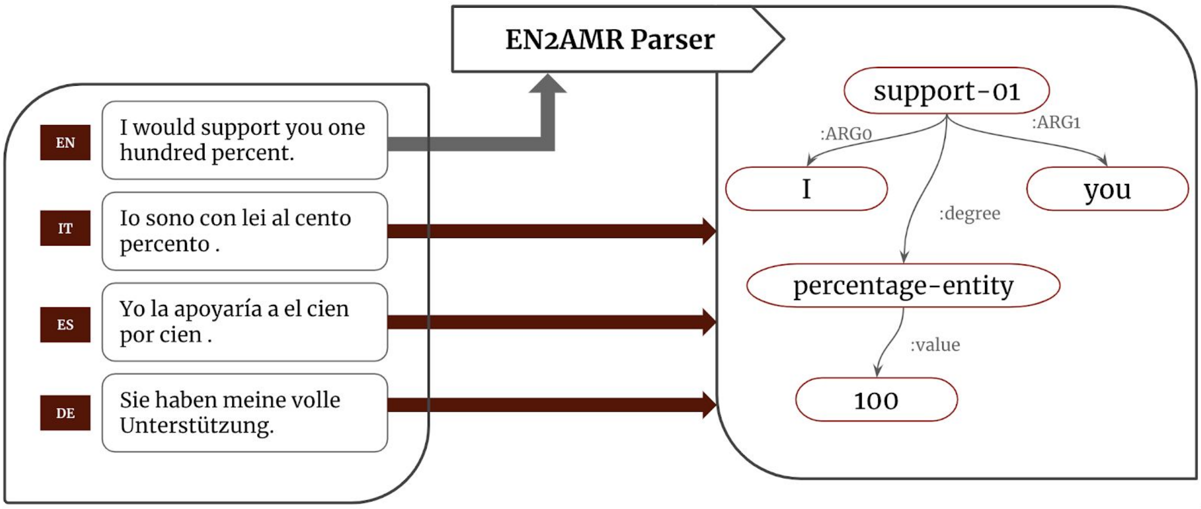


Figure 1: Annotation Projection through Parallel Sentences

5.2.2 Through Machine Translated Sentences (GoldAMRSilverTrans)

For this approach we are given an annotated English AMR corpus and machine translation models for the languages of interest. We use AMR 2.0 (LDC2017T10)[1] and the available machine translation models from the literature of OpusMT (Tiedemann and Thottingal, 2020) for German, Italian and Spanish and MASS (Song et al., 2019a) for Chinese.

In Figure 2, we illustrate the steps of this approach. The first step is to translate the English sentence into all the languages. To ensure quality we translate back to English and perform a filtering step that filters out the sentences for which the English translation is not close to the

[1] https://catalog.ldc.upenn.edu/LDC2017T10

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

original sentence. Finally we associate the sentences which pass the filtering step with the gold AMR graph of the original English sentence.

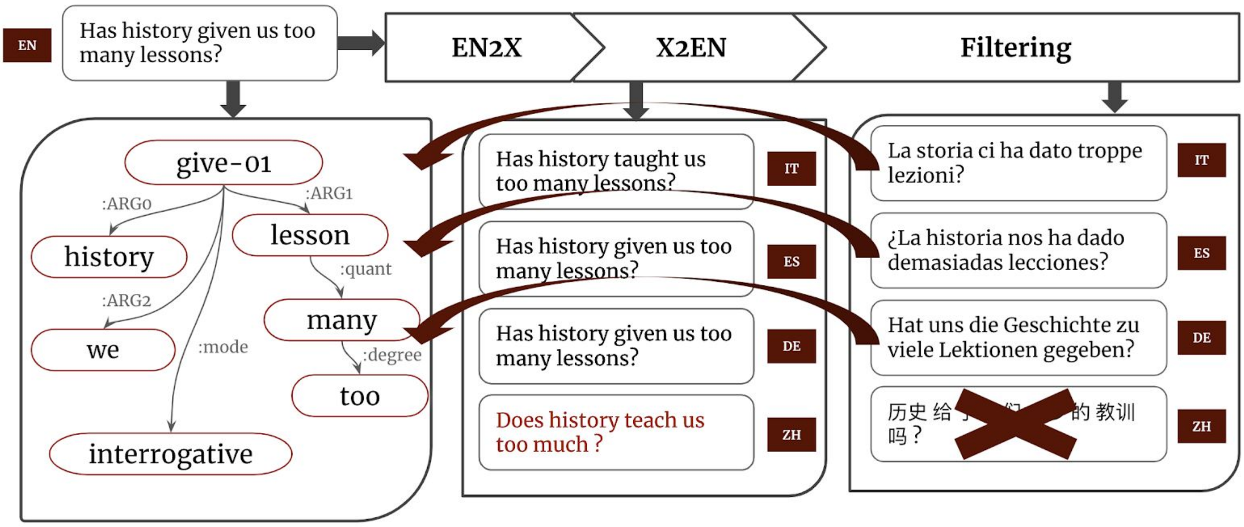


Figure 2: Annotation Projection through Machine Translated Sentences

For the filtering step we use the LASER (Artetxe and Schwenk, 2019b) sentence embedding model from the literature to embed the original English sentence and the backtranslated sentence. Then we perform a 1st Nearest Neighbor (1-NN) algorithm that returns the closest original sentence for each back-translated sentence based on the cosine similarity of sentences' LASER representations. All sentences for which the nearest neighbor is not their corresponding original sentence, are filtered out from the corpus.

5.2.3 Dataset Statistics

In Table 5, we show the number of training and development set instances, their quality and the source where we take the sentences and AMR graphs from.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

Dataset	Lang	Train Insts	Dev Insts	Source
Gold	EN	36521	1368	AMR 2.0
PARSENTS SILVERAMR	DE	20000	2000	Europarl
	EN	20000	2000	Europarl
	ES	20000	2000	Europarl
	IT	20000	2000	Europarl
GOLDAMR SILVERTRNS	DE	34415	1319	AMR 2.0
	ES	34552	1325	AMR 2.0
	IT	34521	1322	AMR 2.0
	ZH	32154	1276	AMR 2.0

Table 5: Dataset statistics.

6 Experimental Setup

In order to assess the effectiveness of XL-AMR model when trained on our created silver data through different approaches, we conduct a set of experiments with different XL-AMR variants and compare with another cross-lingual AMR parser from the literature. Finally, we show how XL-AMR handles translation divergences between languages.

6.1 Evaluation Benchmark

We evaluate on the Abstract Meaning Representation 2.0 - Four Translations[2] (Damonte and Cohen, 2020), a corpus containing translations of the test split of 1371 sentences from the AMR 2.0 (LDC2017T10), in Chinese (ZH), German (DE), Italian (IT) and Spanish (ES). This data is designed for use in cross-lingual AMR parsing.

6.2 Comparison System

We compare the results of the XL-AMR variants with the projection method of AMREager multilingual (henceforth AMREager) on the gold dataset, i.e., AMR 2.0 - Four Translations. It is an existing cross-lingual AMR parser from the literature which employs a transition-based parser and relies on AMR alignments projecting them through English.

[2] https://catalog.ldc.upenn.edu/LDC2020T07

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

6.3 XL-AMR Variants

We train different XL-AMR variants, as follows:

1. Zero-shot (Ø-shot) - trained on English data with language independent features;
2. Language-specific - trained on data from the target language only;
3. Multilingual - trained on data from all the available languages;
4. Bilingual - trained on a combination of English and target language data.

7 Results

To evaluate the performance of the models we use Smatch[3] F1 which computes the degree of overlap of two AMR graphs (Cai and Knight, 2013) as the overlap of their triples, i.e, (concept, relation, concept).

In Figure 3 we show the previously reported results by AMREager and only the performances of our best performing models per method which are obtained by the language-specific variant in Chinese and the bilingual variants on the other languages (the performances of all the variants are shown in Table 6).

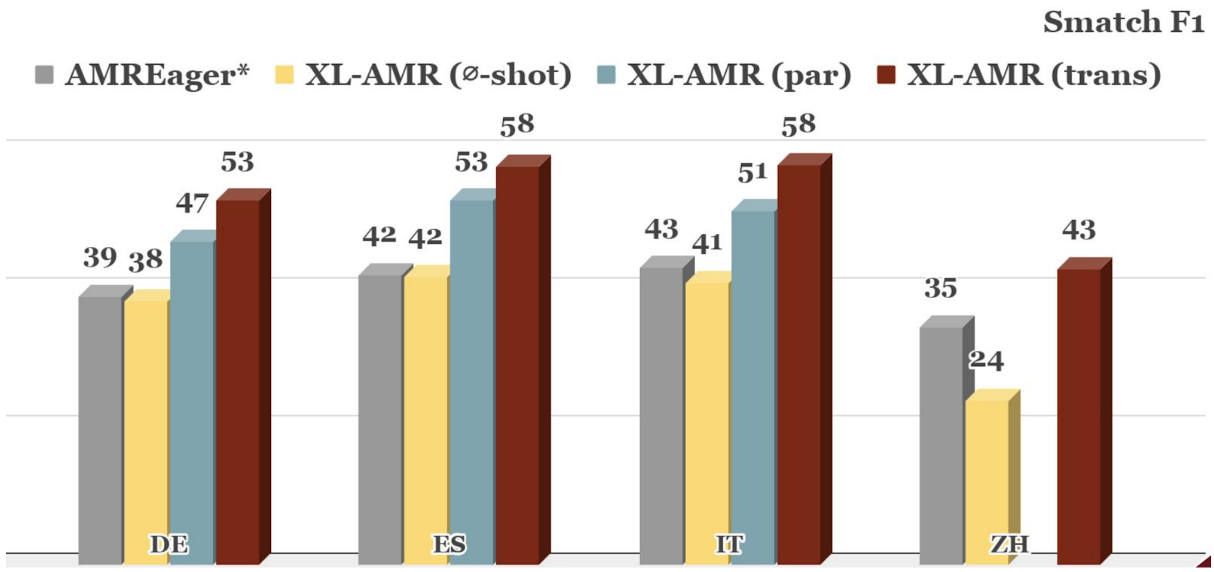


Figure 3: Smatch F1 score of AMREager and XL-AMR models for German (DE), Spanish (ES), Italian (IT) and Chinese (ZH). XL-AMR (par) is the model trained on data from the annotation projection from parallel sentences approach while XL-AMR (trans) refers to the model trained on data created with the annotation projection through machine translated sentences.

[3] https://github.com/snowblink14/smatch

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

From the chart we see that Ø-shot XL-AMR performs on par with AMREager or worse. However, the variant of XL-AMR trained on data from annotation projection through parallel sentences (XL-AMR par) outperforms AMREager by around 8 points per language. Since both are trained on the same set of sentences from Europarl, this improvement is attributed to the disposal of AMR alignments from XL-AMR.

Finally XL-AMR variant trained on data from annotation projection through machine translated sentences (XL-AMR trans) performs best, outperforming the parallel sentences counterpart (XL-AMR par) by 5 to 7 points and AMREager by 8 to 16 points depending on the language. This result is attributed not only to the disposal of AMR alignments but also to the better quality training data.

In Table 6 we show the detailed performances obtained from all the variants of XL-AMR.

Parser	Configuration	DE	ES	IT	ZH
AMREAGER	Lang-Spec.	39.0	42.0	43.0	35.0
XL-AMR$_{\emptyset}^{amr}$	$\emptyset$-shot	32.7	39.1	37.1	25.9
XL-AMR$_{\emptyset}^{par+}$	$\emptyset$-shot	38.3	41.8	41.0	23.9
XL-AMR^{par}	Lang-Spec.	40.8	44.2	43.4	-
	Multiling.	41.5	45.6	45.0	-
	Biling.	42.7	47.9	46.7	-
XL-AMR^{par+}	Multiling.	46.3	51.2	50.9	-
	Biling.	47.0	53.0	51.4	-
XL-AMR^{trans}	Lang-Spec.	51.6	56.1	56.7	**43.1**
	Multiling.	49.9	53.0	54.0	40.0
	Multiling. (-ZH)	51.5	55.5	55.9	-
XL-AMR^{trans+}	Multiling.	49.9	53.2	53.5	41.0
	Multiling. (-ZH)	52.1	56.2	56.7	-
	Biling.	**53.0**	**58.0**	**58.1**	41.5

Table 6: Smatch F1 scores for all the different variants of XL-AMR on DE, ES, IT and ZH. Best scores per language are denoted in **bold**.

8 Translation Divergences

Translation divergences arise when source and target languages have different lexical and syntactic ordering properties. Moreover, some aspects of meaning across languages are lacking in English, therefore it is necessary to observe if an English-centric AMR is able to

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

do away from these distinctions and most importantly, to devise neural algorithms that properly handle translation divergences and achieve high performance despite them.

Dorr (1994) categorized divergences across languages as: i) *thematic*, ii) *promotional*, iii) *demotional*, iv) *structural*, v) *conflational*, vi) *categorial*, vii) *lexical*.

Through a qualitative analysis we manually check predictions of XL-AMR on sentences in which we observe any of these categories.

Consider a *categorical* divergence which happens when the same meaning is expressed by different syntactic categories across languages. In Table 7 we have the Italian and Spanish translation of the English sentence “I agree” in which the verb *agree* is expressed by a noun, *accordo* and *acuerdo*. The resulting AMR graph however, abstracts away from this syntactic divergence and it is the same for all the sentences.

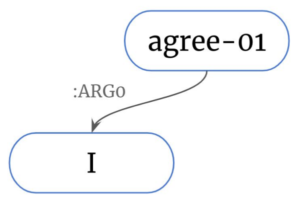
• I agree. (EN) • Sono d’accordo. (IT) • Estoy de acuerdo (ES)	

Table 7: Categorical divergence

In Table 8 we show an example of a *structural* divergence which arises when a verbal object is realized as a noun phrase in one language and as a prepositional phrase in the other. The sentence I saw John is translated in Spanish as if it was *I saw to John*. XL-AMR overcomes this divergence as well producing the same graph for both.

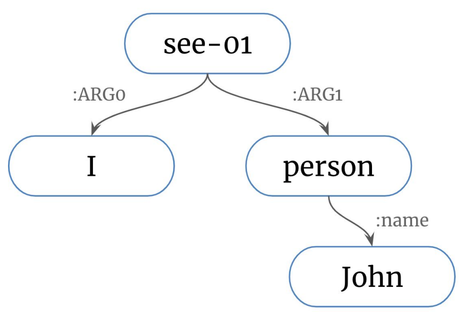
• I saw John. (EN) • Vi a John (ES) *(I saw to John)*	

Table 8: Structural divergence

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

Similarly, from our complete analysis in all the categories we saw that XL-AMR overcomes divergences from *thematic*, *promotional*, *demotional*, *conflational* categories as well, with exception of *lexical* divergence, arising when a verb is translated with a different lexical verb across languages. An example of the latter is shown in Table 9. The verb *break into* is translated in Spanish to *force the entrance*. While the AMR graphs produced for both the sentences are valid, they are not parallel to each other despite their parallel meaning. This arises the need for a higher level of abstraction to handle lexical divergences.

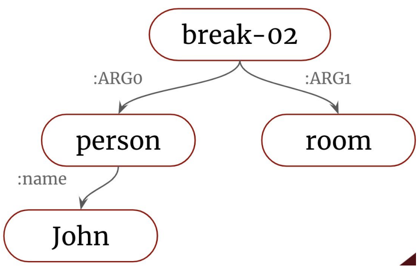
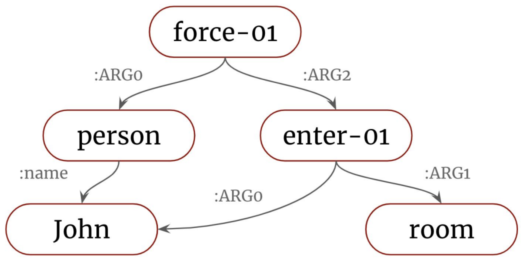
John broke into the room. (EN)	John forzó la entrada al cuarto. (ES) *(John forced the entrance into the room)*
	

Table 9: Lexical divergence

9 Discussion and Future Work

The main insights we get from the experiments in cross-lingual AMR parsing are two-fold:

i) the need for creating higher quality data that have parallel meaning across language is crucial for further improving the performance of the cross-lingual AMR and

ii) the need for a higher level of abstraction to handle lexical divergences among languages to obtain parallel meaning representations. The following resources have a high potential for satisfying the foregoing needs which we intend to exploit in the future towards the goal of this task.

9.1 ELEXIS Dictionary Matrix

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

The availability of an interlinked resource among languages that brings together definitions and examples with comparable meanings across languages could be a great source for automatically creating higher quality training data for supervised multilingual semantic parsers. As future work, we would like to exploit the lexicographers linked data which besides the quality, offer wider coverage across more languages than what we include in this report. This would satisfy the first above-mentioned need.

9.2 VerbAtlas Exploitation

Another need that arises from the qualitative analysis that we perform is that for a higher level of abstraction to handle lexical divergences across languages which, we recall, happens when a verb is translated with a different lexical verb across languages. AMR is a rooted graph, the root of which is often a predicate, i.e, verb, therefore different lexical verbs across languages lead to undesirable different meaning structures for sentences with parallel meaning. To handle this problem we plan to exploit VerbAtlas. Within VerbAtlas, different lexicalizations of the same verb are clustered together within the same frame and have the same argument structure. Moreover since VerbAtlas is connected to BabelNet, it further clusters together different lexicalizations of verbs across languages. This, in fact, could be what we need to overcome the problem of obtaining non-parallel structures for parallel meaning due to lexical distinctions across languages.

We believe that VerbAtlas, alongside better quality training corpora created by exploiting the lexicographer data, could be a stepping stone towards a unified meaning representation across languages and the goal of Task 2.3.

10 Conclusions

During the first part of the project, in task T3.2 we developed VerbAtlas and addressed the key task of multilingual semantic parsing by enabling high-performance cross-lingual AMR. VerbAtlas is a large-coverage manually-crafted verbal resource which despite its advantages in clustering the predicates into semantically-coherent frames with explicit cross-frame semantic roles, it is scalable across languages thanks to its linkage with BabelNet and Open Multilingual WordNet. We believe this resource can be crucial for future developments

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

towards multilingual semantic representations. In multilingual semantic parsing direction, we tackled the paucity of training data for the task. We explored transfer learning techniques to enable high performance cross-lingual AMR parsing and get a step closer towards multilingual semantic parsing without relying on manually created multilingual data. We created silver data based on annotation projection through parallel sentences and machine translation, on which we trained XL-AMR, a cross-lingual AMR parser that achieves the highest results reported to date on Chinese, German, Italian and Spanish. XL-AMR overcomes most of the translation divergences with the exception of the lexical divergence that persists despite the parser predicting a valid graph. The latter divergence results in non-parallel structures for parallel meanings, and we intend to exploit VerbAtlas for unifying synonyms or related meanings within the AMR formalism. Most importantly, through a set of experiments, we show that it is possible to achieve higher performances and spread across languages easily by paying careful attention to the cross-lingual resources and techniques employed. In fact, we believe that key to the improvement of system performance for the cross-lingual semantic parsing can be an interconnected network of resources for better coverage of different meaning aspects across languages.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

11 References

Omri Abend and Ari Rappoport. 2013. Universal con- ceptual cognitive annotation (UCCA). In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 228–238, Sofia, Bulgaria.

Mikel Artetxe and Holger Schwenk. 2019b. Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. Transactions of the Association for Computational Linguistics, 7:597–610.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In Proceedings of the 17th International Conference on Computational Linguistics, pages 86–90. Association for Computational Linguistics.

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, Rocco Tripodi, and Roberto Navigli. 2020. XL-AMR: Enabling Cross-Lingual AMR Parsing with Transfer Learning Techniques. In Proceedings of the 2020 Empirical Methods for Natural Language Processing (EMNLP2020).

Francis Bond and Ryan Foster. 2013. Linking and extending an open multilingual Wordnet. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, volume 1, pages 1352–1362.

Shu Cai and Kevin Knight. 2013. Smatch: an evaluation metric for semantic feature structures. In Proceedings of the 51st Annual Meeting of the Associa- tion for Computational Linguistics (Volume 2: Short Papers), pages 748–752, Sofia, Bulgaria.

Marco Damonte and Shay Cohen. 2020. Abstract Meaning Representation 2.0 - Four Translations LDC2020T07. Web Download, Philadelphia: Lin- guistic Data Consortium.

Marco Damonte and Shay B. Cohen. 2018. Cross-lingual abstract meaning representation parsing. 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 1146–1155, New Orleans, Louisiana.

Marco Damonte, Shay B. Cohen, and Giorgio Satta. 2017. An incremental parser for Abstract Meaning Representation. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, pages 536–546, Valencia, Spain.

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.

Timothy Dozat and Christopher D. Manning. 2017. Deep biaffine attention for neural dependency parsing. In 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

Jeffrey Flanigan, Sam Thomson, Jaime Carbonell, Chris Dyer, and Noah A. Smith. 2014. A discriminative graph-based parser for the abstract meaning representation. In Proceedings of the 52nd Annual Meeting of the Association for Computational Lin- guistics (Volume 1: Long Papers), pages 1426–1436, Baltimore, Maryland.

Hajič, J., Hajicová, E., Panevová, J., Sgall, P., Bojar, O., Cinková, S., Fucíková, E., Mikulová, M., Pajas, P., Popelka, J., Semecký, J., Sindlerová, J., Stepánek, J., Toman, J., Uresová, Z., & Žabokrtský, Z. (2012). Announcing Prague Czech-English Dependency Treebank 2.0. LREC.

Jan Hajič, Ondřej Bojar, and Zdeňka Urešová. 2014. Comparing Czech and English AMRs. In Proceedings of Workshop on Lexical and Grammatical Resources for Language Processing, pages 55–64, Dublin, Ireland.

Hardy Hardy and Andreas Vlachos. 2018. Guided neural language generation for abstractive summarization using Abstract Meaning Representation. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 768– 773, Brussels, Belgium.

Paul Kingsbury and Martha Palmer. 2002. From Tree- Bank to PropBank. In Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02), Las Palmas, Canary Islands - Spain.

Karin Kipper-Schuler. 2005. VerbNet: A broad-coverage, comprehensive verb lexicon. University of Pensylvania.

Eliyahu Kiperwasser and Yoav Goldberg. 2016. Simple and accurate dependency parsing using bidirectional LSTM feature representations. Transactions of the Association for Computational Linguistics, 4:313–327.

Philipp Koehn. 2005. Europarl: A Parallel Corpus for Statistical Machine Translation. In Conference Proceedings: the tenth Machine Translation Summit, pages 79–86, Phuket, Thailand. AAMT.

Kexin Liao, Logan Lebanoff, and Fei Liu. 2018. Abstract Meaning Representation for multi-document summarization. In Proceedings of the 27th International Conference on Computational Linguistics, pages 1178–1190, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Roberto Navigli and Simone Paolo Ponzetto. 2010. BabelNet: Building a very large multilingual semantic network. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pages 216–225. Association for Computational Linguistics.

Stephan Oepen and Jan Tore Lønning. 2006. Discriminant-based MRS banking. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy.

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

Nima Pourdamghani, Yang Gao, Ulf Hermjakob, and Kevin Knight. 2014. Aligning English strings with abstract meaning representation graphs. In Proceed- ings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 425–429, Doha, Qatar.

James Pustejovsky. 1995. The Generative Lexicon. MIT Press, Cambridge MA.

Sudha Rao, Daniel Marcu, Kevin Knight, and Hal Daume´ III. 2017. Biomedical event extraction using Abstract Meaning Representation. In BioNLP 2017, pages 126–135, Vancouver, Canada.

Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. 2019a. MASS: masked sequence to sequence pre-training for language generation. In Proceedings of the 36th International Conference on Machine Learning,

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

ICML 2019, 9-15 June 2019, Long Beach, California, USA, volume 97 of Proceedings of Machine Learning Research, pages 5926–5936.

Linfeng Song, Daniel Gildea, Yue Zhang, Zhiguo Wang, and Jinsong Su. 2019b. Semantic neural machine translation using AMR. *Transactions of the Association for Computational Linguistics*, 7:19–31.

Jörg Tiedemann and Santhosh Thottingal. 2020. OPUS-MT – building open translation services for the world. In Proceedings of the 22nd Annual Con- ference of the European Association for Machine Translation, pages 479–480, Lisboa, Portugal.

Aaron Steven White, Drew Reisinger, Keisuke Sakaguchi, Tim Vieira, Sheng Zhang, Rachel Rudinger, Kyle Rawlins, and Benjamin Van Durme. 2016. Universal decompositional semantics on universal dependencies. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1713–1723, Austin, Texas.

Sheng Zhang, Xutai Ma, Kevin Duh, and Benjamin Van Durme. 2019. AMR parsing as sequence-to- graph transduction. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 80–94, Florence, Italy.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.