

SGL: Speaking the Graph Languages of Semantic Parsing via Multilingual Translation

Luigi Procopio Rocco Tripodi Roberto Navigli

Sapienza NLP Group

Department of Computer Science, Sapienza University of Rome

{procopio, tripodi, navigli}@di.uniroma1.it

Abstract

Graph-based semantic parsing aims to represent textual meaning through directed graphs. As one of the most promising general-purpose meaning representations, these structures and their parsing have gained a significant interest *momentum* during recent years, with several diverse formalisms being proposed. Yet, owing to this very heterogeneity, most of the research effort has focused mainly on solutions specific to a given formalism. In this work, instead, we reframe semantic parsing towards multiple formalisms as Multilingual Neural Machine Translation (MNMT), and propose SGL, a many-to-many *seq2seq* architecture trained with an MNMT objective. Backed by several experiments, we show that this framework is indeed effective once the learning procedure is enhanced with large parallel corpora coming from Machine Translation: we report competitive performances on AMR and UCCA parsing, especially once paired with pre-trained architectures. Furthermore, we find that models trained under this configuration scale remarkably well to tasks such as cross-lingual AMR parsing: SGL outperforms all its competitors by a large margin without even explicitly seeing non-English to AMR examples at training time and, once these examples are included as well, sets an unprecedented state of the art in this task. We release our code and our models for research purposes at <https://github.com/SapienzaNLP/sgl>.

1 Introduction

Being able to associate natural language text with well-defined and machine-actionable meaning representations, i.e. the task of semantic parsing (SP), is one of the holy grails in Natural Language Processing (NLP) and Understanding (Navigli, 2018). Considering how a breakthrough in this direction would empower NLP systems to explicitly *make sense* of natural language, the ever-growing interest semantic parsing has been receiving really comes as no surprise. Graph-based

formalisms such as Abstract Meaning Representation (Banarescu et al., 2013, AMR), Elementary Dependency Structures (Oepen and Lønning, 2006, EDS), Prague Tectogrammatical Graphs (Hajič et al., 2012, PTG), Universal Conceptual Cognitive Annotation (Abend and Rappoport, 2013, UCCA), *inter alia*, are emerging as the *de facto* standard for general-purpose meaning representations and have shown potential in Machine Translation (Song et al., 2019), Text Summarization (Hardy and Vlachos, 2018), Human-Robot Interaction (Bonial et al., 2020), and as evaluation metrics (Sulem et al., 2018; Xu et al., 2020b). These formalisms propose encoding meaning through directed graphs, however, each of them builds upon different linguistic assumptions, aims to target different objectives and, at a more practical level, assigns different functions to nodes and edges. For instance, while AMR uses nodes to encode concepts and edges to express the semantic relations between them, UCCA proposes using text tokens as terminal nodes and building graphs on top of them.

As a result of this heterogeneous landscape, often referred to as *framework-specific balkanization* (Oepen et al., 2020), graph-based semantic parsing has seen a proliferation of framework-specific solutions. However, approaches capable of competitively scaling across formalisms represent a natural *desideratum*, and recent works have started to explore this direction, examining the usage of multi-task learning in different architectures (Herscovich et al., 2018; Oepen et al., 2019), or casting different formalisms under a unified framework where models can be trained to perform graph transduction (Zhang et al., 2019b). Nevertheless, despite achieving promising results, research in this direction has been hindered by the general lack of training data that afflicts semantic parsing. Indeed, due to the inherent complexity of this task, annotated corpora are still scarce and prohibitively expensive to expand.

In this work, we set ourselves to address these issues and, in particular, we propose *Speak the Graph Languages* (SGL), a many-to-many *seq2seq* architecture which we show to competitively scale across formalisms and across languages.¹ The key idea is to train a *seq2seq* model with a Multilingual Neural Machine Translation (MNMT) objective, where, given an input text and an identifier denoting the desired output formalism, a single shared model has to learn to translate towards the corresponding linearized graph. We use AMR and UCCA as our cases in point to show the effectiveness of this framework. In particular, we show that, once the learning procedure also considers large parallel corpora coming from Machine Translation (MT), this configuration becomes an effective approach for framework-independent parsing via a single model. Even more interestingly, this model scales surprisingly well to cross-lingual parsing and is capable of navigating through translation paths like IT \rightarrow AMR,² which it has never seen during training. The contributions of this work are therefore as follows:

- We reframe semantic parsing towards multiple formalisms and from multiple languages as multilingual machine translation;
- On AMR parsing, our framework achieves competitive performances, surpassing most of its current competitors once paired with a pre-trained Transformer;
- We outperform all current alternatives in cross-lingual AMR parsing without ever seeing non-English to AMR examples at training time and push the current state of the art even further once we include these examples;
- On UCCA parsing, we reach competitive results, outperforming a strong BERT-powered baseline (Hershcovich and Arviv, 2019).

We release our code and our models for research purposes.

2 Related Work

Our work is mainly concerned with semantic parsing in UCCA and AMR, considering also the cross-

¹By *across languages*, we mean that the model is capable of performing cross-lingual semantic parsing as defined for AMR by Damonte and Cohen (2018). Unless otherwise specified, we will follow this perspective throughout the paper.

²IT stands for Italian.

lingual setting introduced by Damonte and Cohen (2018) for the latter.

Semantic Parsing Arguably among the formalisms that have drawn the most interest, AMR has seen the emergence of a rich yet dedicated literature, with recent approaches that can be roughly clustered into two groups. On the one hand, several graph-based solutions have been proposed (Lyu and Titov, 2018; Zhang et al., 2019a,b; Zhou et al., 2020; Cai and Lam, 2020); among these solutions, Zhou et al. (2020) show the effectiveness of enhancing an aligner-free parser with latent syntactic information, whereas Cai and Lam (2020) present an iterative method to build and refine AMR graphs incrementally. On the other hand, translation-based approaches, where *seq2seq* models are trained to translate from natural language text to linearized graphs, have been shown to reach competitive performances, despite the scarcity of training data (Konstas et al., 2017; van Noord and Bos, 2017; Ge et al., 2019). Continuing this latter direction and arguably closest to our work, Xu et al. (2020a) and Bevilacqua et al. (2021) show that these models, once paired with adequate pre-training, can behave on par or better than dedicated and more sophisticated graph-based alternatives, surpassing the performances of Cai and Lam (2020). In particular, similarly to our work, Xu et al. (2020a) leverage a multilingual framework inspired by Johnson et al. (2017) and explore the possibility of pre-training on a range of related tasks, including MT; however, their focus is limited to showing the effectiveness of transfer learning from related tasks to English AMR parsing.

Conversely, here we show that the benefits of multilingual *seq2seq* frameworks are not limited to English TEXT-to-AMR but, rather, that they enable astonishing performances on unseen translation paths such as IT \rightarrow AMR and competitive results on other frameworks, using UCCA as our case in point. In this sense, we continue the recent cross-framework trend formally started by the shared task of Oepen et al. (2019), exploring the possibility of using translation-based approaches for framework-independent parsing, as opposed to the transition-based parsers proposed in that seminal work. Our findings are in line with the recent results reported by Oepen et al. (2020) and, in particular, by Ozaki et al. (2020), where the authors cast semantic parsing in multiple formalisms as translation towards a novel Plain Graph Notation

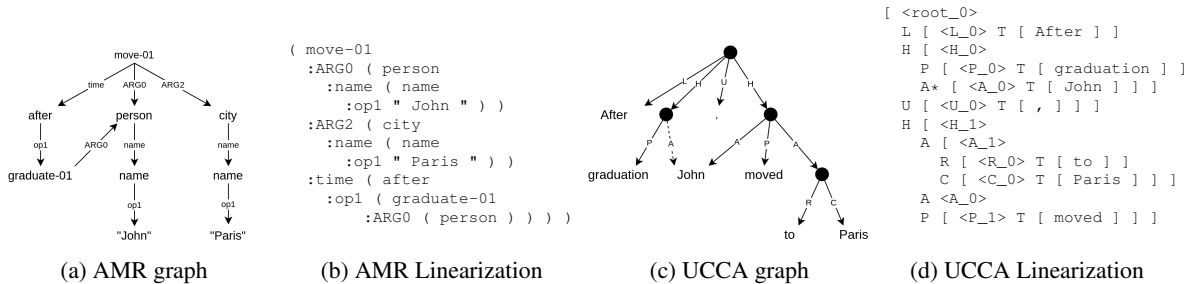


Figure 1: AMR and UCCA graphs, along with their linearizations, for the sentence "After graduation, John moved to Paris". To ease readability, linearizations are shown with newlines and indentation; however, when fed to the neural model, they are in a single-line single-space format.

(PGN) they devise. However, whereas they train different independent models for each framework, we explore the possibility of using a single *multi-lingual* model.

Cross-lingual AMR While most of the research effort in the AMR community has been focused on English only, the seminal work of Damonte and Cohen (2018) gave rise to an interesting new direction, i.e. exploring the extent to which AMR can act as an interlingua. The authors introduced a new problem, cross-lingual AMR parsing, and defined it as the task of recovering, given a sentence in any language, the AMR graph corresponding to its English translation. Using an adapted version of the transition-based parser originally proposed by Damonte et al. (2017) and training it on silver data generated through annotation projection, they examined whether AMR graphs could be recovered starting from non-English sentences. Even though their models fell short when compared to MT alternatives,³ their work showed promising results and suggested that, despite translation divergences, AMR could act effectively as an interlingua.

Annotation projection has been focal in subsequent work as well. Blloshmi et al. (2020) propose an aligner-free cross-lingual parser, thus disposing of the need for word alignments in the annotation projection pipeline; their parser manages to outperform MT alternatives when both annotation projection and these baselines have access to comparable amounts of data. Conversely, Sheth et al. (2021) leverage powerful contextualized word embeddings to improve the foreign-text-to-English-AMR alignments, surpassing all previous approaches and, most importantly, the yet-unbeaten MT baselines that have access to larger amounts of data.

³The input sentence is first translated towards English and, then, an English parser is used.

We stand out from previous research and show that, as a matter of fact, annotation projection techniques are not needed to perform cross-lingual AMR parsing. By jointly training on parallel corpora from MT and the EN \rightarrow SP data we have, we find that a multilingual model can navigate unseen translation paths such as IT \rightarrow AMR effectively, outperforming all current approaches by a significant margin; yet, annotation projection is naturally beneficial and, when its training data are taken into account as well, SGL pushes performances even further.

3 Speak the Graph Languages (SGL)

In this section, we describe SGL, our proposed approach to graph-based semantic parsing. We first explain the graph linearizations we employ for AMR and UCCA, along with their delinearizations (§3.1). We then describe the *seq2seq* modelling approach we use (§3.2) and, finally, we present our multilingual framework (§3.3).

3.1 Graph Linearizations

We now describe how we convert the considered meaning representations into translatable text sequences (*linearization*), along with their reverse process (*delinearization*).

For AMR parsing, as in van Noord and Bos (2017), we first simplify AMR graphs by removing variables and wiki links. We then convert these stripped AMR graphs into trees by duplicating co-referring nodes. At this point, in order to obtain the final linearized version of a given AMR, we concatenate all the lines of its PENMAN notation (Goodman, 2020) together, replacing newlines and multiple spaces with single spaces (Figure 1a and 1b). Conversely, delinearization is performed by assigning a variable to each predicted concept, per-

forming Wikification,⁴ restoring co-referring nodes and, where possible, repairing any syntactically malformed subgraph.⁵ For both phases, we use the scripts released by van Noord and Bos (2017).⁶

For UCCA parsing, we employ a Depth-First Search (DFS) approach: starting from the root, we navigate the graph, using square brackets to delimit subgraph boundaries and special variables to denote terminal and non-terminal nodes; remote edges are denoted by a special modifier appended to their labels, while re-entrancies, that is, edges whose target is a node already seen, are handled by simply entering the respective variable (Figure 1c and 1d). Similarly to AMR, delinearization is performed by back-parsing this sequence into a UCCA graph, repairing malformed subgraphs when possible;⁷ additionally, as terminal nodes are anchored in UCCA, we remove those whose anchoring is impossible. The linearization and delinearization scripts for this schema are released along with the rest of our code.

3.2 Sequence-to-sequence Modelling

We employ neural *seq2seq* models based upon the *Transformer* architecture (Vaswani et al., 2017). This architecture is essentially composed of two building blocks, namely, a *Transformer encoder* and a *Transformer decoder*. The encoder is a stack of N identical layers, each made up of two sub-layers: the first is a multi-head self-attention mechanism, while the second is a position-wise fully connected feed-forward network. The decoder follows a similar architecture, presenting, however, an additional sub-layer that performs multi-head attention over the output of the encoder.

Within this work, we use two different kinds of Transformer architecture, *Cross* and *mBART* (Liu et al., 2020). *Cross* is a randomly initialized Transformer closely following the architecture depicted by Vaswani et al. (2017), except for a significant difference: we leverage a factorized embedding parameterization (Lan et al., 2020), that is, we decompose the large vocabulary embedding matrix into two smaller matrices. While the first of these represents the actual embedding matrix and projects one-hot vectors into an embedding space

⁴We use DBpedia Spotlight API (Daiber et al., 2013).

⁵Although trained to generate syntactically correct graphs, the outputs *seq2seq* models produce may contain syntactic errors, such as brackets that do not match.

⁶<https://github.com/RikVN/AMR>

⁷Should repairing fail, the faulty subgraph is discarded altogether.

whose dimension is lower than the Transformer hidden size, the second one takes care of projecting these intermediate representations towards the actual Transformer hidden space. This technique significantly reduces the number of parameters and, within the context of our experiments, did not show any significant performance penalty.

On the other hand, *mBART* is a multilingual Transformer pre-trained in many languages over large-scale monolingual corpora. As AMR and UCCA are naturally not included among the supported languages in the vocabulary, we apply an architectural change to *mBART* and increase its vocabulary with two new language ids. More specifically, we augment its embedding matrix by adding two additional vectors, which we randomly initialize as in Tang et al. (2021).

3.3 Multilingual Framework

In order to empower our models to support translation from and towards multiple languages, we employ a data-driven approach: we replace the start token of the decoder with a special tag specifying the language the encoder representations should be unrolled towards. Figure 2 shows an example of this schema. It is worth pointing out that, while for *Cross* we do not feed the source language to the encoder, when using the *mBART* model we follow its input format and do provide it.

Once data have been tagged according to this schema, we train a many-to-many translation model on both the semantic parsing and English-centric parallel corpora.⁸ Considering that our focus is on semantic parsing, we perform oversampling on the AMR and UCCA datasets. Furthermore, when considering the parallel corpora from MT, we flip the training direction with probability 0.5, hence allowing our model to see at training time both the $X \rightarrow EN$ and $EN \rightarrow X$ training directions; we argue that this stochastic flip benefits our models in multiple ways:

- As $EN \rightarrow X$ shares the source language with both $EN \rightarrow AMR$ and $EN \rightarrow UCCA$, this results in positive transfer;
- As AMR, UCCA and EN are significantly related, $X \rightarrow EN$ also results in positive transfer (similar target language);

⁸Henceforth, without loss of generality, we will use English as the source language of the MT data and denote by X all the target-side languages.

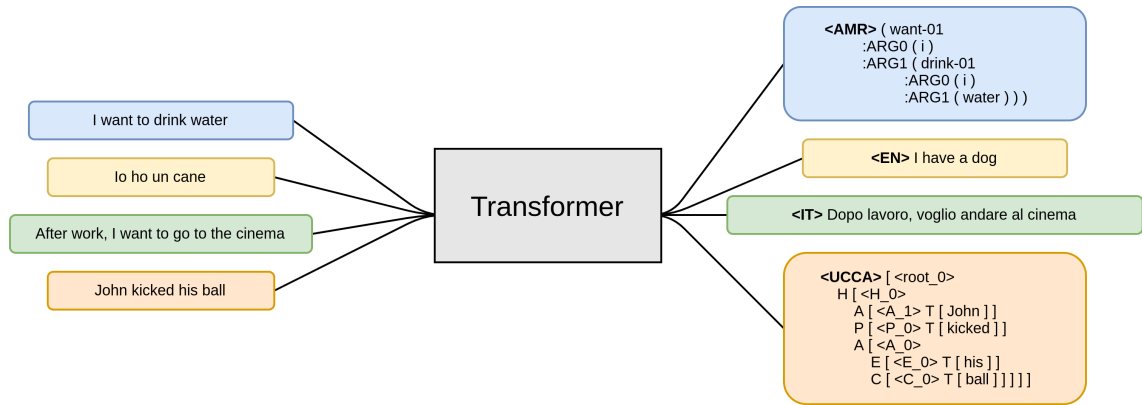


Figure 2: Our SGL multilingual translation framework. Best seen in color.

- Finally, $X \rightarrow EN$ allows our model to navigate unseen translation paths (i.e. *zero-shot*) such as $IT \rightarrow AMR$ and thus tackle tasks like cross-lingual AMR parsing.

4 Experimental Setup

We assess the effectiveness of our proposed approach by evaluating its performance on all translation paths where the target language is a graph formalism, the only exception being $X \rightarrow UCCA$, with X any language but English. This choice is motivated by the fact that, differently from AMR where cross-lingual AMR aims to produce English-based meaning representations (Damonte and Cohen, 2018), UCCA builds graphs on top of its tokens which are, consequently, inherently in the same language as the input text (Hershcovich et al., 2019); we leave exploring this direction to future work.

4.1 Models

We choose to use both *Cross*, a randomly initialized Transformer, and *mBART*, a multilingual pre-trained Transformer, to better grasp the effects of this joint multilingual framework in different regimes. In particular, we explore the following configurations:

- models trained only on a single semantic parsing task (AMR or UCCA parsing) and without considering any parallel data, denoted by $Cross_{st}$ and $mBART_{st}$;
- models trained on both semantic parsing tasks and the MT data, denoted by $Cross_{mt}$ and $mBART_{mt}$.

Furthermore, so as to explore whether the training schedules we use result in underfitting for AMR and

UCCA, we also consider $Cross_{mt}^{ft}$ and $mBART_{mt}^{ft}$, that is, $Cross_{mt}$ and $mBART_{mt}$ fine-tuned with a training schedule biased towards the semantic parsing formalism that is being considered.⁹

4.2 Datasets and Preprocessing

AMR For AMR parsing, we use AMR-2.0 (LDC2017T10) and its recently released expansion, AMR-3.0 (LDC2020T02), amounting, respectively, to 39 260 and 59 255 manually-created sentence-graph pairs.

Cross-Lingual AMR We use *Abstract Meaning Representation 2.0 - Four Translations* (Damonte and Cohen, 2020) to investigate the performance of SGL on cross-lingual AMR parsing. This corpus contains translations of the sentences in the test set of AMR-2.0 in Chinese (ZH), German (DE), Italian (IT) and Spanish (ES).

UCCA We replicate the setting of the CoNLL 2019 Shared Task (Oepen et al., 2019). We train our models using the freely available¹⁰ UCCA portion of the training data; this corpus amounts to 6 572 sentence-graph pairs, drawn from the English Web Treebank (2012T13) and English Wikipedia articles on celebrities. As no official development set was included in the data release, following Hershcovich and Arviv (2019), we reserve 500 instances and use them as the validation set. To the best of our knowledge, the full evaluation data have not been released yet and, therefore, we compare with state-of-the-art alternatives and report results only on *The Little Prince*, a released subset consisting of 100 manually-tagged sentences sampled from the homonymous novel.

⁹We report further details on schedules, models and the training procedure in Appendices A and B.

¹⁰Available on the UCCA website.

	Model	Smatch	Unlabeled	No-WSD	Concepts	Wiki	NER	Reentrancies	Negations	SRL
AMR-2.0	Ge et al. (2019)	74.3	77.3	74.8	84.2	71.3	82.4	58.3	64.0	70.4
	Zhang et al. (2019b)	77.0	80.0	78.0	86.0	86.0	79.0	61.0	77.0	71.0
	Zhou et al. (2020)	77.5	80.4	78.2	85.9	86.5	78.8	61.1	76.1	71.0
	Cai and Lam (2020)	80.2	82.8	80.0	88.1	86.3	81.1	64.6	78.9	74.2
	Xu et al. (2020a)	80.2	83.7	80.8	87.4	75.1	85.4	66.5	71.5	78.9
	SPRING _{bart}	83.8	86.1	84.4	90.2	84.3	90.6	70.8	74.4	79.6
	SPRING	84.5	86.7	84.9	89.6	87.3	83.7	72.3	79.9	79.7
	<i>Cross_{st}</i>	70.7	75.1	71.3	80.3	75.7	78.9	58.9	58.6	68.5
	<i>Cross_{mt}</i>	78.1	82.1	78.7	85.1	80.6	85.0	66.6	71.5	75.2
	<i>Cross_{mt}^{ft}</i>	79.5	83.2	80.2	86.5	80.7	85.9	68.4	71.8	77.1
	<i>mBART_{st}</i>	81.7	85.1	82.1	88.4	81.9	90.3	70.7	73.4	79.7
	<i>mBART_{mt}</i>	81.9	85.3	82.3	88.5	81.0	89.4	71.0	75.3	80.0
	<i>mBART_{mt}^{ft}</i>	82.3	85.7	82.8	88.9	82.3	89.3	71.5	73.7	80.4
	SPRING	83.0	85.4	83.5	89.8	82.7	87.2	70.4	73.0	78.9
AMR-3.0	<i>Cross_{mt}^{ft}</i>	78.1	81.9	78.7	85.3	76.6	81.3	67.6	68.5	76.2
	<i>mBART_{st}</i>	80.0	83.2	80.5	86.6	77.2	86.3	70.0	68.5	78.4
	<i>mBART_{mt}^{ft}</i>	81.2	84.4	81.6	88.4	77.7	86.5	71.1	69.7	79.7

Table 1: Smatch and fine-grained results on AMR-2.0 (top) and AMR-3.0 (bottom).

Parallel Data We use English-centric parallel corpora in four languages, namely, Chinese, German, Italian and Spanish; we employ *MultitUN* (Tiedemann, 2012) for Chinese and Spanish, *ParaCrawl* (Esplà et al., 2019) for German, and *Europarl* (Tiedemann, 2012) for Italian. We perform a mild filtering over all the available parallel sentences and then take the first $5M$ out of these.¹¹

Preprocessing We do not perform any preprocessing or tokenization, except for the graph linearizations explained in §3.1 and Chinese simplification.¹² Instead, we directly apply subword tokenization with a Unigram Model (Kudo, 2018). When working with *Cross* in a single-task setting on AMR or UCCA, we follow Ge et al. (2019) and use a vocabulary size of $20k$ subwords; instead, when working in the multilingual setting, we increase this value to $50k$ so as to better accommodate the increased amount of languages. Conversely, when using *mBART*, we always use the original vocabulary consisting of $250k$ subwords.

4.3 Evaluation

We evaluate AMR and cross-lingual AMR parsing by using the Smatch score¹³ (Cai and Knight, 2013), a metric that computes the overlap between two graphs. Furthermore, in order to have a better picture of the systems’ performances, we also re-

port the fine-grained scores as computed by the evaluation toolkit¹⁴ of Damonte et al. (2017). For UCCA parsing, we employ the official evaluation metric¹⁵ of the shared task, conceptually similar to the Smatch score.

5 Results

We now report the results SGL achieves focusing on the following translation paths: i) $EN \rightarrow AMR$ (§5.1); ii) $X \rightarrow AMR$, with X any language among Chinese, German, Italian and Spanish (§5.2); iii) $EN \rightarrow UCCA$ (§5.3).

5.1 AMR Parsing

We report the Smatch and fine-grained scores that SGL and its current state-of-the-art alternatives attain on AMR-2.0 in Table 1 (top). Among the competing systems considered, for Bevilacqua et al. (2021) we report their *BART*-powered baseline (SPRING_{bart}) and their best performing model (SPRING).

As a first result, we want to highlight the significant boost that jointly training within our proposed framework on MT data provides; *Cross_{mt}* outperforms *Cross_{st}* by more than 7 points and reaches competitive performances when compared with current state-of-the-art approaches. Furthermore, the gap of 1.4 points between *Cross_{mt}* and *Cross_{mt}^{ft}* shows that the training schedule we use for *Cross*

¹¹See Appendix C for further details.

¹²We use the *hanziconv* library (<https://github.com/berniey/hanziconv>).

¹³<https://github.com/snowblink14/smatch>

¹⁴<https://github.com/mdtux89/amr-evaluation>

¹⁵<https://github.com/cfmrp/mtool>

does indeed result in underfitting for AMR and that further training is beneficial; this fine-tuned alternative achieves 79.5 Smatch score, less than one point behind Xu et al. (2020a). Considering the similarity between the two approaches, this difference is likely caused by the increased number of tasks our model is asked to handle.

Once we replace *Cross* with *mBART*, all performances rise significantly. In particular, even *mBART_{st}*, a single-task variant with no additional data, outperforms all its alternatives except for SPRING and SPRING_{bart} (Bevilacqua et al., 2021), highlighting the potential of fully pre-trained Transformer language models for translation-based approaches. *mBART_{mt}* and *mBART_{mt}^{ft}* push performances further up, showing that the MT data are beneficial even in this pre-trained setting and that the multi-task training set, which enables a single shared model to scale across formalisms and languages, is not detrimental to English AMR parsing.

However, arguably more interesting is the comparison between the performances of *mBART* models and SPRING, which, in contrast, builds upon the English-only *BART* (Lewis et al., 2020). In particular, as SPRING_{bart} outperforms even *mBART_{mt}^{ft}*, this finding suggests that, as expected, *BART* is more suitable than *mBART* when dealing with English AMR. However, as we show in §5.2, our choice is beneficial for cross-lingual AMR parsing and results in an interesting trade-off.

Finally, we also evaluate SGL on AMR-3.0 and report the results of *Cross_{mt}^{ft}*, *mBART_{st}* and *mBART_{mt}^{ft}* when trained on this dataset (Figure 1 bottom). Overall, we witness a similar trend compared to AMR-2.0.

5.2 Cross-lingual AMR Parsing

We now show the performances of SGL on cross-lingual AMR parsing in terms of Smatch score over Chinese (ZH), German (DE), Italian (IT) and Spanish (ES). For comparison, we report the results of the systems proposed by Damonte and Cohen (2018, AMREAGER), Blloshmi et al. (2020, XL-AMR) and Sheth et al. (2021); along with their best systems, we also show the strongest MT baseline reported in Damonte and Cohen (2018, AMREAGER_{MT}) and the zero-shot configuration explored in Blloshmi et al. (2020, XL-AMR₀).

Table 2 (top) shows a very interesting trend. First of all, *Cross_{mt}^{ft}* achieves competitive performances,

	Model	DE	ES	IT	ZH
	AMREAGER	39.0	42.0	43.0	35.0
	AMREAGER _{MT}	57.0	60.0	58.0	50.0
	XL-AMR ₀	32.7	39.1	37.1	25.9
	XL-AMR	53.0	58.0	58.1	41.5
	Sheth et al. (2021)	62.7	67.9	67.4	—
HT	<i>Cross_{mt}</i>	60.8	62.9	63.2	51.8
	<i>Cross_{mt}^{ft}</i>	61.8	63.7	64.1	52.6
	<i>mBART_{st}</i>	54.8	60.4	63.6	47.8
	<i>mBART_{mt}</i>	66.3	69.0	69.8	55.4
	<i>mBART_{mt}^{ft}</i>	65.8	69.2	69.6	54.8
	<i>mBART_{mt}^{ft}</i> + AP	69.8	72.4	72.3	58.0
MT	Sheth et al. (2021)	66.9	69.6	71.0	—
	<i>mBART_{mt}^{ft}</i> + AP	73.3	73.9	73.4	64.9

Table 2: Smatch scores on cross-lingual AMR parsing for both human (top, HT) and machine (bottom, MT) translations of the test set.

falling short only when compared to the recent work of Sheth et al. (2021); in particular, it surpasses the strong AMREAGER_{MT} baseline. The most interesting aspect of this result is that *Cross_{mt}^{ft}* attains these performances without ever seeing at training time any $X \rightarrow$ AMR translation path; this is in marked contrast with all previous literature and with the systems we report in Table 2. This finding clearly highlights the effectiveness of transfer learning and, by extension, of our proposed framework in this setting.

Secondly, the performances *mBART_{st}* achieve are astounding under multiple perspectives. First, to the best of our knowledge, it is the first reported result of AMR systems achieving competitive performances on cross-lingual AMR parsing in a fully zero-shot configuration: *mBART_{st}* is fine-tuned solely on $EN \rightarrow$ AMR and then applied directly to $X \rightarrow$ AMR translation; especially when compared to XL-AMR₀, the only similar approach we are aware of, the gap is significant. Second, among the languages we consider, the case of Chinese is especially interesting as it appears to require constrained decoding in order to work properly: in particular, we restrict the model to generate only subwords whose characters belong to the English alphabet.¹⁶ If we were to perform $ZH \rightarrow$ AMR parsing with no additional decoding machinery, as for the other languages, performances would be significantly lower, with *mBART_{st}* attaining only 31.9. This performance drop is caused by

¹⁶The reported results on Chinese of all *mBART* models have been computed using this form of decoding.

the model leaving some nodes of the graph untranslated, i.e. named entities left written in Chinese (奥巴马 rather than *Obama*), which disrupts the auto-regressive nature of the decoding procedure and, besides, eventually results in a penalized Smatch score. Finally, despite the larger amount of pre-training $mBART$ has been exposed to, its bigger capacity and better Smatch score on English, $mBART_{st}$ still falls short when compared to $Cross_{mt}^{ft}$, highlighting the benefits of seeing related translation directions at training time.

$mBART_{mt}$ pushes the bar further up, with performances on German, Spanish and Italian that are now only roughly 10 points behind their English counterparts. As $mBART_{mt}$ significantly outperforms $mBART_{st}$, this result shows that, despite the massive pretraining, parallel data are still beneficial for cross-lingual AMR. Moreover, differently from English AMR, $mBART_{mt}^{ft}$ does not yield an improvement and, in fact, performances slightly drop on average.

While the scores $mBART_{mt}$ attains are already unprecedented, it is natural to wonder whether annotation projection (AP) might yield a further beneficial effect. To this end, similarly to [Blloshmi et al. \(2020\)](#), we translate the input sentences of AMR-2.0 into the four languages under consideration¹⁷ and build a training set for each language by pairing the translated sentence with the original AMR graph. We further fine-tune $mBART_{mt}^{ft}$, including also these new datasets among the training data. This model, which we denote by $mBART_{mt}^{ft} + AP$, surpasses further $mBART_{mt}$, clearly underlining the beneficial effect of this technique.

Finally, following [Sheth et al. \(2021\)](#), we also report the results of SGL when evaluated on the machine-translated test set,¹⁸ similarly to their findings, we observe that, as the mismatch between the training and test set is reduced, our parser performs better in this setting than on the human-translated one.

5.3 UCCA Parsing

We report in Table 3 the performance of SGL on UCCA parsing. We compare our approach with the original multi-task baseline ([Oepen et al., 2019](#)) and 3 transition-based parsers that participated; in

¹⁷We use the MarianMT models ([Tiedemann and Thottungal, 2020](#)) available in the *HuggingFace Transformers* library ([Wolf et al., 2020](#)).

¹⁸We use the same MT models we utilized for annotation projection.

Model	Type	Score
Oepen et al. (2019)	multi-task	41.0
Herscovich and Arviv (2019)	single-task	82.1
Herscovich and Arviv (2019)	multi-task	73.1
Che et al. (2019)	multi-task	82.6
$Cross_{st}$	single-task	55.7
$Cross_{mt}$	multi-task	72.0
$Cross_{mt}^{ft}$	multi-task	75.1
$mBART_{st}$	single-task	77.0
$mBART_{mt}$	multi-task	79.9
$mBART_{mt}^{ft}$	multi-task	76.9

Table 3: UCCA results on *The Little Prince*.

particular, we report the score of [Che et al. \(2019\)](#), the system that ranked first in both *all-framework* and UCCA parsing.

First of all, we note the result of $Cross_{st}$; while its performance is far below the score [Che et al. \(2019\)](#) achieve, it still outperforms the original proposed baseline by more than 10 points. Furthermore, to the best of our knowledge, apart from the recent works proposed in the latest shared task of [Oepen et al. \(2020\)](#), this is the first reported result of translation-based approaches on UCCA parsing.

Once plugged into our multilingual framework, UCCA benefits from transfer learning to an even greater extent than AMR parsing, likely owing to the smaller amount of training data: $Cross_{mt}$ and especially $Cross_{mt}^{ft}$ significantly reduce the gap between SGL and [Che et al. \(2019\)](#), with $Cross_{mt}^{ft}$ outperforming the multi-task transition-based approach of [Herscovich and Arviv \(2019\)](#). The usage of $mBART$ pushes up the system’s performance further, with $mBART_{st}$ achieving 77.0 and $mBART_{mt}$ 79.9; differently from AMR, $mBART_{mt}^{ft}$ suffers from overfitting and its performance is actually lower than that of $mBART_{mt}$. Even though these scores are lower than those of [Che et al. \(2019\)](#), we argue that such results are still incredibly promising as they demonstrate the effectiveness of SGL in tackling cross-framework semantic parsing. Indeed, these results show that multilingual translation-based approaches allow for a single model to jointly accommodate different formalisms, each potentially linearized according to a different linearization scheme. Furthermore, we believe there is a significant margin for improvement on both the linearization used and the model; for instance, we did not consider node ids such as $\langle root_0 \rangle$ as special tokens, but instead had the unigram tokenizer handle them as if they were normal

Model	EN	DE	AMR			UCCA
			ES	IT	ZH	EN
$Cross_{st}$	70.3	-	-	-	-	55.7
$Cross_{mt}$	78.1	60.8	62.9	63.2	51.8	72.0
$Cross_{mt}^N$	74.5	22.0	22.0	29.4	12.8	68.7

Table 4: Results on AMR, cross-lingual AMR and UCCA parsing of $Cross_{st}$, $Cross_{mt}$ and $Cross_{mt}^N$.

words.

Finally, we wish to point out that direct comparability between our system and those reported is hindered by the fact that our training setting is significantly different from theirs; in particular, we limit ourselves to two frameworks only and leverage resources (the parallel corpora from MT) whose usage was forbidden to the shared task participants.¹⁹ Nevertheless, we believe that their results are needed here to better contextualize the performances SGL obtains.

6 Analysis: is MT the one helping?

Although the performances of $Cross_{mt}$ are remarkable, $mBART_{st}$ achieves competitive results on cross-lingual parsing and fares even better on English parsing. While $mBART_{st}$ admittedly features a massive amount of pre-training, this pre-training is over monolingual corpora and, as such, the model has never seen any parallel data. We therefore wonder to what extent the parallel nature of the additional MT data we use is crucial for $Cross_{mt}$.

To answer this question, we treat our MT corpora as monolingual data by sampling, for each instance, either the source or target side and converting the translation task into a denoising one: given an instance $EN \rightarrow IT$, we sample either EN or IT with equal probability, denoting the result by Z , and convert the instance into $g(Z) \rightarrow Z$, where g is a noising function that corrupts the input text. We follow Lewis et al. (2020) and choose a noising function that masks 35% of the words by random sampling a span length from a Poisson distribution ($\lambda = 3.5$). Applying this noisification scheme to the MT data, we train a model identical to $Cross_{mt}$ and denote it by $Cross_{mt}^N$.

As shown in Table 4, in this data regime, the parallel nature is crucial both for English and, especially, for cross-lingual parsing. While $Cross_{mt}^N$ does yield a significant boost over $Cross_{st}$, when

¹⁹Allowed resources are specified at: <http://svn.nlpl.eu/mrp/2019/public/resources.txt>

compared instead to $Cross_{mt}$, it is 4 points behind on UCCA parsing and only half way on AMR parsing. The difference is even more marked in the cross-lingual setting, where $Cross_{mt}^N$ simply does not work.



7 Conclusion

In this work, we presented SGL, a novel framing of semantic parsing towards multiple formalisms as Multilingual Neural Machine Translation. That is to say, given a sentence and the desired output formalism, a many-to-many neural model has to learn to translate from the input sentence to the corresponding linearized graph. Within this framework, we show that we can address the paucity of annotated data that afflicts semantic parsing effectively by performing the learning procedure jointly on large parallel corpora coming from MT, and leveraging the power of pre-trained Transformer language models.

Using AMR and UCCA as our cases in point, we report competitive performances on their parsing, especially once pre-trained models enter the picture. Furthermore, we find that the benefit MT data provide goes beyond merely improving English-centric parsing, yielding astonishing performances on cross-lingual AMR parsing as well, and allowing SGL to outperform all existing approaches by a large margin. Most interestingly, differently from all previous literature, this result is attained without ever explicitly seeing at training time the translation paths the model is tested upon. Once we use annotation projection and include these data as well, performances rise even further, attaining unprecedented results.

As future work, thanks to the nimbleness with which we can add new *languages*, we plan to assess the scalability of this framework as more formalisms are taken into account.

Acknowledgments

The authors gratefully acknowledge the support of the ERC Consolidator Grant MOUSSE No. 726487.  

This work was partially supported by the MIUR under the grant ‘‘Dipartimenti di eccellenza 2018-2022’’ of the Department of Computer Science of Sapienza University.

References

- Omri Abend and Ari Rappoport. 2013. [Universal Conceptual 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. 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. Association for Computational Linguistics.
- Michele Bevilacqua, Rexhina Blloshmi, and Roberto Navigli. 2021. [One SPRING to rule them both: Symmetric AMR semantic parsing and generation without a complex pipeline](#). In *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence*.
- Rexhina Blloshmi, Rocco Tripodi, and Roberto Navigli. 2020. [XL-AMR: Enabling cross-lingual AMR parsing with transfer learning techniques](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2487–2500, Online. Association for Computational Linguistics.
- Claire Bonial, Lucia Donatelli, Mitchell Abrams, Stephanie M. Lukin, Stephen Tratz, Matthew Marge, Ron Artstein, David Traum, and Clare Voss. 2020. [Dialogue-AMR: Abstract Meaning Representation for dialogue](#). In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 684–695, Marseille, France. European Language Resources Association.
- Deng Cai and Wai Lam. 2020. [AMR parsing via graph-sequence iterative inference](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1290–1301, Online. Association for Computational Linguistics.
- Shu Cai and Kevin Knight. 2013. [Smatch: an evaluation metric for semantic feature structures](#). In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 748–752, Sofia, Bulgaria. Association for Computational Linguistics.
- Wanxiang Che, Longxu Dou, Yang Xu, Yuxuan Wang, Yijia Liu, and Ting Liu. 2019. [HIT-SCIR at MRP 2019: A unified pipeline for meaning representation parsing via efficient training and effective encoding](#). In *Proceedings of the Shared Task on Cross-Framework Meaning Representation Parsing at the 2019 Conference on Natural Language Learning*, pages 76–85, Hong Kong. Association for Computational Linguistics.
- Joachim Daiber, Max Jakob, Chris Hokamp, and Pablo N. Mendes. 2013. [Improving efficiency and accuracy in multilingual entity extraction](#). In *Proceedings of the 9th International Conference on Semantic Systems (I-Semantics)*, page 121–124. Association for Computing Machinery.
- Marco Damonte and Shay Cohen. 2020. [Abstract Meaning Representation 2.0 - Four Translations LDC2020T07](#). Web Download, Philadelphia: Linguistic 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. Association for Computational Linguistics.
- 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. Association for Computational Linguistics.
- Miquel Esplà, Mikel Forcada, Gema Ramírez-Sánchez, and Hieu Hoang. 2019. [ParaCrawl: Web-scale parallel corpora for the languages of the EU](#). In *Proceedings of Machine Translation Summit XVII Volume 2: Translator, Project and User Tracks*, pages 118–119, Dublin, Ireland. European Association for Machine Translation.
- DongLai Ge, Junhui Li, Muhua Zhu, and Shoushan Li. 2019. [Modeling source syntax and semantics for neural AMR parsing](#). In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 4975–4981.
- Michael Wayne Goodman. 2020. [Penman: An open-source library and tool for AMR graphs](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 312–319, Online. Association for Computational Linguistics.
- Jan Hajič, Eva Hajičová, Jarmila Panevová, Petr Sgall, Ondřej Bojar, Silvie Cinková, Eva Fučíková, Marie Mikulová, Petr Pajas, Jan Popelka, Jiří Semecký, Jana Šindlerová, Jan Štěpánek, Josef Toman, Zdeňka Urešová, and Zdeněk Žabokrtský. 2012. [Announcing Prague Czech-English Dependency Treebank 2.0](#). In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)*, pages 3153–3160, Istanbul, Turkey. European Language Resources Association (ELRA).
- Hardy Hardy and Andreas Vlachos. 2018. [Guided neural language generation for abstractive summarization using Abstract Meaning Representation](#). In *Pro-*

- ceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 768–773, Brussels, Belgium. Association for Computational Linguistics.
- 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. Association for Computational Linguistics.
- Daniel Hershcovich, Zohar Aizenbud, Leshem Choshen, Elior Sulem, Ari Rappoport, and Omri Abend. 2019. [SemEval-2019 task 1: Cross-lingual semantic parsing with UCCA](#). In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 1–10, Minneapolis, Minnesota, USA. Association for Computational Linguistics.
- Daniel Hershcovich and Ofir Arviv. 2019. [TUPA at MRP 2019: A multi-task baseline system](#). In *Proceedings of the Shared Task on Cross-Framework Meaning Representation Parsing at the 2019 Conference on Natural Language Learning*, pages 28–39, Hong Kong. Association for Computational Linguistics.
- Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2017. [Google’s multilingual neural machine translation system: Enabling zero-shot translation](#). *Transactions of the Association for Computational Linguistics*, 5:339–351.
- Diederik P. Kingma and Jimmy Ba. 2015. [Adam: A method for stochastic optimization](#). In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.
- Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander Rush. 2017. [OpenNMT: Open-source toolkit for neural machine translation](#). In *Proceedings of ACL 2017, System Demonstrations*, pages 67–72, Vancouver, Canada. Association for Computational Linguistics.
- Ioannis Konstas, Srinivasan Iyer, Mark Yatskar, Yejin Choi, and Luke Zettlemoyer. 2017. [Neural AMR: Sequence-to-sequence models for parsing and generation](#). In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 146–157, Vancouver, Canada. Association for Computational Linguistics.
- Taku Kudo. 2018. [Subword regularization: Improving neural network translation models with multiple subword candidates](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 66–75, Melbourne, Australia. Association for Computational Linguistics.
- Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2020. [ALBERT: A lite BERT for self-supervised learning of language representations](#). In *International Conference on Learning Representations*.
- Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. [BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.
- Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. 2020. [Multilingual denoising pre-training for neural machine translation](#). *Transactions of the Association for Computational Linguistics*, 8:726–742.
- Chunchuan Lyu and Ivan Titov. 2018. [AMR parsing as graph prediction with latent alignment](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 397–407, Melbourne, Australia. Association for Computational Linguistics.
- 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)*, pages 5697–5702.
- Stephan Oepen, Omri Abend, Lasha Abzianidze, Johan Bos, Jan Hajic, Daniel Hershcovich, Bin Li, Tim O’Gorman, Nianwen Xue, and Daniel Zeman. 2020. [MRP 2020: The second shared task on cross-framework and cross-lingual meaning representation parsing](#). In *Proceedings of the CoNLL 2020 Shared Task: Cross-Framework Meaning Representation Parsing*, pages 1–22, Online. Association for Computational Linguistics.
- Stephan Oepen, Omri Abend, Jan Hajic, Daniel Hershcovich, Marco Kuhlmann, Tim O’Gorman, Nianwen Xue, Jayeol Chun, Milan Straka, and Zdenka Uresova. 2019. [MRP 2019: Cross-framework meaning representation parsing](#). In *Proceedings of the Shared Task on Cross-Framework Meaning Representation Parsing at the 2019 Conference on Natural Language Learning*, pages 1–27, Hong Kong. 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. European Language Resources Association (ELRA).
- Hiroaki Ozaki, Gaku Morio, Yuta Koreeda, Terufumi Morishita, and Toshinori Miyoshi. 2020. [Hitachi at](#)

- MRP 2020: Text-to-graph-notation transducer. In *Proceedings of the CoNLL 2020 Shared Task: Cross-Framework Meaning Representation Parsing*, pages 40–52, Online. Association for Computational Linguistics.
- Janaki Sheth, Youngsuk Lee, Ramón Fernández As-tudillo, Tahira Naseem, Radu Florian, Salim Roukos, and Todd Ward. 2021. [Bootstrapping multilingual AMR with contextual word alignments](#). In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*.
- Linfeng Song, Daniel Gildea, Yue Zhang, Zhiguo Wang, and Jinsong Su. 2019. [Semantic neural machine translation using AMR](#). *Transactions of the Association for Computational Linguistics*, 7:19–31.
- Elior Sulem, Omri Abend, and Ari Rappoport. 2018. [Semantic structural evaluation for text simplification](#). 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 685–696, New Orleans, Louisiana. Association for Computational Linguistics.
- Yuqing Tang, Chau Tran, Xian Li, Peng-Jen Chen, Naman Goyal, Vishrav Chaudhary, Jiatao Gu, and Angela Fan. 2021. [Multilingual translation with extensible multilingual pretraining and finetuning](#). *ArXiv*, abs/2008.00401.
- Jörg Tiedemann. 2012. [Parallel data, tools and interfaces in OPUS](#). In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)*, pages 2214–2218, Istanbul, Turkey. European Language Resources Association (ELRA).
- Jörg Tiedemann and Santhosh Thottingal. 2020. [OPUS-MT – building open translation services for the world](#). In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, pages 479–480, Lisboa, Portugal. European Association for Machine Translation.
- Rik van Noord and Johan Bos. 2017. [Neural semantic parsing by character-based translation: Experiments with abstract meaning representations](#). *Computational Linguistics in the Netherlands Journal*, 7:93–108.
- 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*, volume 30, pages 5998–6008. Curran Associates, Inc.
- Thomas Wolf, Julien Chaumond, Lysandre Debut, Victor Sanh, Clement Delangue, Anthony Moi, Pierric Cistac, Morgan Funtowicz, Joe Davison, Sam Shleifer, Remi Louf, Patrick von Platen, Tim Rault, Yacine Jernite, Teven Le Scao, Sylvain Gugger, Julien Plu, Clara Ma, Canwei Shen, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. [Transformers: State-of-the-art natural language processing](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.
- Dongqin Xu, Junhui Li, Muhua Zhu, Min Zhang, and Guodong Zhou. 2020a. [Improving AMR parsing with sequence-to-sequence pre-training](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2501–2511, Online. Association for Computational Linguistics.
- Jin Xu, Yinuo Guo, and Junfeng Hu. 2020b. [Incorporate semantic structures into machine translation evaluation via UCCA](#). In *Proceedings of the Fifth Conference on Machine Translation*, pages 934–939, Online. Association for Computational Linguistics.
- Sheng Zhang, Xutai Ma, Kevin Duh, and Benjamin Van Durme. 2019a. [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. Association for Computational Linguistics.
- Sheng Zhang, Xutai Ma, Kevin Duh, and Benjamin Van Durme. 2019b. [Broad-coverage semantic parsing as transduction](#). 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 3786–3798, Hong Kong, China. Association for Computational Linguistics.
- Qiji Zhou, Yue Zhang, Donghong Ji, and Hao Tang. 2020. [AMR parsing with latent structural information](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4306–4319, Online. Association for Computational Linguistics.

A Model and Training Details

Cross For *Cross*, we leverage the implementation available in *OpenNMT-py*²⁰ (Klein et al., 2017) and define a Transformer model closely following Vaswani et al. (2017), except for the embedding modification we described. We use 128 as the embedding size and 512 as the Transformer hidden size when training in single-task settings; when scaling to the multilingual framework, we augment the hidden size to 768 so as to increase the model capacity. The number of layers in both the encoder and the decoder is set to 6, while the number of attention heads to 8; therefore, *Cross_{st}* contains

²⁰<https://github.com/OpenNMT/OpenNMT-py>

46M trainable parameters, while $Cross_{mt}$ 105M. We optimize the models parameters using Adam (Kingma and Ba, 2015) and the original scheduler of Vaswani et al. (2017). We train with an effective batch size of 8000 tokens and for a maximum of 300k steps on a NVIDIA GeForce RTX 2080 Ti, using semantic parsing accuracy as the early stopping criterion with 25k steps of patience; training lasted 1 day for $Cross_{st}$ and roughly 4 days for $Cross_{mt}$. We did not perform any significant tuning of decoding time parameters: we use 5 beams and, following Ge et al. (2019), we set the $GNMT$ length penalty parameter $alpha$ to 1.0.

MBart We use the *mbart-large-cc25* model provided by the Huggingface’s *transformers* library (Wolf et al., 2020) and follow the specifications given in the original paper for the training configuration: we use 2500 warm-up steps, $3e - 5$ maximum learning rate, Adam as the optimizer and 0.3 dropout. We used 8000 effective token batch size and carried out training on a cloud-provided NVIDIA GeForce 3090, using semantic parsing accuracy as the early stopping criterion with 25k steps of patience; training lasted slightly less than 2 days for $mBART_{st}$ and around 4 days for $mBART_{mt}$. As we fine-tune the original model, the number of parameters is unchanged (but for the two additional vectors for UCCA and AMR in the embedding matrix), with both $mBART_{st}$ and $mBART_{mt}$ featuring 610M trainable parameters.

B Training Task Scheduler

When training $Cross_{mt}$ and $mBART_{mt}$, we oversample semantic parsing instances by sampling from the concatenation of AMR and UCCA with probability $p_1 = 0.15$ and $p_1 = 0.4$ respectively; the remaining probability p_2 is uniformly divided among the MT corpora. When considering $Cross_{mt}^{ft}$ and $mBART_{mt}^{ft}$, we bias the schedule towards the semantic formalism under consideration by bringing its probability to $p_3 = 0.8$; we assign the remaining $p_4 = 0.2$ uniformly to the MT corpora.

C Filtering over Parallel Corpora

We perform a mild filtering over parallel sentences, enforcing reasonable minimum, maximum and relative lengths; in particular, we discard sentences:

- that are shorter than 25 characters;
- that are longer than 1 000 characters;

- where the relative character ratio between source and target is bigger than 3.0.

We then take the first 5M out of these. This process results in 5M parallel sentences for all four translation paths except EN \rightarrow IT, where, owing to the smaller size of *Europarl*, it results in only 1.6M sentence pairs.