Accepted refereed manuscript of: Bölücü N, Can B & Artuner H (2023) A Siamese Neural Network for Learning Semantically-Informed Sentence Embeddings. *Expert Systems with Applications*, 214, Art. No.: 119103.
https://doi.org/10.1016/j.eswa.2022.119103

A Siamese Neural Network for Learning Semantically-Informed Sentence Embeddings

Necva Bölücü[a,c], Burcu Can[b], Harun Artuner[a]

[a]Department of Computer Engineering, Hacettepe University, Ankara, Turkey.

[c]Graduate School of Science and Engineering, Hacettepe University, Ankara, Turkey.

Email: necva@cs.hacettepe.edu.tr

[b] Research Institute of Information and Language Processing, RGCL, University of Wolverhampton, UK. Email: b.can@wlv.ac.uk

[a]Department of Computer Engineering, Hacettepe University, Ankara, Turkey.

Email: artuner@hacettepe.edu.tr

Corresponding Author:

Email address: necva@cs.hacettepe.edu.tr (Necva Bölücü)

A Siamese Neural Network for Learning Semantically-Informed Sentence Embeddings

Necva Bölücü[a,c,*], Burcu Can[b], Harun Artuner[a]

[a] *Department of Computer Engineering, Hacettepe University, Ankara, Turkey*
[b] *Research Institute of Information and Language Processing, University of Wolverhampton*
[c] *Graduate School of Science and Engineering, Hacettepe University, Ankara, Turkey*

Abstract

Semantic representation is a way of expressing the meaning of a text that can be processed by a machine to serve a particular natural language processing (NLP) task that usually requires meaning comprehension such as text summarisation, question answering or machine translation. In this paper, we present a semantic parsing model based on neural networks to obtain semantic representation of a given sentence. We utilise semantic representation of each sentence to generate semantically informed sentence embeddings for extrinsic evaluation of the proposed semantic parser, in particular for the semantic textual similarity task. Our neural parser utilises self-attention mechanism to learn semantic relations between words in a sentence to generate semantic representation of a sentence in UCCA (Universal Conceptual Cognitive Annotation) semantic annotation framework (Abend & Rappoport, 2013), which is a cross-linguistically applicable graph-based semantic representation. The UCCA representations are conveyed into a Siamese Neural Network built on top of two Recursive Neural Networks (Siamese-RvNN) to derive semantically informed sentence embeddings which are evaluated on semantic textual similarity task. We conduct both single-lingual and cross-lingual experiments with zero-shot and few-shot learning, which have shown superior performance even in low-resource scenario. The

*Corresponding author.
Email addresses: nbolucu@cs.hacettepe.edu.tr (Necva Bölücü), b.can@wlv.ac.uk (Burcu Can), artuner@hacettepe.edu.tr (Harun Artuner)

experimental results show that the proposed self-attentive neural parser outperforms the other parsers in the literature on English and German, and shows significant improvement in the cross-lingual setting for French which has comparatively low sources. Moreover, the results obtained from other downstream tasks such as sentiment analysis confirm that semantically informed sentence embeddings provide higher-quality embeddings compared to other pre-trained models such as SBERT (Reimers et al., 2019) or SimCSE (Gao et al., 2021), which do not utilise such structured information.
Keywords: Semantic parsing, UCCA, self-attention, semantic textual similarity, Siamese Network, Recursive Neural Network

1. Introduction

Semantics is concerned with meaning defined by relations between words in a sentence. Revealing semantic relations between words or a group of words in a sentence helps to better understand natural languages and this will eventually aid in developing linguistically motivated semantic models in natural language processing (NLP) applications. While semantics is the study of the relations between the building blocks of a sentence (i.e. words or a group of words) and their implicit meaning, semantic representation reflects the meaning of the text in a rather structured form (e.g. graph-based or tree-based representation) (Abend & Rappoport, 2017). Semantic parsing is the task of mapping a text given in a natural language to its formal representation which provides an abstraction of its meaning that can be easily processed by a machine to serve a particular NLP task. Semantic parsing has recently been dominated by tree-structured representations[1].

Graphs have been receiving increasing attention in NLP in recent years due

[1] "Embedding" and "representation" are used interchangeably in the literature; "representation" is also used for semantic parsing. In this paper, "representation" is used to refer to semantic parse tree, whereas "embedding" is used to refer to the distributed low-dimensional vectors to distinguish the two terms.

to their ability to express and generate adequate target structures, especially for sentence-level syntactic analysis and semantic representation of a text. The increasing popularity of graph-based semantic representations has led to the proposal of various semantic representation frameworks such as Abstract Meaning Representation (AMR) (Banarescu et al., 2013), Universal Conceptual Cognitive Annotation (UCCA) (Abend & Rappoport, 2013), bilexical Semantic Dependencies (SDP) (Oepen et al., 2016), Universal Decompositional Semantics (UDS) (White et al., 2016), and Parallel Meaning Bank (PMB) (Abzianidze et al., 2017). These graph-based representations have proven to be beneficial in Natural Language Understanding (NLU) tasks and have already demonstrated their applicability in a variety of NLP tasks such as summarisation (Dohare & Karnick, 2017; Liu et al., 2018), paraphrase detection (Issa et al., 2018; Blloshmi et al., 2020), machine translation (Song et al., 2019; Sulem et al., 2020), question answering (Kapanipathi et al., 2020; Xu et al., 2021), and text simplification (Sulem et al., 2018).

In the last decade, representation of words in a low-dimensional space has provided a profound way of expressing meaning in a compact vector-based form, that are well-known as word embeddings in the literature. Many neural approaches have been introduced to that end (Devlin et al., 2019; Liu et al., 2019; Peters et al., 2018), all competing with each other for better embeddings (usually evaluated on an independent task). Advances in neural word representation techniques have paved the way for representative embeddings for also larger units such as phrases and sentences (Kiros et al., 2015; Hill et al., 2016; Conneau et al., 2017; Logeswaran & Lee, 2018; Cer et al., 2018; Reimers et al., 2019). Sentence embeddings, one of the concerns of this research, are utilised in many NLP applications such as question answering (Yao et al., 2013; Severyn & Moschitti, 2013), short answer grading (Ramachandran et al., 2015), text summarisation (Wang et al., 2016; Nova, 2019), evaluation of machine translation models (Chan & Ng, 2008; Liu et al., 2011; Wang et al., 2017). The aforementioned NLP applications generally benefit from sentence embeddings to assess the semantic similarity between two phrases or sentences. This is also known as

Semantic Textual Similarity (STS) that is the evaluation of pairs of sentences or phrases according to their degree of semantic similarity.

Our main research question is whether semantic representation can be used to learn better sentence embeddings that lead to more accurate sentence similarity evaluation. We introduce a neural semantic parser model that generates a semantic representation of a sentence, and employ the obtained semantic representations to evaluate the semantic similarity between two sentences in semantic textual similarity task.

Graph-based UCCA (Universal Conceptual Cognitive Annotation) (Abend & Rappoport, 2013) semantic representation is one of the graph-based semantic representations that has recently gained attention. It is a cross-linguistically applicable semantic annotation scheme (Abend & Rappoport, 2013) that can be learned as a universal representation across languages. The UCCA representation is built upon a multi-layer structure where each layer specifies the semantic relations it encodes between the building blocks of a sentence. Since the UCCA representation framework is introduced as a paragraph-level annotation, it enables extending the sentence-level annotation to paragraph-level annotation. The UCCA semantic annotation is illustrated by directed acyclic graphs (DAGs) whose leaves, called terminals, correspond to word tokens and multi-tokens (not necessarily corresponding to syntactically complete phrases) in a sentence. In this work, we use UCCA-based semantic representation to generate semantically informed sentence embeddings, which has not been studied before in the literature.

In order to learn UCCA-based semantic representations, we introduce a neural semantic parser that approaches the semantic parsing task as a constituency parsing problem[2]. Our model is inspired by the non-projective dependency parser of Nilsson & Nivre (2005), which has been used in semantic parsing (Jiang et al., 2019; Zhang et al., 2019) but not for UCCA-based an-

[2]Constituency parsing is the process of extracting grammatical categories which are usually a group of words that belong to a phrase such as a noun phrase or a verb phrase in a sentence.

notation before. To learn sentence embeddings, we adopt a Siamese Recursive Neural Network (Siamese-RvNN), which is a combination of a Siamese Network (Chopra et al., 2005) and a Recursive Neural Network (Socher et al., 2010). Siamese networks (Chopra et al., 2005) are dual-branch networks with bound weights having built on the same network, which is copied and merged with an energy function. In the proposed Siamese structure, the same set of weights is used recursively for each UCCA-based representation in the form of a directed acyclic graph. To the best of our knowledge, this is the first attempt to generate sentence embeddings based on UCCA-based semantic representations using Siamese neural networks.

We evaluated both our proposed semantic parser model for learning UCCA-based semantic graphs and Siamese-RvNN for learning sentence embeddings. Our semantic parser model outperformed the other participants in SemEval 2019 (Hershcovich et al., 2019) in both English and German, for both labeled and unlabeled annotation. We also obtained the best results on semantic textual similarity using our proposed Siamese-RvNN model. The results show that using semantically informed sentence embeddings is superior to even recent sentence embedding approaches such as SBERT (Reimers et al., 2019) and SimCSE (Gao et al., 2021).

In summary, the contribution of this study is 5-fold:

- We propose a neural semantic parser model that learns UCCA-based semantic representations of sentences.
- We introduce a neural network architecture for evaluating semantic textual similarity between two sentences.
- We conducted single-lingual and cross-lingual experiments for the semantic parsing task. For the cross-lingual experiments, we performed both few-shot and zero-shot learning due to the insufficient size of the available training data in French.
- We obtained state-of-the-art semantic parsing results in English and Ger-

man in single-lingual setting. The results show that the cross-lingual model performs better in low-resource languages.

- Our proposed Siamese-RvNN model outperforms other approaches in the semantic textual similarity task. We also obtained competitive results in transfer learning tasks by using semantically informed sentence embeddings in downstream NLP tasks such as sentiment analysis.

The paper is organised as follows: Section 2 briefly describes the UCCA semantic representation and annotation, Section 3 discusses related work on both UCCA parsing and semantic textual similarity, Section 4 explains the proposed semantic parser used to learn UCCA semantic-graphs and the Siamese networks used to learn sentence embeddings based on their UCCA semantic-graphs, Section 5 presents our experimental results along with the datasets, evaluation measures, and experimental setup for both tasks, and finally Section 6 concludes the paper with the potential future goals.

2. UCCA Semantic Representation

UCCA is both cognitively and linguistically inspired semantic representation framework Abend & Rappoport (2013). The UCCA representation of a sentence includes some relations and arguments, which makes it deviate from syntactic analysis; e.g. dependency parsing, where syntactic roles matter in annotation but not the actual semantic relations between the arguments. The UCCA semantic representation of a sentence is basically a DAG, where a node can be either a terminal or a non-terminal compromising several tokens that are jointly viewed as a single entity according to some semantic or cognitive consideration. However, those joint units may not directly correspond to syntactically complete phrases as in syntactic parsing but rather are related to each other based on their semantic roles.

The edges of the graphs refer to the role of the child in the relation (i.e. semantic categories) such as scene elements (Process, State, Participant, Ad-

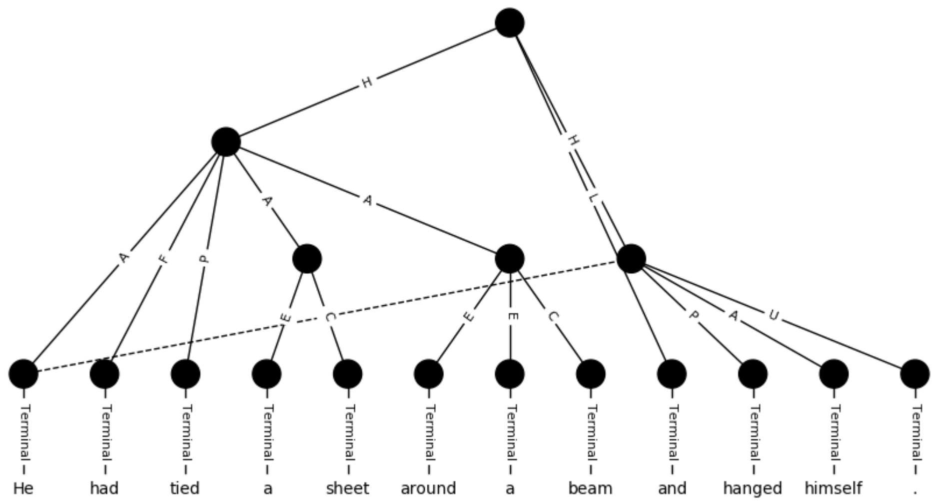


Figure 1: An Example of UCCA annotation for the sentence “He had tied a sheet around a beam and hanged himself.”

verbial), elements of the non-scene units (Center, Elaborator, Connector, Relator), and inter-scene relations (Parallel Scene, Linker, Ground), and other roles (Function).

An example UCCA representation is illustrated in Figure 1. In the example, there are two Scenes having a relation called **Process** that correspond to two actions: “tied” and “hanged”. “and” is a **Linker** between the two Scenes. **Participant** of the Process is the terminal “He”, who is affected by the Processes. “He” has got two parents connected by a primary and a remote edge (dashed) for both Scenes. It is an example of a discontinuous unit (due to “He”), which is a Participant of the two Scenes.

3. Related Work

The related work on both UCCA-based semantic parsing and semantic textual similarity is given separately in the following two subsections.

3.1. UCCA-based Semantic Parsing

TUPA parser (Hershcovich et al., 2017) is the first parser proposed for generating UCCA representations. It is a neural transition-based parser model that includes additional transition actions and features to handle discontinuous and remote nodes in UCCA graphs. Hershcovich et al. (2018) extend the TUPA parser with multi-task learning by utilising other semantic graph representations, namely AMR (Banarescu et al., 2013), UD (Nivre et al., 2020, 2016), and SDP (Oepen et al., 2016).

The UCCA framework has been the main theme in some share tasks; e.g. "Cross-lingual Semantic Parsing with UCCA" at SemEval 2019 (Hershcovich et al., 2019) and "Meaning Representation Parsing (MRP)" cross-framework shared task (Oepen et al., 2020, 2019) in 2019 and 2020.

Current UCCA-based semantic parsers can be categorised based on their approaches as follows: (i) transition-based (Hershcovich et al., 2017; Pütz & Glocker, 2019; Lyu et al., 2019), (ii) graph-based (Cao et al., 2019; Droganova et al., 2019; Koreeda et al., 2019; Li et al., 2019; Na et al., 2019; Wang et al., 2019a; Zhang et al., 2019) , (iii) composition-based (Che et al.; Donatelli et al., 2019; Oepen & Flickinger, 2019), and (iv) encoder-decoder based (Dou et al., 2020; Na & Min, 2020; Cai & Lam, 2020).

Transition-based approaches (Bai & Zhao, 2019; Che et al.; Lai et al., 2019; Straka & Straková, 2019) define a sequence of actions that eventually build semantic graphs. Some of the transition-based approaches extend the models by adding extra actions (Arviv et al., 2020; Lai et al., 2019), adding new features (Pütz & Glocker, 2019; Bai & Zhao, 2019), or layers (Lyu et al., 2019; Che et al.).

Graph-based approaches (Cao et al., 2019; Droganova et al., 2019; Koreeda et al., 2019; Li et al., 2019; Na et al., 2019; Wang et al., 2019a; Zhang et al., 2019) generally tackle the task as a search problem to find the graph with the highest score among all possible graphs for a given input. Some of the approaches use the existing neural parser architectures introduced especially for dependency parsing (NeurboParser (Peng et al., 2017), JAMR (Flanigan et al., 2014), and UDPipe (Straka & Straková, 2017)), whereas others introduce other neural architectures for UCCA-based semantic representation using a graph-based approach (Li et al., 2019; Zhang et al., 2019; Jiang et al., 2019; Koreeda et al., 2019; Na et al., 2019).

Composition-based approaches follow the compositionality principle and performs semantic parsing as a result of a derivation process in which both lexical and syntactic-semantic rules are incorporated to develop a semantic graph parser (Che et al.; Donatelli et al., 2019; Oepen & Flickinger, 2019).

Finally, encoder-decoder approaches use an encoder-decoder architecture to convert an input sentence into a semantic graph (Dou et al., 2020; Na & Min, 2020; Cai & Lam, 2020) as performed in machine translation.

3.2. Semantic Textual Similarity

The identification of STS in short texts was proposed in 2006 (Li et al., 2006; Mihalcea et al., 2006), where the goal was to identify whether two text segments are paraphrases of each other or not. Between 2012 and 2017, the semantic similarity task has been one of the main tasks in SemEval (Agirre et al., 2012, 2013, 2014, 2015, 2016) and the proposed models based on neural networks (Šarić et al., 2012; Afzal et al., 2016; He et al., 2015; He & Lin, 2016a; Shao, 2017) not only were able to identify a similarity between two texts, but also were able to generate a similarity score (usually between 0 and 5).

Measuring semantic similarity between texts has been performed using several methods in the literature (Majumder et al., 2016): (i) topological method, which utilises external semantic resources such as WordNet in order to assess the similarity between two texts using topological distance on such semantic networks (Ramage et al., 2009; Jiang & Conrath, 1997; Sussna, 1993; Sánchez et al., 2012; Gutiérrez et al., 2016), (ii) statistical similarity that exploits mainly statistical vector-based models along with dimension reduction techniques to assess the similarity between two texts (Gabrilovich et al., 2007; Ando, 2000; Jiang & Conrath, 1997) , (iii) semantic-based method that combines a set of similarity measures such as soft cardinality (Jimenez et al., 2012), word n-gram overlap to predict the similarity between texts (Afzal et al., 2016; Sultan et al., 2016), symbolic regression (Martinez-Gil & Chaves-Gonzalez, 2020), (iv) machine learning method, which builds a mathematical model based on lexical, syntactic and semantic features to compute the similarity between given texts (Shao, 2017; Reimers et al., 2019; Gao et al., 2021; Wu et al., 2021; Zhang & Lan, 2021).

Semantic similarity methods have recently made the most out of recent developments in neural networks, especially the recent neural word embedding approaches. The most commonly used neural network architectures for semantic similarity are Convolutional Neural Networks (CNN) (Kim, 2014; Shao, 2017), Long short Term Memory Networks (LSTM) (Tien et al., 2019), Bidirectional Long Short Term Memory (BiLSTM) (He & Lin, 2016b), Recursive Tree LSTMs (Tai et al., 2015) and Decomposable Attention Model (DAM) using n-grams (Lopez-Gazpio et al., 2019).

Since pre-trained language models obtained from BERT (Devlin et al., 2019) and RoBERTa (Liu et al., 2019) achieved state-of-the-art results on sentence-pair regression-classification tasks such as question answering (Qu et al., 2019; Chaybouti et al., 2021), natural language inference (Bowman et al., 2015), they are also applied to the semantic textual similarity task (Reimers et al., 2019; Yin et al., 2020; Li et al., 2020; Cheng, 2021; Xia et al., 2021).

Semantic similarity has also been used in other tasks such as recommendation systems (Bougiatiotis & Giannakopoulos, 2018), code clone detection (Sheneamer, 2021).

4. Self-Attentive UCCA Semantic Parser and Learning Sentence Embeddings using UCCA Representation

In this section, we describe the proposed self-attentive semantic parser model that generates UCCA-based semantic representation of a given sentence and the Siamese Recursive Neural Network (Siamese-RvNN) model that is proposed to assess the semantic similarity between sentences for which UCCA-based semantic graphs are obtained from the proposed neural parser.

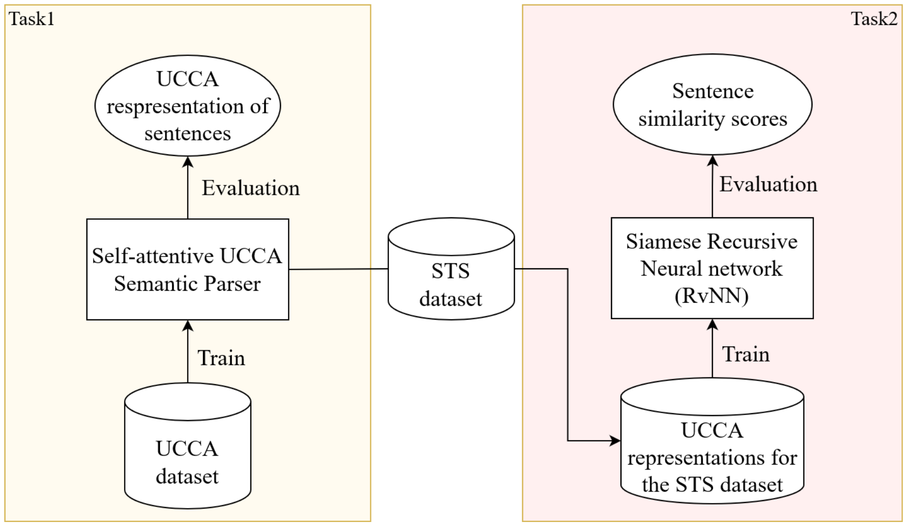


Figure 2: The proposed system's overview

The overview of the proposed model flow is shown in Figure 2. First, we train the self-attentive semantic parser model on the UCCA dataset, which is illustrated as *Task1*. Then, we obtain UCCA representations of the sentences in the STS dataset using the self-attentive semantic parser. Finally, we use the UCCA semantic representations to train universal sentence embeddings to predict the similarity between sentence pairs for the STS task, which is illustrated in *Task2*. The details of each model are given in the following subsections.

4.1. Self-Attentive UCCA Semantic Parser

We adopt the constituency parsing model based on self-attention mechanism proposed by Kitaev & Klein (2018) to learn the UCCA semantic representations of a given text. The parser is built on an encoder-decoder architecture, where the encoder is based on self-attention mechanism and the decoder is based on CYK (Cocke-Younger-Kasami) algorithm (Chappelier & Rajman, 1998). The overall view of the encoder-decoder architecture is given in Figure 3. The parser follows a chart-based constituency parsing approach where the constituency tree T of an input sentence $s = \{w_1, \cdots, w_n\}$ with words w_i is defined as a set of labeled

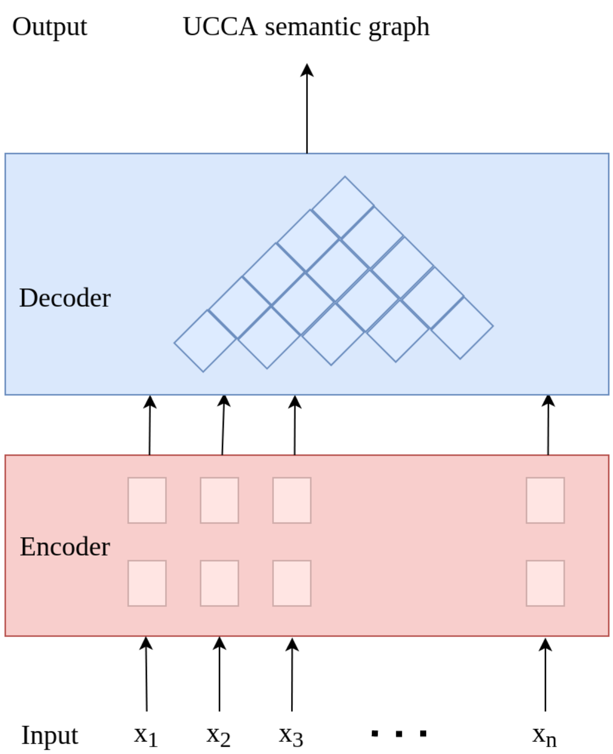


Figure 3: The architecture overview of the self-attentive semantic parser model

spans:

$$T = \{(i_t, j_t, l_t) : t = 1, \cdots, |T|\} \tag{1}$$

where i_t and j_t refer to the beginning and ending positions of the t^{th} span respectively with the label set $l_t \in L$.

We assign a score $s(T)$ to each tree, which is decomposed as follows:

$$s(T) = \sum_{(i,j,l) \in T} s(i, j, l) \tag{2}$$

Here, $s(i, j, l)$ denotes per-span scores predicted by the model.

Each word w_t is mapped into a dense vector x_t which is a concatenation of the word embedding e_{w_t}, PoS tag embedding e_{p_t}, dependency label embedding e_{d_t}, entity type embedding e_{e_t}, and entity iob (inside-outside-beginning) category

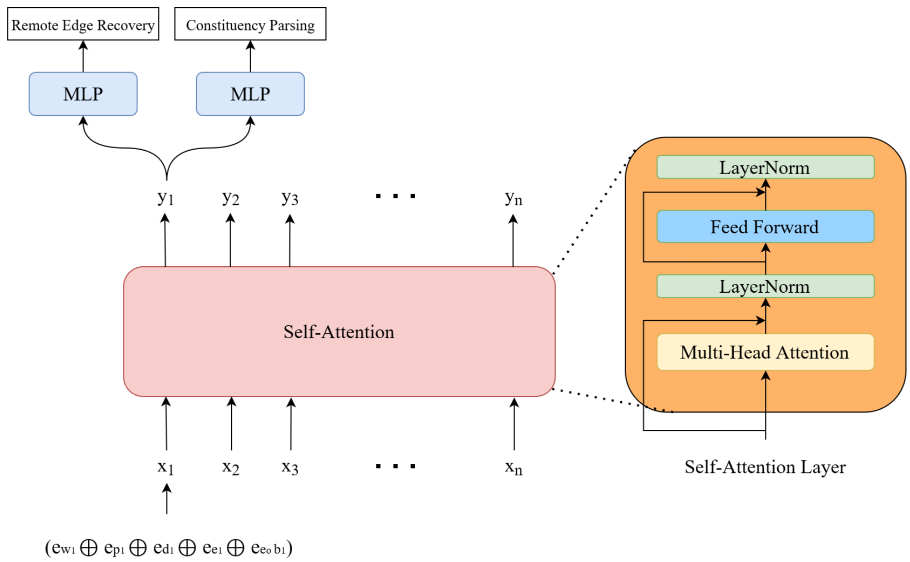


Figure 4: An overview of the self-attention encoder

embedding $e_{e_o b_t}$:

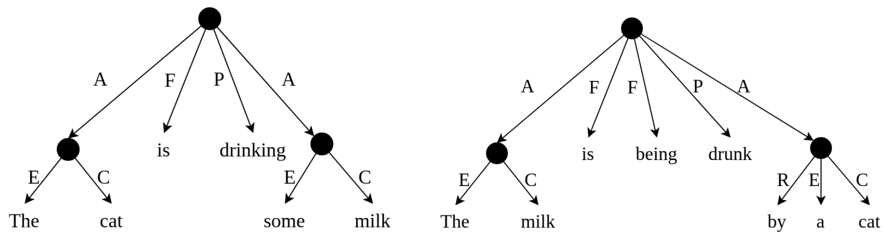
$$x_t = e_{w_t} \oplus e_{p_t} \oplus e_{d_t} \oplus e_{e_t} \oplus e_{e_o b_t} \tag{3}$$

The overview of the encoder along with the remote edge recovery is given in Figure 4. The encoder consists of multiple self-attention layers[3] and only one of them is depicted in the figure for simplicity reasons. The encoder learns a context vector y_t for each position t for a word vector x_t.

An MLP classifier with two fully-connected layers with ReLU activation function assigns labeling scores $s(i, j, l)$ to each span using the encoder output. We integrate remote edge recovery that also shares the same encoder to recover remote edges in trees (Jiang et al., 2019) as shown in Figure 4. Therefore, the model incorporates two independent MLPs to predict remote edges and candidate parent nodes that use the same encoder.

The parsing loss is the sum of the cross-entropy losses introduced by both remote edges L_{remote} and non-terminal node pairs $L_{non-terminal}$ as indicated

[3] In our model, the encoder involves 8 self-attention layers.



(a) The cat is drinking some milk

(b) The milk is being drunk by a cat

Figure 5: UCCA-based semantic representations of a sentence pair (A:Participant, P:Process, F:Function, E:Elaborator, C:Center, R:Relator).

below:

$$L = L_{remote} + L_{non-terminal} \tag{4}$$

As for inference, CYK (Cocke-Younger-Kasami) algorithm (Chappelier & Rajman, 1998) is used to generate a globally optimized tree $\hat{T}$ for each sentence that acts as a decoder in the model (see Figure 3). Therefore the tree with the maximum score is identified by the CYK algorithm as follows:

$$\hat{T} = \arg\max_{T} s(T) \tag{5}$$

The output of the semantic parser model is a UCCA (Abend & Rappoport, 2013) representation of the sentence in the form of a directed graph. An example of a sentence pair annotated by UCCA semantic graph using the self-attentive neural semantic parser is given in Figure 5.

4.2. Siamese Recursive Neural Network (Siamese-RvNN)

Recursive Neural Network (RvNN) (Socher et al., 2010) is a type of neural network in which the same set of weights is applied recursively over a structured input. Each recursive network processes the nodes in topological order in the given structure (in the form of a graph or a tree) and recursively applies transformations to generate further representations from the previously computed representations of children.

We build the RvNN model with graphs constructed by the self-attentive UCCA semantic parser (see Figure 5) with a list of words represented as d-dimensional vectors in a pre-trained word embedding matrix $L \in \mathbb{R}^{\mathrm{d} \times |\mathrm{V}|}$ where $|V|$ is the size of the vocabulary.

As illustrated in Figure 6, we obtain the representation of "The cat" by the composition of "The" and "cat", "some milk" by the composition of "some" and "milk" and the representation of "The cat is drinking some milk" is obtained by the vectors of "The cat", "is", "drinking" and "some milk". The compositional sentence embedding is eventually generated based on the UCCA semantic representation of the sentence, which also gives semantically-informed sentence embeddings.

The composition is applied using a fully connected layer (i.e. one-layer MLP) for each node in the semantic graph. The mean of the input vectors is fed into the MLP since RvNN is adopted for non-binary trees that are the UCCA representation of sentences. For example, in Figure 6, x_1 (The) and x_2 (cat) are combined by the following nonlinear composition with weights W and the parent vector y_1 is computed and y_1, x_3 (is), x_4 (drinking) and y_2 are used to compute y_3 as follows:

$$y_1 = f(Wg(x_1, x_2) + b) \tag{6}$$

$$y_3 = f(Wg(y_1, x_3, x_4, y_2) + b) \tag{7}$$

where f is a nonlinear activation function ReLU, g is the representation extractor function that is mean in the model, W is the weight matrix (with a dimensionality of $d \times d$, where d is the embedding dimension of pre-trained word embedding) and b corresponds to bias vector in that layer. A single MLP is used for the model, therefore the weights are the same for all sentences in the dataset.

Here we combine Siamese Networks with RvNNs. Siamese networks (Chopra et al., 2005) are dual-branch networks with bound weights. In other words, they are built on the same network copied and merged with an energy function.

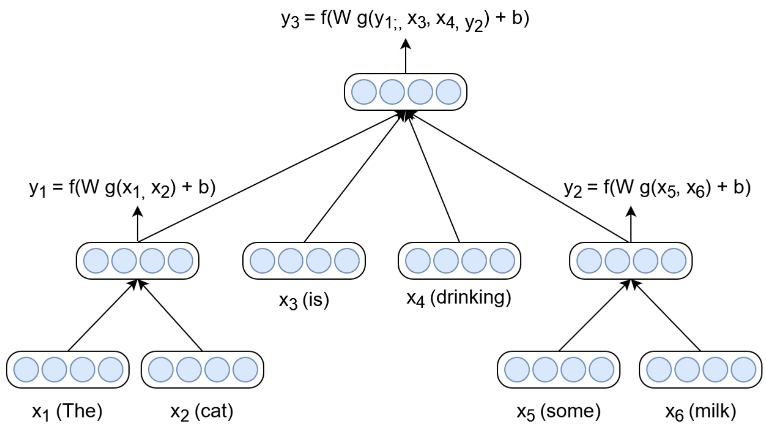


Figure 6: The composition process for "The cat is drinking some milk." using a Recursive Neural Network

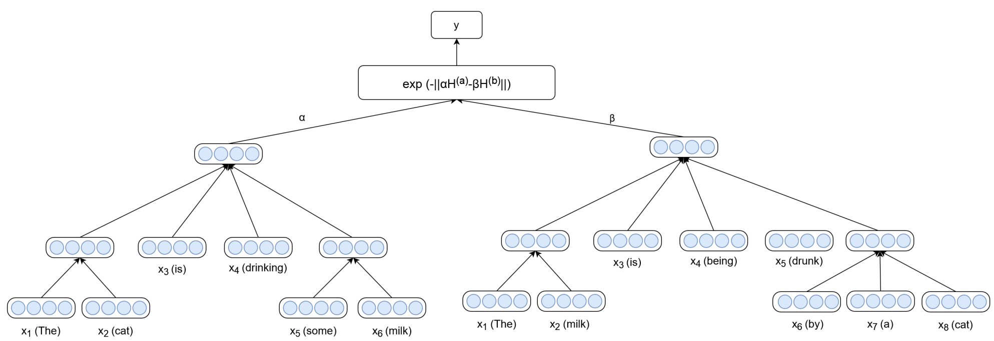


Figure 7: Overview of the Siamese-RvNN model architecture

The Siamese architecture is given in Figure 7. There are two networks $RvNN_a$ and $RvNN_b$ that simultaneously process one of the sentences in a given sentence pair. An example pair of sentences is given in Figure 5. The training set consists of triplets (x_1, x_2, y), where x_1 and x_2 are sentences in a pair in the training set, and y is the similarity score that is between $[0, 5]$ and defines the semantic similarity between the two sentences. The goal is to minimise the distance between semantically similar sentences and maximise the distance between dissimilar sentences in the embedding space for each pair, which is followed during training.

We use the Manhattan distance (Craw, 2017) which performs comparatively better than other distance metrics in Recurrent Neural Networks (Yih et al., 2011; Mueller & Thyagarajan, 2016; Pontes et al., 2018) to measure the similarity between sentences in a pair as follows:

$$g = exp(-\alpha H^{(a)} - \beta H^{(b)}) \in [0, 1] \tag{8}$$

Here, g is computed by the model where $H^{(a)}$ is the output of network $RvNN_a$ and $H^{(b)}$ is the output of network $RvNN_b$. α ve β are two parameters that are used to apply weighting on the output of the two $RvNN$ models: $H^{(a)}$ and $H^{(b)}$. We rescale the output to ensure that the similarity is in the range of $[0, 5]$.

5. Evaluation and Results

In this section, we provide the details of our experimental setting and the results of our evaluation for both semantic parsing and semantic textual similarity.

5.1. Experimental Setting and Results for Semantic Parsing

Datasets We used the SemEval 2019 shared task dataset (Hershcovich et al., 2019) in this study because it is especially built for cross-lingual parsing. The details of the datasets are given in Table 1. The dataset includes English, German, and French languages from *Wikipedia* and *Twenty Thousand Leagues Under the Sea*. We performed two single-lingual experiments similar to (Hershcovich et al., 2019) for English: 1. In-domain setting using English-Wiki corpus for both training and testing purposes (a separate validation set under the same dataset is used for testing) 2. Out-of-domain setting using English-Wiki corpus for training and the English-20K validation set for testing purposes. We only performed in-domain experiments for German and French, since only one dataset is available for both languages.

Since there is not enough training data available for French in the SemEval 2019 dataset, we conducted cross-lingual experiments by merging the training datasets of three languages to train the model in a cross-lingual setting. In this way, we expect the model to utilise all the training data as if it comes from a single language. This helps to share the languages' parameters using the similarities in different languages while learning the language differences concurrently.

	English-Wiki	English-20K	German-20K	French-20K
Train	4,113	0	5,211	15
Validation	514	0	651	238
Test	515	492	652	239

Table 1: Number of sentences in each dataset in Semeval 2019 Dataset (Hershcovich et al., 2019)

Evaluation Metrics We followed the official evaluation metrics (Hershcovich et al., 2019) used in SemEval 2019. The evaluation method measures a matching score between each output graph $G_o = (V_o, E_o, l_o)$ predicted by the model and its corresponding gold graph $G_g = (V_g, E_g, l_g)$ over the same sequence of nodes. Labeled precision and recall metrics are calculated by dividing the number of matching edges in G_o and G_g with their corresponding labels to $|E_o|$ and $|E_g|$ respectively.

F_1 is the harmonic mean of precision and recall:

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision \times Recall} \tag{9}$$

Unlabeled precision, recall, and F_1 are computed analogously, but without requiring a label matching for the edges. In all experiments, we evaluate both primary and remote edges separately.

Hyperparameters and Implementation Details The semantic parser model and the RvNN-Siamese architecture for the semantic similarity task are implemented using PyTorch and both are publicly available at `https://github.com/necvabolucu/UCCA-transformer`. For the encoder, we used a self-attention layer with the same parameter values as in Vaswani et al. (2017).

The word embedding dimensionality is 100 with an embedding dimensionality of 50 for PoS tags, 50 for dependency tags, 25 for entity types, and 25 for the entity iob types. We used Adam optimizer (Kingma & Ba, 2014) and early stopping during training because of the variations in the size of the training sets.

All syntactic embeddings (i.e. word, PoS tags, dependency tags, entity types, and entity iob types) are randomly initialised in the single-lingual experiments. In addition to the syntactic embeddings, we used pre-trained fasttext (Bojanowski et al., 2017a) character n-gram based word embeddings. Additionally, we used BERT embeddings as contextualised embeddings to incorporate contextual information. For the cross-lingual models, we conducted experiments with and without contextual embeddings in addition to the syntactic embeddings.

Results The results of both single-lingual and cross-lingual experiments on SemEval 2019 (Hershcovich et al., 2019) datasets for English, French and German are given in Table 2.

For the single-lingual setting, the use of fasttext embeddings (Bojanowski et al., 2017b) along with BERT contextualised embeddings (Devlin et al., 2019) in addition to syntactic embeddings outperforms the other settings in all languages. For the cross-lingual setting, the results have slightly decreased for all languages except French. However, the results for French have improved significantly. The cross-lingual setting helps predict remote edges in French, while it is not sufficient to predict remote edges in single-lingual setting due to the insufficient amount of training data for French. We did not perform experiments with pre-trained fasttext embeddings as they are trained independently for different languages and are not available as multilingual embeddings. Using BERT improves F1 scores by about 4% for English, 5% for German, and 4% for French in the cross-lingual setting.

Comparative results of our model with other participants of Semeval 2019 (Her-

	Single-Lingual Exp.			Cross-Lingual Exp.		
	Prim.	Rem.	Avg	Prim.	Rem.	Avg
			English-Wiki			
syntactic emb.	74.5	2.1	73.04	74.8	44.7	74.19
syntactic emb. ⊕ fasttext	77.9	53.0	77.4	-	-	-
syntactic emb. ⊕ bert	78.3	52.8	77.79	79.6	48.5	78.97
syntactic emb. ⊕ fasttext ⊕ bert	80.2	55.4	**79.7**	-	-	-
			English-20K			
syntactic emb.	71.0	7.9	68.87	72.7	23.6	71.04
syntactic emb. ⊕ fasttext	73.8	25.0	72.15	-	-	-
syntactic emb. ⊕ bert	75.45	28.6	73.87	75.9	29.4	74.33
syntactic emb. ⊕ fasttext ⊕ bert	76.2	29.3	**74.62**	-	-	-
			German-20K			
syntactic emb.	77.3	31.5	76.09	80.4	49.3	79.58
syntactic emb. ⊕ fasttext	83.6	60.2	82.98	-	-	-
syntactic emb. ⊕ bert	85.1	63.7	84.54	86.2	53.6	85.34
syntactic emb. ⊕ fasttext ⊕ bert	86.7	65.1	**86.13**	-	-	-
			French-20K			
syntactic emb.	43.1	0	41.67	65.4	15.3	63.74
syntactic emb. ⊕ fasttext	43.2	0	41.77	-	-	-
syntactic emb. ⊕ bert	44.5	0	43.02	68.7	45.5	**67.93**
syntactic emb. ⊕ fasttext ⊕ bert	46.2	0	44.67	-	-	-

Table 2: Single-Lingual and Cross-Lingual Experimental Results on Semeval 2019 dataset

shcovich et al., 2019)[4] are given in Table 3. The results show that our model achieves state-of-the-art performance among the other parsers in English and German. The model proposed by Jiang et al. (2019) outperforms the other models in French. However, our results on unlabeled edges are still competitive with that of Jiang et al. (2019).

Error Analysis We characterize the errors of our semantic parser by con-

[4]We report the official results given in Hershcovich et al. (2019)

	English-Wiki					
	Labeled			Unlabeled		
	All	Prim.	Rem.	All	Prim.	Rem.
Tupa ♣	72.8	73.3	47.2	85.0	85.8	48.4
HLT@SUDA ♠	77.4	77.9	52.2	87.2	87.9	52.5
UC Davis ♡	72.2	73.0	0	85.5	86.4	0
CUNY-PekingU ♦	71.8	72.3	49.5	84.5	85.2	50.1
DANGNT@UIT.VNU-HCM ⋆	70.0	70.7	0	81.7	82.6	0
GCN-Sem ▽	65.7	66.4	0	80.9	81.8	0
Self-Attentive UCCA Parser	**79.7**	80.2	55.4	**89.6**	90.3	55.3
	English-20K					
HLT@SUDA ♠	72.7	73.6	31.2	85.2	86.4	32.1
Tupa ♣	67.2	68.2	23.7	82.2	83.5	24.3
CUNY-PekingU ♦	66.9	67.9	27.9	82.3	83.6	29.0
GCN-Sem ▽	62.6	63.7	0	80.0	81.4	0
Self-Attentive UCCA Parser	**74.62**	76.2	29.3	**87.69**	89.7	30.1
	German-20K					
HLT@SUDA ♠	84.9	85.4	64.1	92.8	93.4	64.7
Tupa ♣	79.1	79.6	59.9	90.3	91.0	60.5
TüPa ⊛	78.1	78.8	40.8	89.4	90.3	41.2
XLangMo	78.0	78.4	61.1	89.4	90.1	61.4
MaskParse@Deskiñ ⊘	74.2	74.8	47.3	87.1	88.0	47.6
Self-Attentive UCCA Parser	**86.13**	86.7	65.1	**94.1**	94.4	64.5
	French-20K					
HLT@SUDA ♠	**75.2**	76.0	43.3	**86.0**	87.0	45.1
XLangMo	65.6	66.6	13.3	81.5	82.8	14.1
MaskParse@Deskiñ ⊘	65.4	66.6	24.3	80.9	82.5	25.8
Tupa ♣	48.7	49.6	2.4	74.0	75.3	3.2
TüPa ⊛	45.6	46.4	0	73.4	74.6	0
Self-Attentive UCCA Parser	67.93	68.7	45.5	84.8	85.5	54.6

Table 3: Comparative F-1 results of our model with other participants of UCCA framework at Semeval 2019. (♣: (Hershcovich et al., 2017), ♠: (Jiang et al., 2019), ♡: (Yu & Sagae, 2019), ⋆: Nguyen & Tran (2019), ⊛: Pütz & Glocker (2019), ▽: Taslimipoor et al. (2019), ⊘: Marzinotto et al. (2019)).

ducting further experiments to analyze the effects of structural and linguistic features of sentences on the accuracy of the parser.

- Sentence Length: The results for the different sentence lengths are given in Table 4. The results show that the longer the sentences are, the lower the F-1 scores are for the remote edges except for the English-Wiki dataset. Since UCCA can be extended to represent paragraph-level annotation, the semantic structure of longer sentences can also be efficiently represented using the UCCA framework. The results obtained from the primary edges for longer sentences already confirm this. The frequency of a remote edge is 1 or 0 in each sentence in the dataset, which does not let the model learn the remote edges properly. Therefore, the efficiency of the model is more crucial for primary edges compared to remote edges.

- Semantic Categories: We analyze the results of each semantic category to further evaluate the performance of the model according to each category. The results obtained from each category are given in Table 5. The frequency of Adverbial (A), Function (F), Ground (G), Linker (L), Connector (C) and State (S) are comparatively lower than the other semantic categories in the dataset[5]. While the model struggles with predicting categories with low frequency, more frequent categories are learned more accurately by the model.

Zero-shot and Few-shot Cross-Lingual Model: Zero-shot learning (Wang et al., 2019b; Tran & Bisazza, 2019) and few-shot learning (Lauscher et al., 2020) have recently shown outstanding success in various NLP tasks such as dependency parsing and text classification. Zero-shot cross-lingual model is used when no or few annotated examples are available in the target language. In contrast, few-shot cross-lingual model is used when a small amount of training data is available during training. Due to the insufficient size of the French dataset,

[5]The details of the dataset can be found in (Hershcovich et al., 2019).

Table 4: Results by length

	English-Wiki			English-20K			German-20K			French-20K		
Sent. Len.	Prim.	Rem.	All	Prim.	Rem.	All	Prim.	Rem.	All	Prim.	Rem.	All
<= 10	80.8	50.8	79.76	73.2	30.1	72.33	87.2	66.7	86.61	68.2	45.4	67.42
<= 20	80.4	50.9	79.42	76.9	29.0	75.89	84.5	60.3	83.75	67.5	45.3	66.72
<= 30	79.8	63.0	79.18	75.8	26.6	74.74	83.4	59.2	82.67	69.1	42.7	68.17
<= 40	82.4	61.0	81.68	74.5	26.4	73.58	83.0	48.5	82.01	66.5	41.4	65.58
<= 50	78.5	61.5	77.93	74.2	0	72.55	82.9	49.2	81.84	66.8	39.8	65.83

Table 5: F-1 measure of predicting Primary edges and their labels

	Scene Elements				Non-Scene Elements				Inter-Scene Relations			Other
dataset	P	S	A	D	C	E	N	R	H	L	G	F
English-Wiki	0.68	0.24	0.77	0.67	0.83	0.80	0.87	0.86	0.75	0.56	0.65	0.73
English-20K	0.73	0.23	0.67	0.54	0.81	0.78	0.82	0.86	0.60	0.72	0.25	0.71
German-20K	0.79	0.27	0.81	0.77	0.90	0.87	0.30	0.92	0.79	0.88	0.71	0.88
French-20K	0.68	0.24	0.58	0.32	0.78	0.71	0.75	0.83	0.49	0.59	0.46	0.41

we performed both few-shot and zero-shot learning for French as part of the cross-lingual experiments.

In the zero-shot setting, we performed cross-lingual experiments without using the French dataset during training, whereas we included the French dataset in few-shot learning during training. The results are given in Table 6. The results show that even a small amount of data significantly improves the results in few-shot learning compared to zero-shot learning.

	Labeled			Unlabeled		
	Primary	Remote	Avg	Primary	Remote	Avg
single-lingual	46.2	0	44.67	68.9	0	66.62
zero-shot	57.2	16.2	56.42	78.6	16.2	76.53
few-shot	**68.7**	**45.5**	**67.93**	**85.8**	**54.6**	**84.48**

Table 6: Effect of French dataset on cross-lingual model

5.2. Semantic Textual Similarity

Datasets We evaluated the Siamese-RvNN model on several STS tasks using the output of the semantic parser model. We evaluated on 7 datasets that provide labels between 0 and 5 that correspond to a degree of semantic similarity:

- **SICK Dataset (Marelli et al., 2014)** is compiled for sentence level semantic similarity/relatedness task.
- **STS Datasets (Agirre et al., 2012, 2013, 2014, 2015, 2016; Cer et al., 2017)** involve 6 different datasets released by SemEval in years between 2012 to 2017 for the STS task.

We also evaluated the Siamese-RvNN model on 7 other transfer learning tasks with given datasets:

- Movie Review (MR) (Pang & Lee, 2005) is a dataset annotated for sentiment classification task with 2 classes (binary classification).

- Customer Review (CR) (Hu & Liu, 2004) is a dataset annotated for product review classification task with 2 classes (binary classification).

- Subjectivity / Objectivity (SUBJ) (Pang & Lee, 2004) is a dataset annotated for subjectivity objectivity classification task with 2 classes (binary classification).

- Multi-Perspective Question and Answering (MPQA) (Wiebe et al., 2005) is a dataset annotated for opinion polarity classification task with 2 classes (binary classification).

- Stanford Sentiment Analysis 2 (SST-2) (Socher et al., 2013) is a dataset annotated for sentiment classification task with 2 classes (binary classification).

- Text Retrieval Conference (TREC) (Voorhees & Tice, 2000) is a dataset annotated for question type classification task with 6 classes.

- The Microsoft Research Paraphrase Corpus (MRPC) (Dolan et al., 2004) is a dataset annotated for paraphrase detection task with 2 classes (binary classification).

Evaluation Metric We use the SentEval toolkit (Conneau & Kiela, 2018) to evaluate the results obtained from the STS task and use Spearman's rank correlation ρ as the evaluation metric (Reimers et al., 2016). We use the accuracy metric in all transfer learning tasks and follow the same configurations defined in SentEval [6].

Hyperparameters and Implementation Details For the STS task, we use a combination of SNLI (Bowman et al., 2015) and MNLI (Williams et al., 2018) as in Reimers et al. (2019); Gao et al. (2021) to finetune embeddings. In all STS experiments, we assigned coefficients $\alpha = \beta = 1$. We used a batch size of 16, the Adam optimizer (Kingma & Ba, 2014) for training, BERT (Devlin et al.,

[6] `https://github.com/facebookresearch/SentEval`

2019) (base-uncased) pre-trained embeddings, a dropout of 0.2 and a learning rate of $1e-4$ in the proposed Siamese-RvNN model.

We use semantically-informed sentence embeddings obtained from the Recursive Neural Network to train a logistic regression classifier for the transfer learning tasks. In all transfer-learning experiments, 10-fold cross-validation is performed as used in (Reimers et al., 2019).

Results Table 7 shows the evaluation results for 7 STS tasks. The Siamese-RvNN model significantly improves the results on all datasets except SICK-R. Our model outperforms the previous best average Spearman's correlation with an improvement from 83.76 to 83.98, indicating that semantic annotation with UCCA helps to learn better sentence embeddings than other models such as SBERT (Reimers et al., 2019) that uses pre-trained BERT along with Siamese and triple networks, and SimCSE (Gao et al., 2021), a simple contrastive sentence embedding framework, which uses pre-trained BERT and ROBERTA with an MLP layer that can generate sentence embeddings from either unlabeled or labeled data. These two models use only pretrained language models to capture sentence embeddings without using any semantic structure of the text.

Transfer Learning Task Results Table 8 shows the evaluation results of the transfer learning tasks. Siamese-RvNN achieves the best performance in 2 out of 7 tasks. Although we were not able to outperform the state-of-the-art results on average, we generally achieved competitive results compared to other methods.

6. Conclusion

We propose two models for two subtasks, namely a self-attentive semantic parser that uses both syntactic and semantic embeddings to learn UCCA semantic-graphs, and Siamese Recursive Neural Network (Siamese-RvNN) to generate semantically informed sentence embeddings using UCCA semantic-graphs for the semantic textual similarity task. For the second task, we used the parser model to generate UCCA representations. We obtained state-of-the-

Model	STS12	STS13	STS14	STS15	STS16	STS-B	SICK-R	Avg.
Siamese-RvNN	**78.46**	**87.75**	**82.61**	86.83	**84.39**	**87.03**	80.82	**83.98**
Universal Sentence Encoder ♣	64.49	67.80	64.61	76.83	73.18	74.92	76.69	71.22
SimCSE-BERT$_{base}$ ♠	75.30	84.67	80.19	85.40	80.82	84.25	80.39	81.57
SimCSE-Roberta$_{large}$ ♠	77.46	87.27	82.36	**86.66**	83.93	86.70	**81.95**	83.76
SBERT-NLI-large ♡	72.27	78.46	74.90	80.99	76.25	79.23	73.75	76.55
SRoBERTa-NLI-large ♡	74.53	77.00	73.18	81.85	76.82	79.10	74.29	76.68

Table 7: Task performance on STS tasks (Spearman's correlation X 100). ♣ results from Cer et al. (2018); ♠ results from Gao et al. (2021); ♡ results from Reimers et al. (2019)

Model	MR	CR	SUBJ	MPQA	SST-2	TREC	MRPC	Avg.
Siamese-RvNN	84.25	89.54	**94.68**	89.10	**91.25**	88.42	74.93	87.45
BiLSTM ♢	81.1	86.3	92.4	90.2	-	-	-	-
Universal Sentence Encoder ♣	80.09	85.19	93.98	86.70	86.38	93.2	70.14	85.10
SimCSE-BERT$_{base}$ ♠	83.64	89.43	94.39	89.86	88.96	**89.60**	**76.00**	87.41
SBERT-NLI-large ♡	**84.88**	**90.07**	94.52	**90.33**	90.66	87.4	75.94	**87.69**

Table 8: Transfer test results of the proposed model (measured as accuracy). ♣ results from Cer et al. (2018); ♠ results from Gao et al. (2021); ♡ results from Reimers et al. (2019); ♢ results from Conneau et al. (2017)

art semantic parsing results in English and German. We also observed that the cross-lingual model performs better for low-resource languages. The results for the French are remarkably better with the cross-lingual model.

Our proposed Siamese-RvNN model outperforms other sentence embedding approaches on semantic textual similarity task. We also obtained competitive results on transfer learning tasks.

In the future, we would like to investigate the contribution of phrases to the STS task and perform both semantic parsing and sentence embeddings generation in multilingual setting. In addition, our semantic parsing models suggest that further research in cross-lingual learning has the potential to lead to improvements in creating a dataset for languages without resources.

Acknowledgments

This study was carried out as part of a PhD study pursued by the first author at the Graduate School of Science and Engineering of Hacettepe University.

References

Abend, O., & Rappoport, A. (2013). UCCA: A Semantics-based Grammatical Annotation Scheme. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)–Long Papers* (pp. 1–12).

Abend, O., & Rappoport, A. (2017). The state of the art in semantic representation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 77–89).

Abzianidze, L., Bjerva, J., Evang, K., Haagsma, H., van Noord, R., Ludmann, P., Nguyen, D.-D., & Bos, J. (2017). The Parallel Meaning Bank: Towards a Multilingual Corpus of Translations Annotated with Compositional Meaning Representations. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers* (pp. 242–247).

Afzal, N., Wang, Y., & Liu, H. (2016). MayoNLP at Semeval-2016 Task 1: Semantic textual similarity based on lexical semantic net and deep learning semantic model. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)* (pp. 674–679).

Agirre, E., Banea, C., Cardie, C., Cer, D., Diab, M., Gonzalez-Agirre, A., Guo, W., Lopez-Gazpio, I., Maritxalar, M., Mihalcea, R. et al. (2015). Semeval-2015 task 2: Semantic textual similarity, English, Spanish and pilot on interpretability. In *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)* (pp. 252–263).

Agirre, E., Banea, C., Cardie, C., Cer, D., Diab, M., Gonzalez-Agirre, A., Guo, W., Mihalcea, R., Rigau, G., & Wiebe, J. (2014). SemEval-2014 Task 10: Multilingual Semantic Textual Similarity. In *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014)* (pp. 81–91).

Agirre, E., Banea, C., Cer, D., Diab, M., Gonzalez Agirre, A., Mihalcea, R., Rigau Claramunt, G., & Wiebe, J. (2016). Semeval-2016 task 1: Semantic textual similarity, monolingual and cross-lingual evaluation. In *SemEval-2016. 10th International Workshop on Semantic Evaluation; 2016 Jun 16-17; San Diego, CA. Stroudsburg (PA): ACL; 2016. p. 497-511.*. ACL (Association for Computational Linguistics).

Agirre, E., Cer, D., Diab, M., & Gonzalez-Agirre, A. (2012). Semeval-2012 task 6: A pilot on semantic textual similarity. In ** SEM 2012: The First Joint Conference on Lexical and Computational Semantics–Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)* (pp. 385–393).

Agirre, E., Cer, D., Diab, M., Gonzalez-Agirre, A., & Guo, W. (2013). SEM 2013 shared task: Semantic textual similarity. In *Second joint conference on lexical and computational semantics (* SEM), volume 1: proceedings of* *the Main conference and the shared task: semantic textual similarity* (pp. 32–43).

Ando, R. K. (2000). Latent semantic space: Iterative scaling improves precision of inter-document similarity measurement. In *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 216–223).

Arviv, O., Cui, R., & Hershcovich, D. (2020). HUJI-KU at MRP 2020: Two Transition-based Neural Parsers. In *Proceedings of the CoNLL 2020 Shared Task: Cross-Framework Meaning Representation Parsing* (pp. 73–82).

Bai, H., & Zhao, H. (2019). SJTU at MRP 2019: A transition-based multi-task parser for 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* (pp. 86–94).

Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., & Schneider, N. (2013). Abstract meaning representation for sembanking. In *Proceedings of the 7th linguistic annotation workshop and interoperability with discourse* (pp. 178–186).

Blloshmi, R., Tripodi, R., & Navigli, R. (2020). Enabling Cross-Lingual AMR Parsing with Transfer Learning Techniques. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 2487–2500).

Bojanowski, P., Grave, É., Joulin, A., & Mikolov, T. (2017a). Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, *5*, 135–146.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017b). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, *5*, 135–146.

Bougiatiotis, K., & Giannakopoulos, T. (2018). Enhanced movie content similarity based on textual, auditory and visual information. *Expert Systems with Applications*, *96*, 86–102.

Bowman, S., Angeli, G., Potts, C., & Manning, C. D. (2015). A large annotated corpus for learning natural language inference. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 632–642).

Cai, D., & Lam, W. (2020). AMR Parsing via Graph-Sequence Iterative Inference. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 1290–1301).

Cao, J., Zhang, Y., Youssef, A., & Srikumar, V. (2019). Amazon at MRP 2019: Parsing Meaning Representations with Lexical and Phrasal Anchoring. In *Proceedings of the Shared Task on Cross-Framework Meaning Representation Parsing at the 2019 Conference on Natural Language Learning* (pp. 138–148).

Cer, D., Diab, M., Agirre, E., Lopez-Gazpio, I., & Specia, L. (2017). SemEval-2017 Task 1: Semantic Textual Similarity Multilingual and Crosslingual Focused Evaluation. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)* (pp. 1–14).

Cer, D., Yang, Y., Kong, S., Hua, N., Limtiaco, N., John, R. S., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Sung, Y., Strope, B., & Kurzweil, R. (2018). Universal Sentence Encoder. *CoRR*, *abs/1803.11175*. URL: `http://arxiv.org/abs/1803.11175`. `arXiv:1803.11175`.

Chan, Y. S., & Ng, H. T. (2008). MAXSIM: A maximum similarity metric for machine translation evaluation. In *Proceedings of ACL-08: HLT* (pp. 55–62).

Chappelier, J.-C., & Rajman, M. (1998). A generalized CYK algorithm for parsing stochastic CFG. In *Proc. of 1st Workshop on Tabulation in Parsing and Deduction (TAPD'98)* CONF (pp. 133–137).

Chaybouti, S., Saghe, A., & Shabou, A. (2021). EfficientQA: a RoBERTa Based Phrase-Indexed Question-Answering System. *CoRR*, *abs/2101.02157*. URL: `https://arxiv.org/abs/2101.02157`. `arXiv:2101.02157`.

Che, W., Dou, L., Xu, Y., Wang, Y., Liu, Y., & Liu, T. (). 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* (pp. 76–85).

Cheng, X. (2021). Dual-View Distilled BERT for Sentence Embedding. *CoRR*, *abs/2104.08675*. URL: `https://arxiv.org/abs/2104.08675`. `arXiv:2104.08675`.

Chopra, S., Hadsell, R., & LeCun, Y. (2005). Learning a similarity metric discriminatively, with application to face verification. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (pp. 539–546 vol. 1). volume 1. doi:`10.1109/CVPR.2005.202`.

Conneau, A., & Kiela, D. (2018). SentEval: An Evaluation Toolkit for Universal Sentence Representations. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Conneau, A., Kiela, D., Schwenk, H., Barrault, L., & Bordes, A. (2017). Supervised Learning of Universal Sentence Representations from Natural Language Inference Data. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 670–680).

Craw, S. (2017). Manhattan Distance. In C. Sammut, & G. I. Webb (Eds.), *Encyclopedia of Machine Learning and Data Mining* (pp. 790–791). Boston, MA: Springer US. URL: `https://doi.org/10.1007/978-1-4899-7687-1_511`. doi:`10.1007/978-1-4899-7687-1_511`.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171–4186).

Dohare, S., & Karnick, H. (2017). Text Summarization using Abstract Meaning Representation. *CoRR*, *abs/1706.01678*. URL: `http://arxiv.org/abs/1706.01678`. `arXiv:1706.01678`.

Dolan, W. B., Quirk, C., & Brockett, C. (2004). Unsupervised Construction of Large Paraphrase Corpora: Exploiting Massively Parallel News Sources. In *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics* (pp. 350–356).

Donatelli, L., Fowlie, M., Groschwitz, J., Koller, A., Lindemann, M., Mina, M., & Weißenhorn, P. (2019). Saarland at MRP 2019: Compositional parsing across all graphbanks. In *Proceedings of the Shared Task on Cross-Framework Meaning Representation Parsing at the 2019 Conference on Natural Language Learning* (pp. 66–75).

Dou, L., Feng, Y., Ji, Y., Che, W., & Liu, T. (2020). HIT-SCIR at MRP 2020: Transition-based parser and iterative inference parser. In *Proceedings of the CoNLL 2020 Shared Task: Cross-Framework Meaning Representation Parsing* (pp. 65–72).

Droganova, K., Kutuzov, A., Mediankin, N., & Zeman, D. (2019). ÚFAL-Oslo at MRP 2019: Garage Sale Semantic Parsing. In *Proceedings of the Shared Task on Cross-Framework Meaning Representation Parsing at the 2019 Conference on Natural Language Learning* (pp. 158–165).

Flanigan, J., Thomson, S., Carbonell, J. G., Dyer, C., & Smith, N. A. (2014). A discriminative graph-based parser for the abstract meaning representation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1426–1436).

Gabrilovich, E., Markovitch, S. et al. (2007). Computing semantic relatedness using Wikipedia-based explicit semantic analysis. In *IJcAI* (pp. 1606–1611). volume 7.

Gao, T., Yao, X., & Chen, D. (2021). SimCSE: Simple Contrastive Learning of Sentence Embeddings. *CoRR*, *abs/2104.08821*. URL: `https://arxiv.org/abs/2104.08821`. `arXiv:2104.08821`.

Gutiérrez, Y., Vázquez, S., & Montoyo, A. (2016). A semantic framework for textual data enrichment. *Expert Systems with Applications*, *57*, 248–269.

He, H., Gimpel, K., & Lin, J. (2015). Multi-perspective sentence similarity modeling with convolutional neural networks. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 1576–1586).

He, H., & Lin, J. (2016a). Pairwise word interaction modeling with deep neural networks for semantic similarity measurement. In *Proceedings of the 2016 conference of the north American chapter of the Association for Computational Linguistics: human language technologies* (pp. 937–948).

He, H., & Lin, J. (2016b). Pairwise Word Interaction Modeling with Deep Neural Networks for Semantic Similarity Measurement. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 937–948). San Diego, California: Association for Computational Linguistics. URL: `https://www.aclweb.org/anthology/N16-1108`. doi:`10.18653/v1/N16-1108`.

Hershcovich, D., Abend, O., & Rappoport, A. (2017). A Transition-Based Directed Acyclic Graph Parser for UCCA. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1127–1138).

Hershcovich, D., Abend, O., & Rappoport, A. (2018). Multitask Parsing Across Semantic Representations. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 373–385).

Hershcovich, D., Aizenbud, Z., Choshen, L., Sulem, E., Rappoport, A., & Abend, O. (2019). SemEval-2019 Task 1: Cross-lingual Semantic Parsing with UCCA. In *Proceedings of the 13th International Workshop on Semantic Evaluation* (pp. 1–10).

Hill, F., Cho, K., & Korhonen, A. (2016). Learning Distributed Representations of Sentences from Unlabelled Data. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1367–1377).

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168–177).

Issa, F., Damonte, M., Cohen, S. B., Yan, X., & Chang, Y. (2018). Abstract meaning representation for paraphrase detection. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)* (pp. 442–452).

Jiang, J. J., & Conrath, D. W. (1997). Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy. In *Proceedings of the 10th Research on Computational Linguistics International Conference* (pp. 19–33).

Jiang, W., Li, Z., Zhang, Y., & Zhang, M. (2019). HLT@ SUDA at SemEval-2019 Task 1: UCCA Graph Parsing as eConstituent Tree Parsing. In *Proceedings of the 13th International Workshop on Semantic Evaluation* (pp. 11–15).

Jimenez, S., Becerra, C., & Gelbukh, A. (2012). Soft cardinality: A parameterized similarity function for text comparison. In ** SEM 2012: The First Joint Conference on Lexical and Computational Semantics–Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)* (pp. 449–453).

Kapanipathi, P., Abdelaziz, I., Ravishankar, S., Roukos, S., Gray, A. G., Astudillo, R. F., Chang, M., Cornelio, C., Dana, S., Fokoue, A., Garg, D., Gliozzo, A., Gurajada, S., Karanam, H., Khan, N., Khandelwal, D., Lee, Y., Li, Y., Luus, F. P. S., Makondo, N., Mihindukulasooriya, N., Naseem, T., Neelam, S., Popa, L., Reddy, R. G., Riegel, R., Rossiello, G., Sharma, U., Bhargav, G. P. S., & Yu, M. (2020). Question Answering over Knowledge Bases by Leveraging Semantic Parsing and neuro-symbolic reasoning. *CoRR*, *abs/2012.01707*. URL: `https://arxiv.org/abs/2012.01707`. `arXiv:2012.01707`.

Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. *CoRR*, *abs/1408.5882*. URL: `http://arxiv.org/abs/1408.5882`. `arXiv:1408.5882`.

Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. URL: `https://arxiv.org/abs/1412.6980`. doi:`10.48550/ARXIV.1412.6980`.

Kiros, R., Zhu, Y., Salakhutdinov, R., Zemel, R. S., Torralba, A., Urtasun, R., & Fidler, S. (2015). Skip-thought vectors. In *Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 2* (pp. 3294–3302).

Kitaev, N., & Klein, D. (2018). Multilingual Constituency Parsing with Self-Attention and Pre-Training. *CoRR*, *abs/1812.11760*. URL: `http://arxiv.org/abs/1812.11760`. `arXiv:1812.11760`.

Koreeda, Y., Morio, G., Morishita, T., Ozaki, H., & Yanai, K. (2019). Hitachi at MRP 2019: Unified Encoder-to-Biaffine Network for 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* (pp. 114–126).

Lai, S., Lo, C. H., Leung, K. S., & Leung, Y. (2019). CUHK at MRP 2019: Transition-based parser with cross-framework variable-arity resolve action. In *Proceedings of the Shared Task on Cross-Framework Meaning Representation Parsing at the 2019 Conference on Natural Language Learning* (pp. 104–113).

Lauscher, A., Ravishankar, V., Vulić, I., & Glavaš, G. (2020). From Zero to Hero: On the Limitations of Zero-Shot Language Transfer with Multilingual Transformers. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 4483–4499).

Li, B., Zhou, H., He, J., Wang, M., Yang, Y., & Li, L. (2020). On the Sentence Embeddings from BERT for Semantic Textual S0imilarity. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 9119–9130).

Li, Y., McLean, D., Bandar, Z. A., O'shea, J. D., & Crockett, K. (2006). Sentence similarity based on semantic nets and corpus statistics. *IEEE transactions on knowledge and data engineering*, *18*, 1138–1150.

Li, Z., Zhao, H., Zhang, Z., Wang, R., Utiyama, M., & Sumita, E. (2019). SJTU-NICT at MRP 2019: Multi-Task Learning for End-to-End Uniform Semantic Graph Parsing. In *Proceedings of the Shared Task on Cross-Framework Meaning Representation Parsing at the 2019 Conference on Natural Language Learning* (pp. 45–54).

Liu, C., Dahlmeier, D., & Ng, H. T. (2011). Better evaluation metrics lead to better machine translation. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (pp. 375–384).

Liu, F., Flanigan, J., Thomson, S., Sadeh, N. M., & Smith, N. A. (2018). Toward Abstractive Summarization Using Semantic Representations. *CoRR*, *abs/1805.10399*. URL: `http://arxiv.org/abs/1805.10399`. `arXiv:1805.10399`.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *CoRR*, *abs/1907.11692*. URL: `http://arxiv.org/abs/1907.11692`. `arXiv:1907.11692`.

Logeswaran, L., & Lee, H. (2018). An efficient framework for learning sentence representations. *CoRR*, *abs/1803.02893*. URL: `http://arxiv.org/abs/1803.02893`. `arXiv:1803.02893`.

Lopez-Gazpio, I., Maritxalar, M., Lapata, M., & Agirre, E. (2019). Word n-gram attention models for sentence similarity and inference. *Expert Systems with Applications*, *132*, 1–11.

Lyu, W., Huang, S., Khan, A. R., Zhang, S., Sun, W., & Xu, J. (2019). CUNY-PKU Parser at SemEval-2019 Task 1: Cross-Lingual Semantic Parsing with UCCA. In *Proceedings of the 13th International Workshop on Semantic Evaluation* (pp. 92–96).

Majumder, G., Pakray, P., Gelbukh, A., & Pinto, D. (2016). Semantic textual similarity methods, tools, and applications: A survey. *Computación y Sistemas*, *20*, 647–665.

Marelli, M., Menini, S., Baroni, M., Bentivogli, L., Bernardi, R., & Zamparelli, R. (2014). A SICK cure for the evaluation of compositional distributional semantic models. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavik, Iceland: European Language Resources Association (ELRA).

Martinez-Gil, J., & Chaves-Gonzalez, J. M. (2020). A novel method based on symbolic regression for interpretable semantic similarity measurement. *Expert Systems with Applications*, *160*, 113663.

Marzinotto, G., Heinecke, J., & Damnati, G. (2019). MaskParse@ Deskin at SemEval-2019 Task 1: Cross-lingual UCCA Semantic Parsing using Recursive Masked Sequence Tagging. In *Proceedings of the 13th International Workshop on Semantic Evaluation* (pp. 107–112).

Mihalcea, R., Corley, C., Strapparava, C. et al. (2006). Corpus-based and knowledge-based measures of text semantic similarity. In *AAAI* (pp. 775–780). volume 6.

Mueller, J., & Thyagarajan, A. (2016). Siamese recurrent architectures for learning sentence similarity. In *Proceedings of the AAAI conference on artificial intelligence*. volume 30.

Na, S.-H., & Min, J. (2020). JBNU at MRP2020: AMR parsing using a joint state model for graph-sequence iterative inference. In *Proceedings of the CoNLL 2020 Shared Task: Cross-Framework Meaning Representation Parsing* (pp. 83–87).

Na, S.-H., Min, J., Park, K., Shin, J.-H., & Kim, Y.-G. (2019). JBNU at MRP 2019: Multi-level Biaffine Attention for Semantic Dependency Parsing. In *Proceedings of the Shared Task on Cross-Framework Meaning Representation Parsing at the 2019 Conference on Natural Language Learning* (pp. 95–103).

Nguyen, D. T., & Tran, T. (2019). DANGNT@ UIT. VNU-HCM at SemEval 2019 Task 1: Graph Transformation System from Stanford Basic Dependencies to Universal Conceptual Cognitive Annotation (UCCA). In *Proceedings of the 13th International Workshop on Semantic Evaluation* (pp. 97–101).

Nilsson, J., & Nivre, J. (2005). Pseudo-projective dependency parsing. In *Proceedings of ACL* (pp. 99–106).

Nivre, J., De Marneffe, M.-C., Ginter, F., Goldberg, Y., Hajic, J., Manning, C. D., McDonald, R., Petrov, S., Pyysalo, S., Silveira, N. et al. (2016). Universal dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 1659–1666).

Nivre, J., de Marneffe, M., Ginter, F., Hajic, J., Manning, C. D., Pyysalo, S., Schuster, S., Tyers, F. M., & Zeman, D. (2020). Universal Dependencies v2: An Evergrowing Multilingual Treebank Collection. *CoRR*, *abs/2004.10643*. URL: `https://arxiv.org/abs/2004.10643`. `arXiv:2004.10643`.

Nova, E. (2019). STRASS: A Light and Effective Method for Extractive Summarization Based on Sentence Embeddings. *ACL 2019*, *1*, 243.

Oepen, S., Abend, O., Abzianidze, L., Bos, J., Hajic, J., Hershcovich, D., Li, B., O'Gorman, T., Xue, N., & Zeman, D. (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* (pp. 1–22).

Oepen, S., Abend, O., Hajic, J., Hershcovich, D., Kuhlmann, M., O'Gorman, T., Xue, N., Chun, J., Straka, M., & Urešová, Z. (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* (pp. 1–27).

Oepen, S., & Flickinger, D. (2019). The ERG at MRP 2019: Radically compositional semantic dependencies. In *Proceedings of the Shared Task on Cross-Framework Meaning Representation Parsing at the 2019 Conference on Natural Language Learning* (pp. 40–44).

Oepen, S., Kuhlmann, M., Miyao, Y., Zeman, D., Cinková, S., Flickinger, D., Hajic, J., Ivanova, A., & Uresova, Z. (2016). Towards comparability of linguistic graph banks for semantic parsing. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 3991–3995).

Pang, B., & Lee, L. (2004). A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04)* (pp. 271–278).

Pang, B., & Lee, L. (2005). Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)* (pp. 115–124).

Peng, H., Thomson, S., & Smith, N. A. (2017). Deep Multitask Learning for Semantic Dependency Parsing. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 2037–2048).

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep Contextualized Word Representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*.

Pontes, E. L., Huet, S., Linhares, A. C., & Torres-Moreno, J. (2018). Predicting the semantic textual similarity with siamese CNN and LSTM. *CoRR*, *abs/1810.10641*. URL: `http://arxiv.org/abs/1810.10641`. `arXiv:1810.10641`.

Pütz, T., & Glocker, K. (2019). Tüpa at SemEval-2019 Task1:(Almost) feature-free Semantic Parsing. In *Proceedings of the 13th International Workshop on Semantic Evaluation* (pp. 113–118).

Qu, C., Yang, L., Qiu, M., Croft, W. B., Zhang, Y., & Iyyer, M. (2019). BERT with history answer embedding for conversational question answering. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 1133–1136).

Ramachandran, L., Cheng, J., & Foltz, P. (2015). Identifying patterns for short answer scoring using graph-based lexico-semantic text matching. In *Proceedings of the Tenth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 97–106).

Ramage, D., Rafferty, A. N., & Manning, C. D. (2009). Random walks for text semantic similarity. In *Proceedings of the 2009 workshop on graph-based methods for natural language processing (TextGraphs-4)* (pp. 23–31).

Reimers, N., Beyer, P., & Gurevych, I. (2016). Task-oriented intrinsic evaluation of semantic textual similarity. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers* (pp. 87–96).

Reimers, N., Gurevych, I., Reimers, N., Gurevych, I., Thakur, N., Reimers, N., Daxenberger, J., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Sánchez, D., Batet, M., Isern, D., & Valls, A. (2012). Ontology-based semantic similarity: A new feature-based approach. *Expert systems with applications*, *39*, 7718–7728.

Šarić, F., Glavaš, G., Karan, M., Šnajder, J., & Bašić, B. D. (2012). Takelab: Systems for measuring semantic text similarity. In ** SEM 2012: The First Joint Conference on Lexical and Computational Semantics–Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)* (pp. 441–448).

Severyn, A., & Moschitti, A. (2013). Automatic feature engineering for answer selection and extraction. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 458–467).

Shao, Y. (2017). HCTI at SemEval-2017 Task 1: Use convolutional neural network to evaluate semantic textual similarity. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)* (pp. 130–133).

Sheneamer, A. M. (2021). Multiple Similarity-based Features Blending for Detecting Code Clones using Consensus-Driven Classification. *Expert Systems with Applications*, *183*, 115364.

Socher, R., Manning, C. D., & Ng, A. Y. (2010). Learning continuous phrase representations and syntactic parsing with recursive neural networks. In *Proceedings of the NIPS-2010 deep learning and unsupervised feature learning workshop* (pp. 1–9). volume 2010.

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing* (pp. 1631–1642).

Song, L., Gildea, D., Zhang, Y., Wang, Z., & Su, J. (2019). Semantic neural machine translation using AMR. *Transactions of the Association for Computational Linguistics*, *7*, 19–31.

Straka, M., & Straková, J. (2017). Tokenizing, pos tagging, lemmatizing and parsing ud 2.0 with udpipe. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies* (pp. 88–99).

Straka, M., & Straková, J. (2019). ÚFAL MRPipe at MRP 2019: UDPipe Goes Semantic in the Meaning Representation Parsing Shared Task. In

Proceedings of the Shared Task on Cross-Framework Meaning Representation Parsing at the 2019 Conference on Natural Language Learning (pp. 127–137).

Sulem, E., Abend, O., & Rappoport, A. (2018). Simple and Effective Text Simplification Using Semantic and Neural Methods. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 162–173).

Sulem, E., Abend, O., & Rappoport, A. (2020). Semantic structural decomposition for neural machine translation. In *Proceedings of the Ninth Joint Conference on Lexical and Computational Semantics* (pp. 50–57).

Sultan, M. A., Bethard, S., & Sumner, T. (2016). DLS@ CU at Semeval-2016 Task 1: Supervised models of sentence similarity. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)* (pp. 650–655).

Sussna, M. (1993). Word sense disambiguation for free-text indexing using a massive semantic network. In *Proceedings of the second international conference on Information and knowledge management* (pp. 67–74).

Tai, K. S., Socher, R., & Manning, C. D. (2015). Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 1556–1566). Beijing, China: Association for Computational Linguistics. URL: `https://www.aclweb.org/anthology/P15-1150`. doi:`10.3115/v1/P15-1150`.

Taslimipoor, S., Rohanian, O., & Može, S. (2019). GCN-Sem at SemEval-2019 Task 1: Semantic Parsing using Graph Convolutional and Recurrent Neural Networks. In *Proceedings of the 13th International Workshop on Semantic Evaluation* (pp. 102–106).

Tien, N. H., Le, N. M., Tomohiro, Y., & Tatsuya, I. (2019). Sentence Modeling via Multiple Word Embeddings and Multi-level Comparison for Semantic Textual Similarity. *Information Processing & Management*, *56*, 102090.

Tran, K. M., & Bisazza, A. (2019). Zero-shot Dependency Parsing with Pre-trained Multilingual Sentence Representations. In *Proceedings of the 2nd Workshop on Deep Learning Approaches for Low-Resource NLP (DeepLo 2019)* (pp. 281–288).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998–6008).

Voorhees, E. M., & Tice, D. M. (2000). Building a question answering test collection. In *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 200–207).

Wang, L., Raghavan, H., Castelli, V., Florian, R., & Cardie, C. (2016). A Sentence Compression Based Framework to Query-Focused Multi-Document summarization. *CoRR*, *abs/1606.07548*. URL: `http://arxiv.org/abs/1606.07548`. `arXiv:1606.07548`.

Wang, R., Finch, A., Utiyama, M., & Sumita, E. (2017). Sentence embedding for neural machine translation domain adaptation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 560–566).

Wang, X., Huang, J., & Tu, K. (2019a). Second-Order Semantic Dependency Parsing with End-to-End Neural Networks. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 4609–4618).

Wang, Y., Che, W., Guo, J., Liu, Y., & Liu, T. (2019b). Cross-Lingual BERT Transformation for Zero-Shot Dependency Parsing. 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) (pp. 5725–5731).

White, A. S., Reisinger, D., Sakaguchi, K., Vieira, T., Zhang, S., Rudinger, R., Rawlins, K., & Van Durme, B. (2016). Universal decompositional semantics on universal dependencies. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 1713–1723).

Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, *39*, 165–210.

Williams, A., Nangia, N., & Bowman, S. (2018). A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)* (pp. 1112–1122).

Wu, X., Gao, C., Zang, L., Han, J., Wang, Z., & Hu, S. (2021). ESimCSE: Enhanced Sample Building Method for Contrastive Learning of unsupervised sentence embedding. *CoRR*, *abs/2109.04380*. URL: `https://arxiv.org/abs/2109.04380`. `arXiv:2109.04380`.

Xia, T., Wang, Y., Tian, Y., & Chang, Y. (2021). Using Prior Knowledge to Guide BERT's Attention in Semantic Textual Matching Tasks. In *Proceedings of the Web Conference 2021* (pp. 2466–2475).

Xu, W., Zhang, H., Cai, D., & Lam, W. (2021). Dynamic Semantic Graph Construction and Reasoning for Explainable Multi-hop Science Question Answering. *CoRR*, *abs/2105.11776*. URL: `https://arxiv.org/abs/2105.11776`. `arXiv:2105.11776`.

Yao, X., Van Durme, B., Callison-Burch, C., & Clark, P. (2013). Answer extraction as sequence tagging with tree edit distance. In *Proceedings of the*

2013 conference of the North American chapter of the association for computational linguistics: human language technologies (pp. 858–867).

Yih, W.-t., Toutanova, K., Platt, J. C., & Meek, C. (2011). Learning discriminative projections for text similarity measures. In *Proceedings of the fifteenth conference on computational natural language learning* (pp. 247–256).

Yin, D., Meng, T., & Chang, K.-W. (2020). SentiBERT: A Transferable Transformer-Based Architecture for Compositional Sentiment Semantics. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 3695–3706).

Yu, D., & Sagae, K. (2019). UC Davis at Semeval-2019 Task 1: DAG semantic parsing with attention-based decoder. In *Proceedings of the 13th International Workshop on Semantic Evaluation* (pp. 119–124).

Zhang, J., & Lan, Z. (2021). S-SimCSE: Sampled Sub-networks for Contrastive Learning of Sentence embedding. *CoRR*, *abs/2111.11750*. URL: `https://arxiv.org/abs/2111.11750`. `arXiv:2111.11750`.

Zhang, Y., Jiang, W., Xia, Q., Cao, J., Wang, R., Li, Z., & Zhang, M. (2019). SUDA-Alibaba at MRP 2019: Graph-Based Models with BERT. In *Proceedings of the Shared Task on Cross-Framework Meaning Representation Parsing at the 2019 Conference on Natural Language Learning* (pp. 149–157).