Deep Semantic Role Labeling for Tweets using 5W1H: *Who, What, When, Where, Why and How*

Kunal Chakma[1], Amitava Das[2], Swapan Debbarma[3]
kchakma.cse@nita.ac.in, amitava.das@iiits.in, and swapanxavier@gmail.com

[1] National Institute of Technology Agartala, India
[2] Indian Institute of Information Technology, Sricity, India
[3] National Institute of Technology Agartala, India

Abstract. Semantic Role Labeling (SRL) is considered one of the important tasks in natural language understanding and widely studied by the research community. State-of-the-art lexical resources have been in existence for defining the semantic role arguments with respect to the predicates. However, such lexical resources are complex in nature which is difficult to understand. Therefore, instead of the classical semantic role arguments, we adopted the concept of 5W1H (*Who, What, When, Where, Why and How*) for SRL. The 5W1H concept is widely used in journalism and it is much simpler and easier to understand as compared to the classical SRL lexical resources. In the recent years, recurrent neural networks (RNN) based end-to-end SRL systems have gained significant attention. However, all recent works have been developed for formal texts. This paper reports on the implementation of a deep neural network using the attention mechanism for extracting the 5W1H from tweets. Our implementation reports an F-1 score of **88.21** which outperforms other recent Twitter SRL system by **28.72**.

1 Introduction

Semantic Role Labeling (SRL) is a natural language understanding task that extracts semantic constituents of a sentence for answering *who* did *what* to *whom*. SRL is a shallow semantic parsing task whose primary goal is to identify the semantic roles and the relationship among them and therefore, has wide application in other Natural Language Processing (NLP)tasks such as Information Extraction [1] ,Question Answering([2], [3], [**?**]), Machine Translation ([4], [5] and [6]) and Multi-document Abstractive Summarization [7].

The study of semantic roles was first introduced by the Indian grammarian Panini [8]. in his "*Karaka*" theory. *Karaka* theory assigns generic semantic roles to words in a natural language sentence. The relationship of the arguments with the verb is described using relations called *Karaka* relations. *Karaka* relations describe the way in which arguments participate in the action described by the verb. Gildea and Jurafsky [9] developed the first automatic semantic role labeling system based on FrameNet [10].Subsequent works ([11],[12] and [13]) are considered as traditional approaches that explored the syntactic features for capturing the overall sentence structure. There are several lexical resources available for SRL such as PropBank [14], FramNet [10] and VerbNet [15] that define different semantic role sets. Most of the SRL works are based on the PropBank [14] role set and use the CoNLL-2005 [16] shared task datasets. These datasets are mainly sections from the World Street Journal (WSJ) articles. Though there have been significant developments in studying SRL, most of the state-of-the-art SRL systems have been developed for formal texts only. There exists only a few articles([17], [18], [19] and [20]) on Twitter SRL.

Twitter is a micro-blogging site that allows a user to post texts (often known as *tweets*) within the limit of 280 characters. Tweets are often found to be informal in nature and tend to be without proper grammatical structures. Use of phonetic typing, abbreviations, word play and emoticons are very common in tweets. Therefore, performing SRL on such informal texts is a difficult task.

For performing SRL on tweets, annoted corpora is a prerequisite. Annotation based on the PropBank role set requires sufficient knowledge about the constituent arguments of a predicate. Therefore, instead of using the PropBank role set, we adopted the concept of 5W1H (*Who, What, When, Where, Why, How*) as described in [21]. 5W1H concept is widely used in journalism because an article is considered complete only when all the 5W1H are present. The concept of 5W1H is similar to the *Karaka* relations and easy to understand.
The major contributions of this paper are:

- Development of a corpus for SRL on tweets.
- Development of a Deep Neural Network for SRL on tweets.

2 Proposed Work

2.1 SRL

Most of the state-of-the-art SRL systems are based on the PropBank argument role-set. For a given sentence, the goal of SRL is to identify and classify the various SRL arguments of each target verb (also known as *predicate*) as semantic roles. For example, for a given sentence say "*Hillary lost the elections in 2016*", a ProbBank based SRL system yields the following output:

$$[Hillary]_{\mathbf{ARG0}}\,[lost]_{\mathbf{V}}\,[the\ elections]_{\mathbf{ARG1}}\,[in\ 2016]_{\mathbf{ARG-TMP}}$$

In the above example, **V** represents the verb *lost* for which argument **ARG0** represents the *loser*, **ARG1** represents the *thing lost* and **ARG** – **TMP** is an adjunct that represents the timing of the event.

2.2 Defining the 5W1H

Let $w = \{w_1, w_2, ..., w_n\}$ be the sequence of words in a tweet and X be the attribute to which w is to be mapped. We therefore, assume a tweet as $\langle w, X\rangle$, where, X is the tuple $\langle WHO, WHAT, WHEN, WHERE, WHY, HOW\rangle$ in 5W1H described below.

5W1H definition:
Definition 1:Who. *It is the set of words that refer to a person, a group of people or an institution responsible for causing an action.*
Definition 2:What. *It is the set of words that refer to the people, things or abstract concepts which undergo the change of state or being affected by an action.*
Definition 3:When. *It is the set of words that refer to temporal characteristics. In tweets, the notion of time may be the days, weeks, months and year of a calender or the tick of a clock. It also refers to the observations made either before, after or during the occurance of events such as festivals, ceremonies, elections etc.*

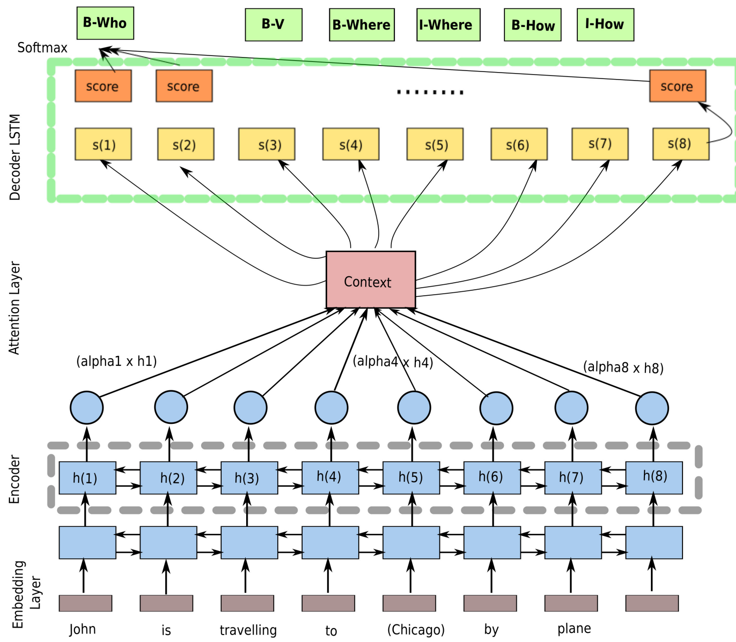


FIG. 1: Deep Neural Attention Network

Definition 4:Where. *It is the set of the words that refer to locative markers in a tweet. The notion of location is not restricted to physical locations but it also refers to abstract locations.*
Definition 5:Why. *It is the set of words that refer to the cause of an action.*
Definition 6:How. *It is the set of words that refer to the manner in which an action is performed.*

We denote $\psi_X(w)$ to represent the set of words contained in the text w and classified to the attribute X where, $X \epsilon 5W1H$. According to the Definition 1 to 6, the 5W1H model of tweets can be represented as

$$\psi_{5W1H}(w) = \bigcup_{X \epsilon 5W1H} \psi_X(w) \tag{1}$$

2.3 PropBank vs. 5W1H

The PropBank role set is categorized into basic role arguments and adjuncts. Arguments ARG0 to ARG4 are the basic roles whereas ARG-TMP, ARG-LOC, ARG-CAU etc. are the adjuncts. To perform SRL annotation based on such role set requires sufficient knowledge about them. On the other hand, the 5W1H scheme is simple to understand and convenient to annotate sentences. However, in certain cases, the 5W1H scheme does not provide fine grain details of the semantic roles. For instance, the below given tweet could be annotated with PropBank role set and 5W1H as

Tweet:

#MeToo : Aamir Khan releases a statement , steps away from doing film with accused

PropBank Role	Who (%)	What (%)	When (%)	Where (%)	Why (%)
ARG0	**41.69**	4.44	0.90	5.59	7.55
ARG1	29.84	**56.01**	32.43	32.17	41.51
ARG2	1.59	3.09	0.00	6.99	1.89
ARG3	0.00	0.05	0.00	0.00	0.00
ARG4	0.00	0.05	0.00	1.4	0.00
ARG-TMP	1.82	0.65	**50.45**	4.2	0.00
ARG-LOC	0.23	0.27	1.8	**25.87**	0.00
ARG-CAU	0.00	0.00	0.00	0.00	**11.32**
ARG-MOD	0.23	0.98	0.00	0.00	0.00
ARG-DIS	0.23	0.11	0.00	0.00	0.00
ARG-MNR	0.23	0.76	0.00	0.00	0.00
ARG-ADV	0.23	1.14	0.00	0.00	16.98
ARG-NEG	0.46	1.08	0.00	0.00	0.00
ARG-EXT	0.00	0.00	0.00	0.00	0.00
ARG-DIR	0.23	0.05	0.00	0.00	0.00

TABLE 1: Co-relation of 5W1H with PropBank Roles

PropBank tagged:

- Predicate:**releases**:
  # MeToo : $[Aamir\ Khan]_{\mathbf{ARG0}}$ $[releases]_{\mathbf{V}}$ $[a\ statement]_{\mathbf{ARG1}}$, steps away from doing film with accused
- Predicate:**steps**:
  #MeToo : $[Aamir\ Khan]_{\mathbf{ARG1}}$ releases a statement , $[steps]_{\mathbf{V}}$ $[away\ from\ doing\ film\ with\ accused]_{\mathbf{ARG\text{-}DIR}}$
- Predicate:**doing**:
  #MeToo : $[Aamir\ Khan]_{\mathbf{ARG0}}$ releases a statement , steps away from $[doing]_{\mathbf{V}}$ $[film\ with\ accused]_{\mathbf{ARG1}}$
- Predicate:**accused**:
  #MeToo : Aamir Khan releases a statement , steps away from doing film with $[accused]_{\mathbf{V}}$

5W1H tagged:

- Predicate:**releases**:
  # MeToo : $[Aamir\ Khan]_{\mathbf{Who}}$ $[releases]_{\mathbf{V}}$ $[a\ statement]_{\mathbf{What}}$, steps away from doing film with accused
- Predicate:**steps**:
  #MeToo : $[Aamir\ Khan]_{\mathbf{Who}}$ releases a statement , $[steps]_{\mathbf{V}}$ $[away\ from\ doing\ film\ with\ accused]_{\mathbf{What}}$
- Predicate:**doing**:
  #MeToo : $[Aamir\ Khan]_{\mathbf{Who}}$ releases a statement , steps away from $[doing]_{\mathbf{V}}$ $[film\ with\ accused]_{\mathbf{What}}$
- Predicate:**accused**:
  #MeToo : Aamir Khan releases a statement , steps away from doing film with $[accused]_{\mathbf{V}}$

From the above example, it is observed that adjuncts in PropBank are simply represented by one of the 5Ws or 1H.

2.4 Deep Attention Neural Network for 5W1H Extraction

In this section, we describe our deep neural network implementation based on attention mechanism (Bahdanau et al., 2015).

It is understandable from the example in section 2.1 that SRL is a sequence labeling task. The **seq2seq** model ([22]) was developed to transform an input sequence (source) to a new one (target) and both sequences can be of arbitrary lengths. The **seq2seq** model is basically an *encoder-decoder* architecture. An *encoder* encodes an input sequence and compresses the information into a context vector of a fixed length. A *decoder* is initialized with the context vector to emit the transformed output. In such architecture, only the last state of the encoder network is used as the decoder initial state. One major disadvantage of this fixed-length context vector design is incapability of remembering long sentences. In SRL, an argument may span a long sequence in a sentence. In such a scenario, the **seq2seq** model is not suitable because it often forgets earlier parts once it completes processing the whole input. The attention mechanism was introduced by (Bahdanau et al., 2015) to resolve this problem. Rather than building a single context vector out of the encoder's last hidden state, attention creates shortcuts between the context vector and the entire source input. The weights of these shortcut connections are customizable for each output element.

The input sentence is represented as a sequence of dense word vectors. These word vectors are fed to a Bi-LSTM encoder to produce a series of hidden states that represent the input. Let us consider a source sequence x of length n and try to output a target sequence y of length m:

$$\boldsymbol{x} = [x_i, x_2, ..., x_n]$$
$$\boldsymbol{y} = [y_i, y_2, ..., y_n]$$

As in the attention model of Bahdanau et al. [23], we implemented an *encoder* which is a bi-directional LSTM with forward hidden state $\underset{\boldsymbol{h_i}}{\rightarrow}$ and a backward hidden state $\underset{\boldsymbol{h_i}}{\leftarrow}$. An *encoder* state is represented by the concatenation of the hidden states:

$$\boldsymbol{h_i} = \left[\underset{\boldsymbol{h_i}}{\rightarrow}^T : \underset{\boldsymbol{h_i}}{\leftarrow}^T\right]^T, \; i = 1, \; ..., \; n \tag{2}$$

The context vector c_t depends on a sequence $(h_1, ..., h_n)$ to which an encoder maps the input sentence. The *decoder* has hidden states $s_t = f\left(s_{t-1}, \; y_{t-1}, \; c_t\right)$ for the output word at position t, $t = 1, ..., m$, where the context vector c_t is a sum of hidden states of the input sequence, weighted by alignment scores:

$$c_t = \sum_{i=1}^{n} \alpha_{t,i} \boldsymbol{h_i} \tag{3}$$

$$\alpha_{t,i} = \frac{exp\left(score\left(s_{t-1}, \boldsymbol{h_i}\right)\right)}{\sum_{i'=1}^{n} exp\left(score\left(st-1, \boldsymbol{h_{i'}}\right)\right)} \tag{4}$$

$\alpha_{t,i}$ is the alignment score assigned to the pair of input at position i and output at position t, $\left(y_t, x_i\right)$, based on how well they match. $\{\alpha_{t,i}\}$ the set of weights defining how much of each source hidden state

	Predicate	Tweet
Original tweet		Apple CEO Tim Cook attempts to unify staff in wake of Trump victory
Repeat 1	attempt	$[Apple\ CEO\ Tim\ Cook]_{\mathbf{WHO}}$ $[\mathbf{attempts}]_{\mathbf{V}}$ $[to\ unify\ staff\ in\ wake\ of\ Trump\ victory]_{\mathbf{WHAT}}$
Repeat 2	unify	$[Apple\ CEO\ Tim\ Cook]$ attempts to $[\mathbf{unify}]_{\mathbf{V}}$ $[staff\ in\ wake\ of\ Trump\ victory]_{\mathbf{WHAT}}$

TABLE 2: Dataset structure

should be considered for each output. The alignment score α is parameterized by a feed-forward network with a single hidden layer and this network is jointly trained with other parts of the model. The score function is therefore in the following form, given that $tanh$ is used as the non-linear activation function:

$$\text{score}(\boldsymbol{s}_t, \boldsymbol{h}_i) = \boldsymbol{v}_a^\top \tanh(\boldsymbol{W_a}[\boldsymbol{s_t}; \boldsymbol{h}_i]) \tag{5}$$

where both $\boldsymbol{v_a}$ and $\boldsymbol{W_a}$ are weight matrices to be learned in the alignment model.

3 Experiments

3.1 Dataset

We used two different datasets, one based on the US Elections held in November, 2016 and the other based on the #MeToo[4] campaign. The dataset on the US Elections are taken from [20] containing 3000 English tweets. For the #MeToo dataset, we crawled 248,160 tweets using hash tags such as #MeToo, #MeTooCampaign, #MeTooControversy, #MeTooIndia, as query with the *twitter4j*[5] API. After manually removing the re-tweets (tweets with RT prefix) and Non-English tweets, the dataset finally reduced to 8175 tweets . We tokenized all the tweets with CMU tokenizer [**?**]. We prepared the datasets in such a manner that for every tweet that has multiple predicates, the tweet is repeated in the corpus for each predicate (Table 2).

3.2 Model Setup

We setup our model with Keras and initialize the model with pre-trained 300-dimensional GloVe embeddings. Our vocabulary size is set to $|v| \approx 40K$ words with a maximum sequence length = 100. Both the *encoder* and *decoder* are set with latent dimensions of 256. For the *attention* layer, we use Keras RepeatVector set to the maximum length of the input sequence. The *attention* layer has two *dense* layers set with *tanh* and *softmax* respectively.

3.3 Learning

We use Adam optimizer (Kingma and Ba, 2014) and a learning rate $l_r = 0.1$. We experimented with different epochs of 5, 10 and 20 and got the best results with 20 epochs with a batch size, $BATCH\_SIZE = 1000$. The dataset was split into 90% train and 10% test sets.

[4] https://en.wikipedia.org/wiki/Me_Too_movement_(India)
[5] http://twitter4j.org/en/index.html

TABLE 3: Comparison of DRP and our system on the PropBank role identification task for the US Election corpus

System	#Tweets	F-1
DRP	3000	59.76
DeepSRL	3000	**88.48**

TABLE 4: Our System (DeepSRL) for 5W1H extraction on both the US Election and #MeToo corpus

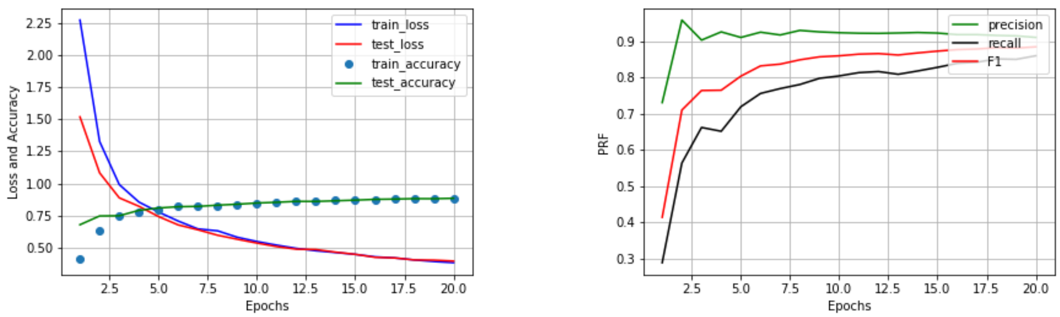
Corpus	Precision	Recall	F-1
US Elections	90.87	86.21	88.48
#MeToo	90.63	85.40	87.94
Average	90.75	85.8	**88.21**

4 Results and Analysis

The objective of our work is to extract the 5W1H from tweets. But for comparison with previous [20] SRL systems on tweets, we evaluated our system (*DeepSRL*) for PropBank role identification task. In Table 3 , we give the comparison of our system (*DeepSRL*) with the SRL system of Rudrapal et. al. [20](*DRP*) on the PropBank role identification task. On the *US Elections 2016* dataset, our system outperformed *DRP* system by overall F-1 of 28.72. In Table 4 , we give the performance of *DeepSRL* for 5W1H extraction on both the two datasets (*US Elections 2016* and *metoo movement*). *DeepSRL* achieves an overall F-1 score of 88.21 in the whole corpus. Fig 2(A) and Fig 3(A)show the loss and accuracy of our model on the train and test sets on both the datasets respectively. The three metrics of precision, recall and F1 score is plotted in fig 2(B) and fig 3(B).

5 Related Work

Though the traditional approaches of Gildea and Jurafsky [9], Pradhan et al. [11],Punyakanok et al. [12] explored the syntactic features, recently, deep neural network based implementations have outperformed the traditional approaches. Zhou and Xu [24] pioneered the work on building an end-to-end system for SRL, where they applied an 8 layered LSTM model which outperformed the previous state-of-the-art system.Roth and Lapata [25], proposed a neural classifier using dependency path embeddings to assign semantic labels to syntactic arguments. He et al. [26] developed a deep neural network with highway LSTMs and constrained decoding that improved over earlier results. Marcheggiani and Titov [27] combine their LSTM model with a graph convolutional network to encode syntactic information at word level, which improves their LSTM classifier results on the dependency-based benchmark dataset (CoNLL-09). Attention mechanism was pioneered by Bahdanau et al. [23]. Cheng et al. used [28] LSTMs and self-attention to facilitate the task of machine reading. Tan et al. [29] implemented a self-attention based neural network for SRL without explicitly modeling any syntax that outperformed the previous state-of-the-art results. Strubell et al. [30] implemented a neural network model that combines multi-head self-attention with multi-task learning across dependency parsing, part-of speech tagging, predicate detection and SRL. Their [30] method achieved the best scores on the ConLL-2005 dataset.
Liu et al. [17] are the first to study SRL on tweets. They considered only those tweets that reported news events. They mapped predicate-argument structures from news sentences to news tweets to get training data, based on which a tweet specific system is trained. They further extended their work in [18]



(A) Loss and Accuracy of train and test

(B) Precision, Recall and F-1

FIG. 2: Model performance on US Elections dataset

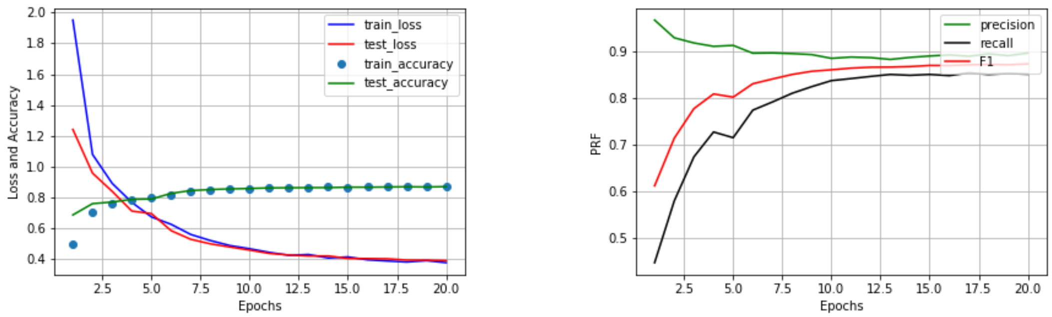
where similar tweets are grouped by clustering. Then for each cluster a two-stage SRL labeling is conducted. [19] describe a system for emotion detection from tweets.Their work mainly focuses on identification of roles for *Experiencer*, *State* and *Stimulus* of an emotion. [20] proposed an SRL system for tweets using sequential minimal optimisation (SMO). Our work adopts the 5W1H extraction for SRL using deep neural network attention mechanism of Bahdanau et al. [23].Our experiments also show the effectiveness of attention mechanism on the sequence labeling task of 5W1H.

6 Conclusion

We proposed a deep attention based neural network for the task of semantic role labeling by extracting the 5W1H from tweets. We trained our SRL models and evaluated them on the 2016 US Elections dataset that was used by a previous SRL system for tweets. We also prepared a new dataset based on the #MeToo campaign and evaluated our models. Our experimental results indicate that our models substantially improve SRL performances on tweets, leading to the new state-of-the-art. However, there are certain limitations in the 5W1H adoption as the fine grain semantic roles are ignored in such an approach.

References

1. E. Bastianelli, G. Castellucci, D.C., Basili, R.: Textual inference and meaning representation in human robot interaction. In: Proceedings of the Joint Symposium on Semantic Processing. Textual Inference and Structures in Corpora. (2013)
2. M. Surdeanu, S. Harabagiu, J.W., Aarseth, P.: Using predicate-argument structures for information extraction. In: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics. (2003)
3. A. Moschitti, P.M., Harabagiu, S.M.: Open domain information extraction via automatic semantic labeling. In: In FLAIRS Conference. (2003)
4. Knight, K., Luk, S.K.: Building a large-scale knowledge base for machine translation. In: In AAAI. (1994)
5. N. Ueffing, G.H., Sarkar, A.: Transductive learning for statistical machine translation. In: In Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics. (2007)
6. Wu, D., Fung, P.: Semantic roles for smt: a hybrid two-pass model. In: In Proceedings of Human Language Technologies, ACL (2009)



(A) Loss and Accuracy of train and test

(B) Precision, Recall and F-1

FIG. 3: Model performance on #MeToo dataset

7. Genest, P., Lapalme, G.: Framework for abstractive summarization using text-to-text generation. In: In Proceedings of the Workshop on Monolingual Text-To-Text Generation, ACL (2011)
8. Bharati, A., Chaitanya, V., Sangal, R.: Natural Languageg Processing a Paninian Perspective. Volume 1. Prentice-Hall, PHI, New Delhi (1994)
9. Gildea, D., Jurafsky, D.: Automatic labeling of semantic roles. In Association for Computational Linguistics **28** (2002) 245–288
10. Baker, C.F., Fillmore, C.J., Lowe, J.B.: The berkeley framenet project. ACL '98 Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics **1** (1998) 86–90
11. S. Pradhan, K. Hacioglu, W.W.K.M., Jurafsky, D.: Semantic role chunking combining complimentary syntactic views. In: In Proceedings of the 9th Conference on Computational Natural Language Learning. (2005)
12. Vasin Punyakanok, D.R., Yih, W.T.: Semantic role chunking combining complimentary syntactic views. In: In Proceedings of the 9th Conference on Computational Natural Language Learning. (2005)
13. M. Palmer, D.G., Xue, N.: Synthesis lectures on human language technology series., Morgan and Claypool (2010)
14. Palmer, M., Gildea, D., Kingsbury, P.: The proposition bank: A corpus annotated with semantic roles. Computational Linguistics Journal **31** (2005) 71–105
15. Schuler, K.K., Palmer, M.S.: Verbnet: a broad-coverage, comprehensive verb lexicon. University of Pennsylvania Philadelphia, PA, USA (2005)
16. Carreras, X., Màrquez, L.: Introduction to the conll-2005 shared task: semantic role labeling. In: In Proceedings of the Ninth Conference on Computational Natural Language Learning, ACL (2005)
17. Liu, X., Li, K., Han, B., Zhou, M., Jiang, L., Xiong, Z., Huang, C.: Semantic role labeling for news tweets. In Coling (2010) 698–706
18. Liu, X., Li, K., Han, B., Zhou, M., Xiong, Z.: Collective semantic role labeling for tweets with clustering. In Proceedings of the Twenty-Second international joint conference on Artificial Intelligence(IJCAI'11). **3** (2010) 1832–1837
19. Mohammad, S.M., Zhu, X., Martin, J.: Semantic role labeling of emotions in tweets. In Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. (2014) 32–41
20. Rudrapal, D., Das, A.: Semantic role labelling of english tweets through sentence boundary detection. International Journal of Intelligent Information and Database Systems **11** (2018) 225–235
21. Chakma, K., Das, A.: A 5w1h based annotation scheme for semantic role labeling of english tweets. Computación y Sistemas **22** (2018) 747–755
22. Ilya Sutskever, O.V., Le, Q.V., McCallum., A.: Sequence to sequence learning with neural networks. In: In Proceedings of NIPS. (2014)

23. Dzmitry Bahdanau, K.C., Bengio., Y.: Neural machine translation by jointly learning to align and translate. In: In Proceedings of the ICLR Conference. (2015)
24. Zhou, J., Xu., W.: End-to-end learning of semantic role labeling using recurrent neural networks. In: In Proc. of the Annual Meeting of the Association for Computational Linguistics, ACL (2015)
25. Roth, M., Lapata., M.: Neural semantic role labeling with dependency path embeddings. In: In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL (2016)
26. He, L., Lee, K., Lewis, M., Zettlemoyer, L.S.: Deep semantic role labeling: What works and what's next. In: ACL. (2017)
27. Marcheggiani, D., Titov., I.: Encoding sentences with graph convolutional networks for semantic role labeling. In: In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing., ACL (2017)
28. J. Cheng, L.D., Lapata., M.: Long short-term memory-networks for machine reading. In: In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing., ACL (2016)
29. Zhixing Tan, Mingxuan Wang, J.X.Y.C., Shi., X.: Deep semantic role labeling with self-attention. In: In Proceedings of AAAI. (2018)
30. Emma Strubell, Patrick Verga, D.A.D.W., McCallum., A.: Linguistically-informed self-attention for semantic role labeling. In: In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing., ACL (2018)