Arab World English Journal
INTERNATIONAL PEER REVIEWED JOURNAL ISSN: 2229-9327
مجلة اللغة الانكليزية في العالم العربي

Arab World English Journal (AWEJ) Volume 10. Number 4 December 2019 Pp. 251-262
DOI: https://dx.doi.org/10.24093/awej/vol10no4.19

Understanding EFL Linguistic Models through Relationship between Natural Language Processing and Artificial Intelligence Applications

Muhammed Salim Keezhatta
Department of English, College of Science and Humanities in Al-Kharj
Prince Sattam Bin Abdulaziz University, Saudi Arabia

Abstract

Natural Language Processing (NLP) platforms have recently reported a higher adoption rate of Artificial Intelligence (AI) applications. The purpose of this research is to examine the relationship between NLP and AI in the application of linguistic tasks related to morphology, parsing, and semantics. To achieve this objective, a theoretical framework was designed to investigate the direct and indirect impact of the relationship between NLP platforms and AI applications, such as machine learning and deep learning. Theoretically, this study contributes to examining the relationship between NLP platforms and AI applications through selected linguistic models from the English as a Foreign Language (EFL) perspective. Practical implications are derived from syntactic and semantic variables when AI applications are used. The results of this study suggest that AI applications can use to support NLP tasks, particularly the adaptation of deep learning applications that can prove useful in extracting analytical inferences and enhancing NLP approaches applied to EFL texts. The conclusion drawn of this study is that if NLP caters to knowledge-rich AI techniques, it can make significant advances in the linguistics disciplines of morphology, parsing, and semantics.

Keywords: Artificial intelligence, deep learning, English as a Foreign Language**,** linguistic models, machine learning, morphology, natural language processing, parsing, semantics

Cite as: Keezhatta, M. S. (2019). Understanding EFL Linguistic Models through Relationship between Natural Language Processing and Artificial Intelligence Applications. *Arab World English Journal*, *10* (4) 251-262. DOI: https://dx.doi.org/10.24093/awej/vol10no4.19

Introduction

Natural Language Processing (NLP) has multifarious applications from computational linguistics to artificial intelligence (AI). It is generally defined as the computational processing of a text in a natural human language (Fukushima & Miyake, 1982). It makes use of arithmetic algorithms in order to process information from a language (LeCun & Bengio, 1995), or design and develop computational techniques to analyze spoken or written texts in EFL context. Contextually, due to the extensive use of machine learning applications, NLP is also accepted as a branch of artificial intelligence, exploited through tools such as speech recognition, tokenization, parsing, information extraction, and many others. In making use of live AI applications, NLP can contribute in several linguistic tasks including text summarization, sentiment analysis, parts-of-speech tagging, stemming, text mining, and automated question answering with dexterous use of predictive analytics and language modeling. In simple terms, AI helps the creation of such language models that improve performance on a variety of NLP tasks

NLP relies on two primary techniques to accomplish its tasks: syntactic analysis and semantic analysis. In syntactic analysis, computer algorithms are used to understand several grammatical rules such as lemmatization, morphological segmentation, word segmentation, part-of-speech tagging, sentence breaking and stemming. Lemmatization involves reducing inflected forms of a word into a single form whereas *morphological segmentation* requires a division of words into morphemes. Simialrly, *word segmentation* requires distributing a text into distinct units and *part-of-speech tagging* requires identifying the part of speech for every word. Likewise, *parsing* refers to grammatical analysis in each sentence, *sentence breaking* means placing sentence boundaries in a large piece of text; and finally, *stemming* refers to reducing the inflected words to their root forms.

On the other hand, semantic analysis requires the use of computer algorithms to interpret words and sentence structures to derive their meaning and relationship. The techniques used in semantic analysis include *named entity recognition* (NER) or identifying such portion of a text that can be categorized into preset groups like names of people or places. Other techniques include *word sense disambiguation*, or offering to a word a meaning based on its context; *natural language generation*, referring to databases and choosing the semantic purpose of the language.

NLP also plays a critical role in supporting machine-human interactions by using AI applications. There is an extensive use of AI application models that use computational architectures and algorithmic methods to provide data-driven statistics. With the help of resources like deep learning and machine learning operations, AI is capable of comprehending NLP operations and tasks (Jones, 1994). This study is an attempt to correlate NLP and Artificial Intelligence in order to ascertain whether AI applications could be useful in understanding the linguistic domains. While NLP has been a much recognized, well established, discipline of computational linguistics, artificial intelligence, on the other hand, is widely used in data mining and data retrievals. While NLP engages in reading, deciphering, and making sense of the human languages by machine operations like online chatbots, text summarization, and auto-generated keyword tabs and even sentiment of a given text, AI applications like deep learning methods promise to offer short and long-range applications.

This study will be an addition to many recent studies (Goldberg, 2017; Young, Hazarika, Poria, & Cambria, 2018) that show AI applications are used to solve current problems in NLP.

Problem statement

A big challenge before NLP is to teach computers the way humans learn and use a language, although NLP has penetrated into common applications such as language translation (e.g. Google Translate) or word processors which employed NLP to check grammatical accuracy of texts (e.g. MS-Word, Grammarly) and Interactive Voice Response (IVR) applications (e.g call centers) or in personal assistant applications (e.g. Siri, Cortana, and Alexa). Despite its wide use, NLP is still a complex phenomenon. It is not the computer applications, but the nature of the human language and the rules that dictate a language which makes tasks difficult for NLP. The reason why a computer fails to understand regulations, because a few of these rules are high-leveled and abstract, making it difficult for computer algorithms to identify, convert and extract information from the unstructured language data made available to the computer.

In any NLP platform, a computer is required to extract meaning and collect essential data from the text provided, for which it utilizes pre-coded algorithms that are often insufficient to make a computer understand the meaning of a sentence. The ambiguity and imprecise characteristics of a natural language make it difficult for the computer to implement, and therefore, it gets obscure results. It was realized in many studies that in order to fully understand a natural language, machines need to take into account not only the literal meaning that semantics provides, but the intended message, or understanding of what the text is trying to achieve. This level is called pragmatic analysis, which is only the beginning of Artificial Intelligence (AI) applications to be introduced into the NLP techniques.

To resolve these issues, NLP techniques are assisted by AI-based neural networks to assess negative/positive/ neutral feelings of a text. These AI applications, assisted by its tools such as deep learning, machine learning, and computer vision assist computers in understanding more complex language inputs. The AI algorithms help reduce human speech into a structured ontology, attempting to make it easier to detect such linguistic characteristics related to intent, timing, locations, and sentiments. This also leads to understanding the fact that in order to be successful, NLP platforms must adopt AI applications in a wide range of fields to make linguistic understanding much comprehensible and cognitive.

This study is an attempt to understand the collaborative aspects of NLP and AI applications such as deep learning as a tool in computational linguistics. No prior study has so far examined this relationship in the context of application of EFL linguistic tasks related to morphology, parsing, and semantics. This study contributes to this research gap, especially in the context of learning English as a foreign language. This proposition is consistent with several studies, including Bengio, Goodfellow, & Courville, (2017) who recommended the use of artificial neural networks (ANNs) and parameters such as machine learning techniques. These applications, with the assistance of associated learning algorithms, build up large datasets with the help of data collection procedures and deep architectures (LeCun, Bengio & Hinton,2015; Schmidhuber, 2015; Ciresan, Meier, Masci, Gambardella, & Schmidhuber, 2011).

Literature Review

i. Artificial intelligence and NLP

AI Algorithms of NLP is basically derived from machine learning approaches, where it uses the machine learning approaches to learn the rules automatically for analyzing large volume of data. Several studies (Bengio, et al, 2017; LeCun et al, 2015; Schmidhuber, 2015; Ciresan et al, 2011) have recommended various real-time applications for NLP tasks. In the case of feature extraction on a huge volume of big data, fast-automatic processing is quite not possible by machine learning approaches. Hence, deep learning is preferred instead of machine learning to provide fast and automatic real-time applications. Deep learning is one of the advanced machine learning approaches that extends the features of artificial neural networks. Deep learning can extract and classify features automatically and fast. The primary objective of a deep learning algorithm is to classify and analyze the different patterns generated out of natural languages. Deep learning provides a multi-layer abstraction approach towards non-linear feature and pattern analysis in the field of Natural Language Processing. Deep learning can able to obtain hidden features on a large volume of data automatically. Deep learning-based NLP follows the mantra "Word2Vec" which reduces the computation and comparison complexity.

ii. Morphology

Many studies have recommended the use of morphological analyzers to accomplish NLP tasks in larger linguistic systems. For instance, Belinkov, Durrani, Dalvi, Sajjad, & Glass (2017) have drawn attention to the use of neural machine translation models where morphological knowledge of a language is first acquired and then utilized to construct translating models from English to French, German, Czech, Arabic, or Hebrew languages. These models acted as encoders and decoders a few of which followed Long short-term memory (LSTM) based systems with attention mechanisms or built upon the WIT3 (Web Inventory of Transcribed and Translated Talks) corpus (Cettolo, Girardi, & Federico, 2012; Cettolo, 2016) as AI applications. LSTM, as well as, WIT3 are widely used applications in NLP, speech recognition, and computer vision over diverse recurrent neural networks (RNNs) or Recursive Neural Networks (RvNN).

To understand their significance in performing NLP tasks, these decoders are replaced with part-of-speech (POS) taggers and morphological taggers, ensuring to preserve the internal representations by managing the weights of the encoders and their effect on the decoders. The study concluded that attention mechanisms limit the performance of encoders in order to increase the performance of decoders. It was also revealed that translating models assisted by AI applications are superior to others for learning morphology and that the output language affects the performance of encoders. Ironically, this hints at the fact that the more morphologically rich the output language, the worse would be the encoders' performance.

Luong, Socher, & Manning (2013) designed a model with RvNN, a pioneer attempt to design a morphological structure of English words by making use of Morfessor for word segmentation (Creutz & Lagus, 2007). The study generated a dataset of rare and obsolete words to construct two models—one using the context of the words and the other not. The first model insensitive to the context did not respond in certain morphological structures while the second one, sensitive to the context, performed better as it recognized the relationships between stems and also accepted such features such as the prefix "un" for constructing antonyms. The model was later tested on several other datasets (Miller & Charles, 1991; Rubenstein & Goodenough, 1965; Huang,2012) and results proved better than previous models.

Morita, Kawahara, & Kurohashi (2015) investigated a similar language model for un-segmented languages. It was built upon RNN systems with a beam search decoder and an automated labeled (Kawahara & Kurohashi, 2006) corpus replacing the earlier manually labeled corpus. The new model was found to be capable of performing morphological analysis, POS tagging, and lemmatization. The model had later been tested on Kyoto Text Corpus (Kawahara, Kurohashi & Hasida 2002) and distinct Web Document Leads corpora (Hangyo, Kawahara, & Kurohashi, 2012) with similar results, and it was reported that it out-performed all manual baselines used earlier to perform tasks.

iii. Parsing

Also, some studies have used applications like deep learning (Dehouck & Denis, 2018) in performing NLP tasks such as universal parsing or dependency parsing. More popular are however graph-based approaches that enable the construction of several parse trees in order to search the correct one. These graph-based approaches use generative models of formal grammar, based on NLP, to construct the trees (Jurafsky, 2000)] and use transition-based approaches. A pioneering work of application of deep learning for NLP tasks is that of (Socher, Perelygin, Wu, Chuang, Manning, Ng & Potts, 2013; Socher, Bauer, Manning, & Ng, 2013), which utilized RNNs with probabilistic context-free grammars (PCFGs) (Zeman et al, 2018; Nivre, 2015; Fujisaki, Jelinek, Cocke et al, 1991; Jelinek, Lafferty, & Mercer,1992; Chi & Geman,1998). However, Le and Zuidema (2014) designed the first-ever neural model to achieve state-of-the-art parsing which employed both inner and outer vector representations enabling both top-down and bottom-up flows of data.

Vinyals, Kaiser, Koo, Petrov, Sutskever, & Hinton (2015) adopted LSTM and used a Recurrent Neural Network with attention mechanism in a syntactic constituency parser, in order to undertake highly focused research on Grammar in a Foreign Language situation. The authors believed that syntactic constituency parsing was a great concern in linguistics and NLP in particular and required a wide range of applications. They highlighted the weaknesses of computational requirements of traditional parsers such as their sentence length, linear-time and like due to which they never matched state-of-the-art. So they recommended the use of more standardized parsers using generic sequence-to-sequence approaches. For this purpose, they rejected the model of Sutskever,Vinyals, & Le (2014) as it was not data-efficient and discovered that the attention model of Bahdanau, Cho, and Bengio. (2014) was highly data-efficient and compatible to even small human-annotated parsing datasets.

Chen and Manning (2014) pioneered the state of the art in both English and Chinese datasets by using a simple feed-forward neural network and a transition-based parser. This enabled them to design statistical models. Weiss, Alberti, Collins, Petrov (2015) extended Chen and Manning's experiment by using a deeper neural network and Andor et al. (2016) also used a feed-forward network for NLP tasks such as part-of-speech tagging, sentence compression, etc. Dyer, Kuncoro, Ballesteros, and Smith (2016) recommended recurrent neural network grammar models for parsing and emphasized upon taking a top-down approach while others took a bottom-up approach. Their model achieved the best results in English generative parsing as well as in single sentence language modeling.

Numerous other studies have investigated various linguistics models and mechanisms. Charniak (2016) viewed parsing as a language modeling issue, and recommended the use of LSTM for parsing; Fried, Stern, & Klein, (2017) tested such linguistic models built upon deep learning applications in order to determine the source of the power of these models. Similarly, Dozat and Manning (2018) analyzed the graph-based approach and found self-attentive networks suitable to parse a natural language. Duong et al. (2018) were innovative enough to use a Transformer architecture, as a possible solution to problems in semantic parsing. Last, but not the least, Tan, Wang, Xie, Chen, and Shi, (2018) suggested a self-attention model for semantic role labeling, a kind of semantic parsing and experimented with hyper-parameters for the self-attention mechanism

iv. Semantics

Critical studies on semantics in the NLP context can be classified into two domains: first studies on comparison of semantic similarity of two texts; second, studies that have examined the use of neural language modeling to understand the meaning of words of a language.

In the first section of semantic comparison, the approach adopted is to test the efficacy of computing semantics mechanisms beyond human efforts; that is, to assess the difference made by humans and a computer program to extract the meaning of two similar phrases or sentences. Hu, Lu, Li and Chen (2014) attempted a semantic comparison with two CNN models: In the first model, each CNN shared the weights equally to evaluate given two sentences while in the second, connections were placed between two sentences, and made use of top-level feature maps in the final stage of the CNNs. The results outperformed a number of previously existing models (Hu et al 2014; Socher, Huang, Pennin, Manning, & Ng, 2011; Kalchbrenner, Grefenstette, & Blunsom, 2014). The second section shows how neural language models captured the meaning of words in vectors. Models prominently used were those of Le and Mikolov (2014), which dealt with paragraphs or larger bodies of text; or of Kalchbrenner et al. (2014) which represents sentences using a dynamic convolutional neural network (DCNN), represented by filters. A study was conducted by Poliak, Belinkov, Glass, & Van, (2018) which trained AI enabled encoders on four different language pairs: English and Arabic, English and Spanish, English and Chinese, and English and German and found that each pair required distinct decoding classifiers. The study concluded that NLP models fail to capture paraphrased information as well as semantic inferences e.g. resolving gender, plurality, etc.

A concurrent work (Poliak, Naradowsky, Haldar, Rudinger, & Van, 2018) had also analyzed similar datasets to draw natural language inferences with similar results. And so were the findings of Herzig and Berant (2017) who found that semantic parsers on a single domain are less effective than when used across many domains. The reasons assigned in this situation are that when a single encoder and single decoder are used, it requires the network itself to determine the domain of the input. Similar conclusions are drawn in Brunner, G., Wang, Wattenhofer, & Weigelt. (2018) which create multi-domain LSTM based encoder-decoder networks and analyze the resulting embedded vectors. It was found that a single encoder could work with four different decoders. When the single encoder accepts English sentences as inputs the first decoder replicates attempting to reproduce the original English input. The second and third decoders attempt to translate the text into German or French. Finally, the fourth decoder serves as a POS tagger. This study proved that logical arithmetic mechanisms of simple AI application can be performed on word embeddings as well as sentence embeddings.

Theoretical Framework

A conceptual framework was designed for this study to analyze whether there could be a relationship between NLP platforms such as Morphology, Parsing and Semantics and AI applications for creating linguistic models and to investigate the direct and indirect impact of the relationship between NLP platforms and such AI applications like machine learning and deep learning. The purpose was to assess how their relationship would result in the creation of language models that could be used for different purposes such as translations, sentiment analysis, and chatbots. Figure 1 illustrates the conceptual framework.

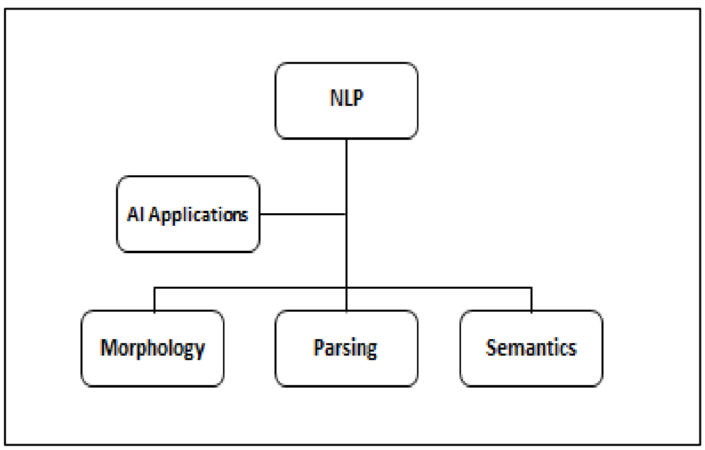

Figure 1. Conceptual Framework

Results and Discussions

The findings and results of this study reveal that there are two kinds of intelligence, "verbal intelligence" and "mathematical intelligence." The former is expressed as human learning output, whereas the latter is machine-learning output. When any English words or sentences are used as inputs in any NLP method, it analyzes the big data freely available on the web. For example, such inputs are associated with syntactic elements, e.g., nouns, verbs, and clauses, or to their semantics, e.g., the individuals, groups in a given domain. The meaning of such inputs varies and is consistent with the domain it belongs to e.g. education, politics, research, government, etc. NLP and its embedded Web technologies can extract meaning for such inputs and represent ontology as linked corpora Data.

The results of this study also suggest that AI applications like machine learning can be used to support many NLP tasks as those applications would utilize the structured data with trained and tested sets through learning algorithms and prediction classifiers and indicators. Such machine learning mechanisms have proved useful prediction, universal search and information retrieval, compliance checking and decision support and also for a better presentation of information. Figure 2 illustrates these machine learning tasks as AI applications. This is an execution of adoption of transfer learning enabled models as AI operation as seen in Figure 2 where datasets can be transferred through algorithms to perform different NLP functions and culminate into a predictable outcome.

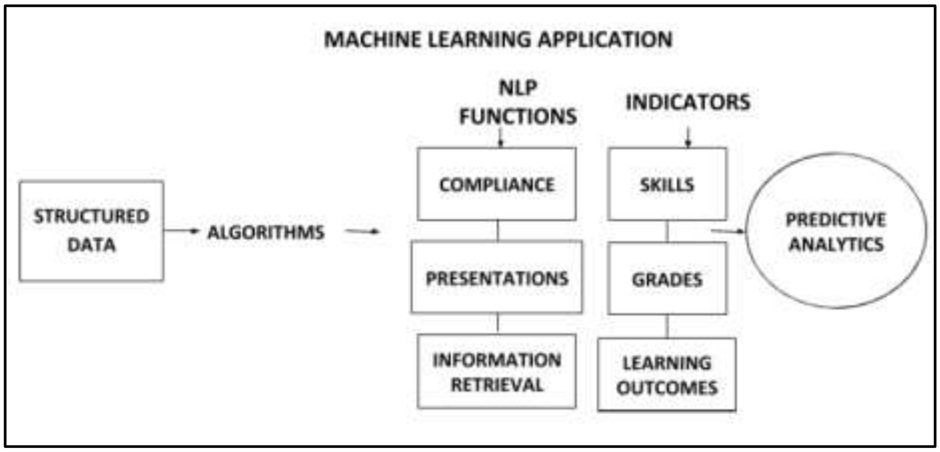

Figure 2. Machine learning approach

Similarly, the adaptation of Deep Learning as one of the AI applications found useful in extracting analytics inferences and enhancing NLP approaches can be applied to EFL texts to address classification, knowledge representation, argument mining, information extraction, information retrieval, ontology population, and multilingualism in specific documents. Figure 3 illustrates how deep learning applications extracts text based or image based data from the unstructured corpora into comprehension and learning.

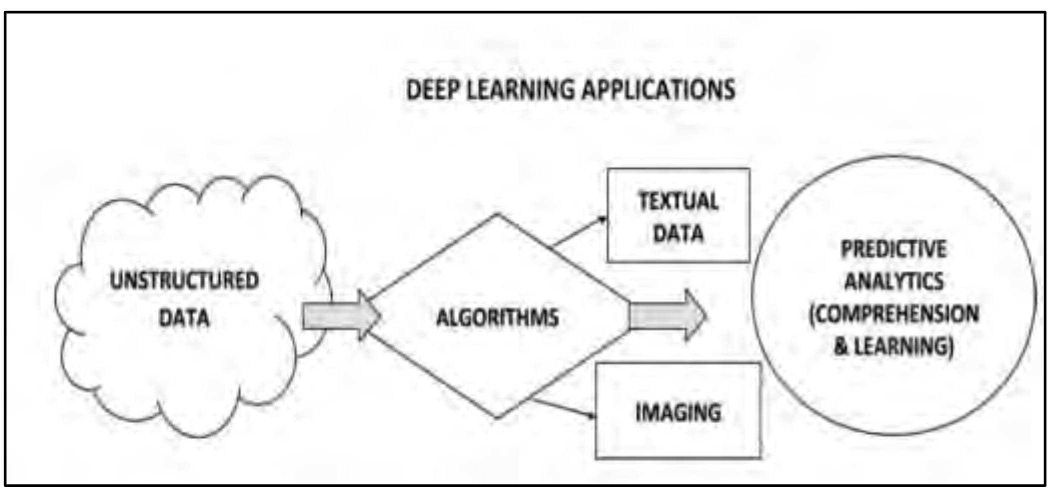

Figure 3. Deep learning approach

Both the AI applications have thus helped build up multi-purpose models in the NLP domain as revealed from the above two models. These models relate to machine translation, question answering systems, chatbots, sentiment analysis and other core issues of language modeling. With the use of AI applications, Google's Transformer architecture too adopted recurrent neural networks (RNN) for language tasks including machine translation and question answering systems and outperformed both RNNs and CNNs (convolutional neural networks). The use of AI applications also reduced the requirement of computational resources for training models due to the use of the self-attention mechanism.

BERT (**B**idirectional **E**ncoder **R**epresentations from Transformers) is also another framework modeled upon AI applications and designed to do multi-task learning and perform different NLP tasks simultaneously. BERT is the first unsupervised, deeply bidirectional system for pre-training NLP models and uses only a plain text corpus The bi-directionality of this framework (the ability to use from both sides; left and right a word or a sentence) helps any linguistic model to gain a much better understanding of the context in which a word or sentence is used, particularly when the Semitic languages are involved.

Conclusion

The study revealed several new avenues that can be made possible through AI applications, namely the use of high- and low-level features in large text corpora (Poliak, Naradowsky et al. 2018)predictive analytics to predict the next word or character in a sequence through applications like Bidirectional Encoder Representations from Transformers (BERT). Additionally, these AI-enabled learning models are also capable of doing multi-task learning, that is, they can perform different NLP applications simultaneously. A few of these applications include building machine translation corpora, statistical parsing, and part-of-speech tagging, to name only a few. Findings of this study have both theoretical and practical implications. Theoretically, this study is a contribution to examine the relationship between NLP platforms and AI applications from EFL perspective. Practical implications can be derived by examining the trends of syntactic and semantic variables when AI applications are utilized. This study faced a few limitations: first, there exist no significant studies on the relationship between NLP platforms and AI applications. Secondly, identifying a sample in a given population was a big challenge as there is a lack of the usage of advanced technologies to sustain AI applications. Moreover, no studies have found to measure the impact of learning models that are powered by NLP applications in the EFL context, mainly to understand their application in the Arabic language. Future studies have a more enormous scope to investigate the use of AI applications in these domains.

About the Author:

Muhammed Salim Keezhatta is an Assistant Professor of English at Prince Sattam Bin Abdulaziz University, Saudi Arabia. He has 12- year experience in teaching English. His academic interests include Applied Linguistics, Theoretical linguistics, Neuro-linguistics, and ELT.
https://orcid.org/0000-0002-4376-3716

References

Andor, D., Alberti, C., Weiss, D., Severyn, A., Presta, A., Ganchev, K., Petrov, S. & Collins, M. (2016). Globally normalized transition-based neural networks. *arXiv preprint arXiv:1603.06042*.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Belinkov, Y., Durrani, N., Dalvi, F., Sajjad, H., & Glass, J. (2017). What do neural machine translation models learn about morphology?. arXiv preprint arXiv:1704.03471.

Bengio, Y., Goodfellow, I., & Courville, A. (2017). *Deep learning* (Vol. 1). MIT press.

Brunner, G., Wang, Y., Wattenhofer, R., & Weigelt, M. (2018). Natural Language Multitasking: Analyzing and Improving Syntactic Saliency of Hidden Representations. *arXiv preprint arXiv:1801.06024*.

Cettolo, M. (2016). An Arabic-Hebrew parallel corpus of TED talks. *arXiv preprint arXiv:1610.00572*.

Cettolo, M., Girardi, C., & Federico, M. (2012). Wit3: Web inventory of transcribed and translated talks. In *Conference of European Association for Machine Translation* (pp. 261-268).

Chen, D., & Manning, C. (2014, October). A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 740-750).

Chi, Z., & Geman, S. (1998). Estimation of probabilistic context-free grammars. *Computational linguistics*, *24*(2), 299-305.

Charniak, E. (2016, November). Parsing as language modeling. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 2331-2336).

Ciresan, D. C., Meier, U., Masci, J., Gambardella, L. M., & Schmidhuber, J. (2011, June). Flexible, high performance convolutional neural networks for image classification. In *Twenty-Second International Joint Conference on Artificial Intelligence*. *22*(1). 1237.

Creutz, M., & Lagus, K. (2007). Unsupervised models for morpheme segmentation and morphology learning. *ACM Transactions on Speech and Language Processing (TSLP)*, *4*(1), 3.

Dehouck, M., & Denis, P. (2018, October). A Framework for Understanding the Role of Morphology in Universal Dependency Parsing.

Dozat, T., & Manning, C. D. (2018). Simpler but more accurate semantic dependency parsing. *arXiv preprint arXiv:1807.01396*.

Duong, L., Afshar, H., Estival, D., Pink, G., Cohen, P., & Johnson, M. (2018, July). Active learning for deep semantic parsing. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 43-48).

Dyer, C., Kuncoro, A., Ballesteros, M., & Smith, N. A. (2016). Recurrent neural network grammars. *arXiv preprint arXiv:1602.07776*.

Fried, D., Stern, M., & Klein, D. (2017). Improving neural parsing by disentangling model combination and reranking effects. *arXiv preprint arXiv:1707.03058*.

Fujisaki, T., Jelinek, F., Cocke, J., Black, E., & Nishino, T. (1991). A probabilistic parsing method for sentence disambiguation. In *Current issues in parsing technology* (pp. 139-152). Springer, Boston, MA.

Fukushima, K., & Miyake, S. (1982). Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position. *Pattern recognition*, *15*(6), 455-469.

Goldberg, Y. (2017). Neural network methods for natural language processing. *Synthesis Lectures on Human Language Technologies*, *10*(1), 1-309.

Hangyo, M., Kawahara, D., & Kurohashi, S. (2012, November). Building a diverse document leads corpus annotated with semantic relations. In *Proceedings of the 26th Pacific Asia Conference on Language, Information, and Computation* (pp. 535-544).

Herzig, J., & Berant, J. (2017). Neural semantic parsing over multiple knowledge-bases. *arXiv preprint arXiv:1702.01569*.

Hu, B., Lu, Z., Li, H., & Chen, Q. (2014). Convolutional neural network architectures for matching natural language sentences. In *Advances in neural information processing systems* (pp. 2042-2050).

Huang, E. H., Socher, R., Manning, C. D., & Ng, A. Y. (2012, July). Improving word representations via global context and multiple word prototypes. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1* (pp. 873-882). Association for Computational Linguistics.

Jelinek, F., Lafferty, J. D., & Mercer, R. L. (1992). Basic methods of probabilistic context free grammars. In *Speech Recognition and Understanding* (pp. 345-360). Springer, Berlin, Heidelberg.

Jones, K. S. (1994). Natural language processing: a historical review. In *Current issues in computational linguistics: in honour of Don Walker* (pp. 3-16). Springer, Dordrecht.

Jurafsky, D. (2000). *Speech & language processing*. Pearson Education India.

Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A convolutional neural network for modelling sentences. *arXiv preprint arXiv:1404.2188*.

Kawahara, D., & Kurohashi, S. (2006, May). Case Frame Compilation from the Web using High-Performance Computing. In *LREC* (pp. 1344-1347).

Kawahara, D., Kurohashi, S., & Hasida, K. (2002, May). Construction of a Japanese Relevance-tagged Corpus. In *LREC*.

Le, Q., & Mikolov, T. (2014, January). Distributed representations of sentences and documents. In *International conference on machine learning* (pp. 1188-1196).

Le, P., & Zuidema, W. (2014, October). The inside-outside recursive neural network model for dependency parsing. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 729-739).

LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, *3361*(10), 1995.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436.

Luong, T., Socher, R., & Manning, C. (2013, August). Better word representations with recursive neural networks for morphology. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning* (pp. 104-113).

Miller, G. A., & Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and cognitive processes*, *6*(1), 1-28.

Morita, H., Kawahara, D., & Kurohashi, S. (2015, September). Morphological analysis for unsegmented languages using recurrent neural network language model. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 2292-2297).

Nivre, J. (2015, April). Towards a universal grammar for natural language processing. In *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 3-16). Springer, Cham.

Poliak, A., Belinkov, Y., Glass, J., & Van Durme, B. (2018). On the evaluation of semantic phenomena in neural machine translation using natural language inference. *arXiv preprint arXiv:1804.09779*.

Poliak, A., Naradowsky, J., Haldar, A., Rudinger, R., & Van Durme, B. (2018). Hypothesis only baselines in natural language inference. *arXiv preprint arXiv:1805.01042*.

Rubenstein, H., & Goodenough, J. B. (1965). Contextual correlates of synonymy. *Communications of the ACM*, *8*(10), 627-633.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85-117.

Socher, R., Huang, E. H., Pennin, J., Manning, C. D., & Ng, A. Y. (2011). Dynamic pooling and unfolding recursive autoencoders for paraphrase detection. In *Advances in neural information processing systems* (pp. 801-809).

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., & Potts, C. (2013, October). 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).

Socher, R., Bauer, J., Manning, C. D., & Ng, A. Y. (2013, August). Parsing with compositional vector grammars. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 455-465).

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Advances in neural information processing systems* (pp. 3104-3112).

Tan, Z., Wang, M., Xie, J., Chen, Y., & Shi, X. (2018, April). Deep semantic role labeling with self-attention. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Vinyals, O., Kaiser, Ł., Koo, T., Petrov, S., Sutskever, I., & Hinton, G. (2015). Grammar as a foreign language. In *Advances in neural information processing systems* (pp. 2773-2781).

Weiss, D., Alberti, C., Collins, M., & Petrov, S. (2015). Structured training for neural network transition-based parsing. *arXiv preprint arXiv:1506.06158*.

Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing. *ieee Computational intelligenCe magazine*, *13*(3), 55-75.

Zeman, D., Hajič, J., Popel, M., Potthast, M., Straka, M., Ginter, F., Nivre, J., & Petrov, S. (2018, October). CoNLL 2018 shared task: multilingual parsing from raw text to universal dependencies. In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies* (pp. 1-21).