



Portuguese POS Tagging Using BLSTM Without Handcrafted Features

Rômulo César Costa de Sousa and Hélio Lopes^(✉)

Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ, Brazil
{rsousa,lopes}@inf.puc-rio.br

Abstract. Training state-of-the-art Part-of-speech (POS) taggers traditionally requires many handcraft features and external data. In this paper, we propose a neural network architecture for POS tagging task for both contemporary and historical Portuguese texts. The proposed architecture does not use the two traditional requirements cited above. It uses word embeddings and character embeddings representations combined with a BLSTM layer. We apply the architecture on three Portuguese corpora and obtaining state-of-the-art accuracy of 97.87% on the Mac-Morpho corpus, 97.62% accuracy on the revised Mac-Morpho and 97.36% on Tycho Brahe. We also improve the tagging accuracy for Out of Vocabulary (OOV) words in the Mac-Morpho corpus and in the revised Mac-Morpho.

Keywords: Part-of-speech tagging · Deep learning · Word embeddings

1 Introduction

Part-of-speech (POS) tagging is a process of labeling each word in a sentence with a morphosyntactic class (verb, noun, adjective and etc). POS tagging is considered a hard task to perform due to the fact that some words could have more than one class, depending on the context that it is used. POS tagging is a fundamental part of the linguistic pipeline [9], most natural language processing (NLP) applications demand, at some step, part-of-speech information [8]. For example, its use can be found in sentiment analysis [4], in machine translation [15] and question answering [27].

Several works have been done to solve this task and they push the state-of-art accuracy to more than 97%. But still, have room for improvement since any small gain can generate an impact on other NLP tasks. Recently, neural network approaches have become very used to solve NLP problems [12] and word embeddings, also have been applied with a great success [16].

In [18], the authors build an LSTM neural network that extracted character embeddings information for using in POS tagging classification. They tested the model on four different languages. In [24], the authors explore the problem of using a deep neural network that uses a convolutional layer to learn character level representation of words, and achieve great results on three Portuguese corpora.

In this work, we propose the use of a neural network architecture that is effective to solve the Portuguese POS tagging task. More precisely, our proposal combines BLSTM with pre-trained word embeddings and character embeddings. We evaluate our model on three Portuguese corpora: the original Mac-Morpho corpus [2], the revised Mac-Morpho [7] and Tycho Brahe [25] corpus. We outperforms previous state-of-the-art results on these corpora, obtaining 97.87% accuracy on the Mac-Morpho corpus, 97.62% accuracy on the revised Mac-Morpho and 97.36% on Tycho Brahe.

The structure of the paper is as follows. Section 2 discusses related work. Section 3 introduces the neural network architecture proposal. Section 4 presents the experiments and the results. Finally, Sect. 5 gives a conclusion and points out our next steps.

2 Related Work

There has been many efforts to build Portuguese POS tagging. In [24], authors built a system without any handcrafted features, using a neural network. They employ a convolutional layer that allows effective feature extraction from words of any size. They combine this representation along with word embeddings to perform POS tagging. They evaluate on three Portuguese corpora: Mac-Morpho-v1, Mac-Morpho-v2 and Tycho Brahe corpus.

In [5], authors used a Large Margin Structured Perceptron algorithm to solve POS tagging. They used a small set (four) of handcraft features to improve their performance. They empirically evaluate their system on the two versions of the Portuguese Mac-Morpho corpus.

Fonseca et al. [6] used a multilayer perceptron neural network for training a POS tagger. This neural network receives word embedding information and handcraft features such as the presence of capital letters and word endings. The authors report state-of-the-art performance on Portuguese corpora. With 97.57% overall accuracy on the Mac-Morpho-v1 corpus, 97.48% on Mac-Morpho-v2, and 97.33% on Mac-Morpho-v3 presented in their work.

There are also several other neural networks proposed for POS tagging in other languages. Wang et al. [26] proposed to use a BLSTM neural network with word embedding for POS tagging with an English corpus. In [23], authors used a BLSTM neural network for POS tagging with an auxiliary loss function that accounts for rare words. They evaluated the neural network across 22 languages.

In [16], Ma and Hovy propose a neural network to solve POS tagging and named entity recognition (NER). They constructed a model by feeding character embeddings and word embeddings to a BLSTM on top of a CRF (Conditional Random Fields) layer. They used a Convolution Neural Network (CNN) for extracting character embeddings representations of words. And evaluate the model on the English Penn Treebank WSJ corpus. There are four main differences between their model and what we are to propose here. We did not employ CNNs to extracting character embeddings information, instead we use a BLSTM

layer. We did not use CRF layer in the end of the model. We use word embeddings information provided by Wang2vec instead of GloVe. And they apply their method for English language, while we are focused on Portuguese.

The work of [20] is what most resembles our neural network proposal. The authors used a BLSTM layer fed by character and word embeddings information to perform POS Tagging. They used a BLSTM for extracting character embeddings representations of words. Our work mainly differs from the model proposed since we use Wang2vec instead of FastText and they apply their method to Italian, and we apply it to Portuguese.

3 Neural Network Architecture Proposal

In this section, we describe our proposal for the neural network architecture, which consists in a BLSTM architecture combined with word embeddings and character embeddings. Figure 1 illustrates our proposal.

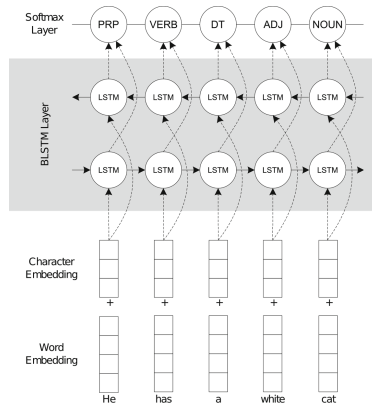


Fig. 1. Neural network architecture for POS tagging

Word-Level Representation. Word embeddings have become commonly used in modern NLP systems [10], this technique represents each word as a vector with real numbers in an d -dimensional space. This allows words with similar meaning to have a similar representation. Word embeddings capture both the semantic and syntactic information of words [13]. This representation have been applied with a great success [16] in many NLP tasks.

In this work, we used pre-trained word embedding models provided by [13]. They collected a large corpus from several sources in order to obtain a multi-genre corpus, representative of the Portuguese language. Seventeen different corpora were used, totaling 1,395,926,282 tokens. They training the word embedding models in algorithms such as Word2vec [21], FastText [3], Wang2vec [14] and

Glove [22]. We tested which algorithm for word embedding is better fit in our neural network.

Character-Level Representation. Morphological information of words can be helpful in POS tagging classification. Suffixes could indicate a word class. For example, the suffix like “ly” in “quietly” indicate an adverb, and a capital letter could suggest that this word is a noun. But, handcrafted this kind of features is costly to develop [17], and make this hard to adapt to other domains or languages. Previous studies have shown that neural networks are a powerful way to extract automatically morphological information. In [24], the authors reached state-of-art results in Portuguese for POS tagging using a convolutional layer to extract this information. And [20] exhibit positive results in Italian for POS tagging by using BLSTM for extract morphological information.

For this subtask, we used a model proposed in [19] that applied a BLSTM layer to produce character embeddings representation. Differently from word embeddings, which are able to capture syntactic and semantic information, character embeddings can capture intra-word morphological and shape information [20]. We define a character’s vocabulary that contains all uppercase and lowercase letter as well as numbers and punctuation present on data. Given a word w as an input, decomposed in m characters $\{c_1, c_2, \dots, c_m\}$, where m is the length of w . Each c_i is encoded as a one hot vector, with one on the index of c_i in vocabulary. The representation of the word w is obtained by combining the forward and backward states from the BLSTM layer.

Network Architecture. Long short-term memory (LSTM) proposed by [11], are a variant of RNNs. This kind of neural network is well known by their power of capturing long-term dependencies, and have been widely used for sequence labeling tasks. LSTM maintain previous information using memory cells. For example, it takes a sequence of vectors, (x_1, x_2, \dots, x_n) as input and produces another sequence (h_1, h_2, \dots, h_n) as output [1]. The LSTM captures just previous information and has no knowledge of what comes next, but for many sequence labeling tasks, it is helpful to have access to both past and future contexts [16]. Bi-directional LSTM (BLSTM) is a solution for filling this gap. BLSTM take information in two separate hidden states one to the past and other to the future and then concatenated the two separate hidden as one final state.

Given a decomposed sentence with n words, $w_1, w_2, w_3, \dots, w_n$, and n tags, $t_1, t_2, t_3, \dots, t_n$, we use the BLSTM main layer to predict the tag probability distribution of each word. As shown in Fig. 1, to perform this prediction, we firstly represent each word as a vector, which is the result of the concatenation of the word embeddings representation and the character embeddings representation. In this way, we can capture semantic and syntactic information (word embeddings) and morphological information (character embeddings) of the words.

4 Experiments

In this section, we provide details about training the neural network and discuss our results. We implemented our model using Keras¹ on top of TensorFlow². In the output layer, we use a softmax activation function.

Corpora. We evaluate the model on three different Portuguese corpora: the original Mac-Morpho (v1) corpus [2], a revised version of Mac-Morpho (v2) [7] and Tycho Brahe corpus [25]. The Mac-Morpho is a large manually POS-tagged corpus Portuguese, collected from Brazilian newspaper articles [2]. The original version has 53,374 sentences and 41 morphosyntactic class, as shown in Table 1, and the revised version has 49,900 sentences and 30 morphosyntactic class. The Tycho Brahe Corpus is composed of historical Portuguese with texts written in Portuguese by authors born between 1380 and 1881. We use the same train/development/train split as [24] and [6] in order to directly compare results.

Table 1. Corpus splits

Corpus		Train	Development	Test	Total	Tagset
Mac-Morpho v1	Sentences	42,022	2,211	9,141	53,374	41
	Tokens	957,439	50,232	213,794	1,221,465	
Mac-Morpho v2	Sentences	42,742	2,249	4,999	49,990	30
	Tokens	807,796	43,141	94,991	945,928	
Tycho Brahe	Sentences	29,162	1,534	10,233	40,929	265
	Tokens	734,889	40,673	259,913	1,035,475	

Hyperparameters. We use the development sets to tune the neural network hyperparameters. First, we examine the size of the hidden layer on the main BLSTM. This feature has a limited impact on results. We also analyze word embedding dimensions, character embedding dimensions and dropout rate on the main BLSTM layer. We used the same set of hyperparameters shown on Table 2 for all corpus and experiments.

Table 2. Hyper-parameters for all experiments

Parameter	Value
Hidden Layer (Main BLSTM)	512
Word embedding dimensions	100
Char. embedding dimensions	50
Dropout (Main BLSTM)	0.6

¹ <https://keras.io/>.

² <https://www.tensorflow.org/>.

Results. Firstly, we test the performance of our model on four different methods to pre-train word embedding. We used as metrics overall accuracy (ALL) and Out of Vocabulary (OOV) accuracy. We run experiments with Word2vec, Fasttext, Wang2vec and Glove. As presented in Table 3, all word embeddings generate higher accuracy. However, Wang2vec produce the highest overall accuracy in our model on all corpora. Wang2vec’s good performance may be explained by its focus in capture syntactic information [14], this is really useful for POS tagging classification.

Table 3. Performance of our model with different word embeddings on the development set

Corpus	Embedding	Accuracy (%)
Mac-Morpho v1	Word2vec	97.84
	Fasttext	97.78
	Wang2vec	97.97
	Glove	97.70
Mac-Morpho v2	Word2vec	97.56
	Fasttext	97.44
	Wang2vec	97.67
	Glove	97.53
Tycho Brahe	Word2vec	96.93
	Fasttext	96.94
	Wang2vec	97.17
	Glove	97.11

In order to measure the impact of word embeddings information and character embeddings, we compare the performance of our model with two baseline systems. Table 4 shown the accuracy and OOV accuracy for the three systems in the three corpora evaluated. The first baseline (Perceptron-handcrafted-features) uses a multilayer perceptron with handcraft features. We utilize as features the presence of capital letters, word suffixes, word prefixes, previous three tokens, and the next three tokens. This model does not use any word embeddings or character embeddings information. The second baseline (BLSTM-WE) was built similar to our main model, but without any character embeddings information, just word embeddings vectors. And the system BLSTM-WE-CE is our model describe in Sect. 3. Both BLSTM-WE and BLSTM-WE-CE use Wang2vec word embeddings and the same hyperparameters as shown in Table 2.

According to the results shown Table 4, BLSTM-WE performed better than Perceptron-handcrafted-features on Mac-Morpho v1 and Mac-Morpho v2. Since Tycho Brahe is formed by history of Portuguese text, many words are not present in the pre-trained word embeddings, and this corpus could not benefit like the

other by using BLSTM-WE. BLSTM-WE-CE system outperforms the BLSTM-WE on the three corpora, especially on OOV accuracy. This result support what already has been demonstrated by [24] and [16], that character embeddings are important for linguistic sequence labeling tasks like POS tagging.

Table 4. Performance of our model and two baseline on the test set

Corpus	System	All (%)	OOV (%)
Mac-Morpho v1	Perceptron-handcrafted features	96.31	88.45
	BLSTM-WE	96.81	90.33
	BLSTM-WE-CE	97.87	95.44
Mac-Morpho v2	Perceptron-handcrafted features	96.28	88.56
	BLSTM-WE	96.58	90.57
	BLSTM-WE-CE	97.62	94.84
Tycho Brahe	Perceptron-handcrafted features	96.04	75.72
	BLSTM-WE	95.88	74.01
	BLSTM-WE-CE	97.36	85.74

Comparisons. We compare our results with two top performance systems. In Table 5, we compare overall accuracy and OOV accuracy on the three corpora. Our system outperforms the previously best systems by improving in 0.23% overall accuracy for the Mac-Morpho v1, 0.31% for Mac-Morpho v2 and 0.19% for Tycho Brahe corpus. We also improve OOV accuracy in 3.18% for the Mac-Morpho v1 and 1.50% in Mac-Morpho v2. These results define a new state-of-the-art of this three corpora and demonstrate effectiveness of using BLSTM in Portuguese POS tagging.

Table 5. Comparison with Portuguese POS taggers

Corpus	Embedding	All (%)	OOV (%)
Mac-Morpho v1	This work	97.87	95.44
	[6]	97.57	93.38
	[24]	97.47	92.49
Mac-Morpho v2	This work	97.62	94.84
	[6]	97.48	94.34
	[24]	97.31	93.43
Tycho Brahe	This work	97.36	85.74
	[6]	96.91	84.14
	[24]	97.17	86.58

Table 6. F_1 per tags in Mac-Morpho v1

Class	Meaning	F_1 (%)
CUR	Currency	100.0
ART	Article	99.0
PREP	Preposition	99.0
V	Verb	99.0
KC	Coord. conjunction	98.0
N	Noun	98.0
PROADJ	Adjective pronoun	98.0
PROPESS	Personal pronoun	98.0
NPROP	Proper noun	97.0
NUM	Number	97.0
PCP	Participle	97.0
VAUX	Auxiliary verb	96.0
ADJ	Adjective	95.0
PRO-KS-REL	Sub. c. r. pronoun	94.0
ADV	Adverb	93.0
KS	Sub. conjunction	91.0
PDEN	Denotative word	91.0
PROSUB	Nominal pronoun	89.0
ADV-KS-REL	Sub. rel. adverb	86.0
PRO-KS	Sub. con. pronoun	75.0
ADV-KS	Sub. con. adverb	55.0
IN	Interjection	52.0

Table 7. F_1 per tags in Mac-Morpho v2

Class	Meaning	F_1 (%)
CUR	Currency	100.0
PREP+PRO-KS-REL	Prep. sub. c. r. pronoun	100.0
PU	Punctuation	100.0
ART	Article	99.0
PREP+ART	Preposition + article	99.0
PREP+PROADJ	Prep. + adjective pronoun	99.0
PREP+PROPESS	Prep. + personal pronouns	99.0
PROPESS	Personal pronouns	99.0
V	Verb	99.0
KC	Coord. conjunction	98.0
N	Noun	98.0
PREP	Preposition	98.0
NPROP	Proper noun	97.0
NUM	Number	97.0
PCP	Participle	97.0
PROADJ	Adjective pronoun	97.0
VAUX	Auxiliary verb	96.0
ADJ	Adjective	95.0
ADV	Adverb	93.0
PREP+ADV	Preposition + adverb	92.0
PRO-KS-REL	Sub. c. r. pronoun	92.0
KS	Sub. conjunction	91.0
PROSUB	Nominal pronoun	91.0
PDEN	Denotative word	90.0
ADV-KS-REL	Sub. rel. adverb	86.0
PREP+PROSUB		83.0
PRO-KS	Sub. con. pronoun	76.0
IN	Interjection	73.0
ADV-KS	Sub. con. adverb	71.0
PREP+PRO-KS	Prep. + s. c. pronoun	00.0

Performance Per Tags. We used F1-score $F_1 = 2 * Precision * Recall / Precision + Recall$ to evaluate our best model (BLSTM-WE-CE). We analyzed the tag-wise performance obtained by BLSTM-WE-CE in the test set for the Mac-Morpho v1 and the Mac-Morpho v2.

Table 6 shows F_1 scores per tag on the Mac-Morpho v1, only non punctuation tags were included. Most of the tags reached of more than 90% of F_1 . The worst results are for IN (Interjection) with 52% and for ADV-KS (Subordinating connective adverb) with 55%. These two tags have a low representation on Mac-Morpho v1 corpus, IN represents just 0.034% of the tokens and ADV-KS just 0.032%. Due to their poor distributions, the model could not learn to classify these classes properly.

In Table 7 we have the F_1 scores per tag on the Mac-Morpho v2. Most of the tags reached more than 90% of F_1 . The model completely failed to predict for PREP+PRO-KS (Preposition + subordinating connective pronoun), the system

is not learning anything due to their poor distributions. PREP+PRO-KS have the lowest representation on the dataset just 0.033%.

5 Conclusions

In this study, We have empirically investigated a new approach that does not require handcrafted features to deal with Portuguese POS tagging task. We show that BLSTM with word embeddings and character embeddings gives superior performance for Portuguese POS tagging reaching state-of-art in overall accuracy in Mac-Morpho v1, Mac-Morpho v2, and Tycho Brahe corpus. As future work, we intend to extend this architecture to solve another natural language processing tasks in Portuguese, such as named entity recognition.

References

1. Alam, F., Chowdhury, S.A., Noori, S.R.H.: Bidirectional LSTMs-CRFs networks for bangla POS tagging. In: 2016 19th International Conference on Computer and Information Technology (ICIT), pp. 377–382. IEEE (2016)
2. Aluísio, S., Pelizzoni, J., Marchi, A.R., de Oliveira, L., Manenti, R., Marquiasável, V.: An account of the challenge of tagging a reference corpus for Brazilian Portuguese. In: Mamede, N.J., Trancoso, I., Baptista, J., das Graças Volpe Nunes, M. (eds.) PROPOR 2003. LNCS (LNAI), vol. 2721, pp. 110–117. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-45011-4_17
3. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. *Trans. Assoc. Comput. Linguist.* **5**, 135–146 (2017). <http://aclweb.org/anthology/Q17-1010>
4. Das, O., Balabantaray, R.C.: Sentiment analysis of movie reviews using POS tags and term frequencies. *Int. J. Comput. Appl.* **96**(25), 36–41 (2014)
5. Fernandes, E.R., Rodrigues, I.M., Milidiú, R.L.: Portuguese part-of-speech tagging with large margin structure learning. In: 2014 Brazilian Conference on Intelligent Systems (BRACIS), pp. 25–30. IEEE (2014)
6. Fonseca, E.R., Rosa, J.L.G., Aluísio, S.M.: Evaluating word embeddings and a revised corpus for part-of-speech tagging in portuguese. *J. Braz. Comput. Soc.* **21**(1), 2 (2015)
7. Fonseca, E.R., Rosa, J.L.G.: Mac-Morpho revisited: towards robust part-of-speech tagging. In: Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology (2013)
8. Giménez, J., Marquez, L.: SVMtool: a general POS tagger generator based on support vector machines. In: Proceedings of the 4th International Conference on Language Resources and Evaluation. Citeseer (2004)
9. Gimpel, K., et al.: Part-of-speech tagging for twitter: annotation, features, and experiments. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2, pp. 42–47. Association for Computational Linguistics (2011)
10. Hartmann, N., Fonseca, E., Shulby, C., Treviso, M., Silva, J., Aluísio, S.: Portuguese word embeddings: evaluating on word analogies and natural language tasks. In: Proceedings of the 11th Brazilian Symposium in Information and Human Language Technology, pp. 122–131 (2017)

11. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997)
12. Jung, S., Lee, C., Hwang, H.: End-to-end Korean part-of-speech tagging using copying mechanism. *ACM Trans. Asian Low-Resour. Lang. Inf. Process. (TALLIP)* **17**(3), 19 (2018)
13. Lai, S., Liu, K., He, S., Zhao, J.: How to generate a good word embedding. *IEEE Intell. Syst.* **31**(6), 5–14 (2016)
14. Ling, W., Dyer, C., Black, A.W., Trancoso, I.: Two/too simple adaptations of word2vec for syntax problems. In: *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1299–1304. Association for Computational Linguistics (2015). <https://doi.org/10.3115/v1/N15-1142>. <http://aclweb.org/anthology/N15-1142>
15. Ma, J., Liu, H., Huang, D., Sheng, W.: An English part-of-speech tagger for machine translation in business domain. In: *2011 7th International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE)*, pp. 183–189. IEEE (2011)
16. Ma, X., Hovy, E.: End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, vol. 1, pp. 1064–1074 (2016)
17. Ma, X., Xia, F.: Unsupervised dependency parsing with transferring distribution via parallel guidance and entropy regularization. In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, vol. 1, pp. 1337–1348 (2014)
18. Makazhanov, A., Yessenbayev, Z.: Character-based feature extraction with LSTM networks for POS-tagging task. In: *2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT)*, pp. 1–5. IEEE (2016)
19. Marujo, W.L.T.L.L., Astudillo, R.F.: Finding function in form: Compositional character models for open vocabulary word representation (2015)
20. Marulli, F., Pota, M., Esposito, M.: A comparison of character and word embeddings in bidirectional LSTMs for POS tagging in Italian. In: De Pietro, G., Gallo, L., Howlett, R.J., Jain, L.C., Vlacic, L. (eds.) *KES-IIMSS-18 2018. SIST*, vol. 98, pp. 14–23. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-92231-7_2
21. Mikolov, T., Chen, K., Corrado, G.S., Dean, J.: Efficient estimation of word representations in vector space. *CoRR abs/1301.3781* (2013)
22. Pennington, J., Socher, R., Manning, C.: Glove: global vectors for word representation. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543 (2014)
23. Plank, B., Søgaard, A., Goldberg, Y.: Multilingual part-of-speech tagging with bidirectional long short-term memory models and auxiliary loss. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 412–418 (2016)
24. dos Santos, C.N., Zadrozny, B.: Training state-of-the-art Portuguese POS taggers without handcrafted features. In: Baptista, J., Mamede, N., Candeias, S., Paraboni, I., Pardo, T.A.S., Volpe Nunes, M.G. (eds.) *PROPOR 2014. LNCS (LNAI)*, vol. 8775, pp. 82–93. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-09761-9_8
25. Temponi, C.N., et al.: O corpus anotado do português histórico: um avanço para as pesquisas em lingüística histórica do português. *Revista Virtual de Estudos da Linguagem: REVEL* **2**(3), 1 (2004)

26. Wang, P., Qian, Y., Soong, F.K., He, L., Zhao, H.: Part-of-speech tagging with bidirectional long short-term memory recurrent neural network. arXiv preprint [arXiv:1510.06168](https://arxiv.org/abs/1510.06168) (2015)
27. Wang, W., Auer, J., Parasuraman, R., Zubarev, I., Brandyberry, D., Harper, M.: A question answering system developed as a project in a natural language processing course. In: Proceedings of the 2000 ANLP/NAACL Workshop on Reading Comprehension Tests as Evaluation for Computer-based Language Understanding Systems-Volume 6, pp. 28–35. Association for Computational Linguistics (2000)