Deep Neural Networks for Part-of-Speech Tagging in Under-Resourced Amazigh

Rkia Bani[1]*, Samir Amri[2], Lahbib Zenkouar[1], Zouhair Guennoun[1]

[1] Smart Communication Research Team (SCRT), Mohammadia School of Engineering (EMI), Mohammed V University, Rabat 10090, Morocco
[2] Department of Computer Science and Mathematics, ENSAM School of Engineering, Moulay Ismail University, Meknes 50000, Morocco

Corresponding Author Email: rkiabani@research.emi.ac.ma

ABSTRACT

Part-of-speech (POS) tagging denotes the assignment of appropriate grammatical categories to individual words within a sentence or text, playing a pivotal role in numerous natural language processing (NLP) tasks. While POS tagging in widely-used languages such as English has reached accuracy levels exceeding 97%, less-resourced languages such as Amazigh have seen limited research and therefore, less accuracy in tagging efforts. This paper aims to bridge this gap by exploring the application of deep learning models for Amazigh POS tagging, specifically focusing on various types of recurrent neural networks (RNNs)—gated recurrent networks, long short-term memory (LSTM) networks, and bidirectional LSTM networks. Despite the relatively small dataset of 60k tokens, a stark contrast to the vast corpuses available for languages with extensive resources, the proposed RNN models have demonstrated significant improvements over existing Amazigh POS taggers. Remarkably, all RNN models tested in this study outperformed traditional machine learning taggers, achieving an accuracy rate of 97%, thus presenting a promising avenue for enhanced POS tagging in under-resourced languages. This research underscores the potential of deep learning approaches in contributing to the advancement of linguistic studies in less-documented languages, such as Amazigh.

1. INTRODUCTION

Part-of-speech tagging (abbreviation POS tagging) is the process of assigning to each word of a sentence or text, its corresponding grammatical class. There are basic POS taggers that classify the words into verbs, nouns, adjectives, particles etc. And other POS taggers can give other specifications like gender, number, state etc. POS tagging is an important preprocessing step for other natural language processing (NLP) tasks and improves their performance. Also, the corpus or dataset which is annotated with POS is largely used in subfields of NLP like information retrieval, sentiment analysis, speech recognition etc. There are several approaches of concepting POS taggers that can mainly be divided into two categories: rules-based models and machine learning models. Rule based POS taggers rely on rules elaborated by humans starting from the characteristics of the language and its grammar, it's a fastidious approach and requires more time. In the other side, there are machine learning POS taggers, they take statistic and probabilistic as source to establish models. Both approaches perform satisfactory results in rich languages like English but not in very low resource language like Amazigh.

1.1 Amazigh language background

The Amazigh language, called Berber or commonly Tamazight, belongs to the Afro-Asiatic family languages [1, 2]. It is widespread in North Africa mainly from the Niger in

the Sahara to the Mediterranean Sea in addition of the Canary Isles. The first perspective of enriching Amazigh language is developing an Amazigh writing system and its standardization. This leads to the creation of the Amazigh graphical system called Tifinagh-IRCAM. The Tifinaghe-IRCAM system includes: 27 consonants (ⵜ, ⵀ, ⵠ, ⵌ, ⴻ, ⵁ, ⵏ, ⵂ, ⵀ, ⵛ, ⵔ, ⵅ, ⵋ, ⴻ, ⵯ, ⵔ, ⵅ, ⵛ, ⵇ, ⵉ, ⵥ, ⵏ, ⵅ, ⵎ, ⴽ, ⵀ, ⴼ), two semi-consonants (ⵢ, ⵡ) and 4 vowels (ⵄ, ⵖ, ⵗ, ⵘ).

The syntactic categories [3, 4] of the Amazigh language includes nouns, verbs, and particles. A noun is a grammatical class, obtained from the combination of roots and patterns. It could be in a simple form (ⵄⵎⵔⵗⵛ 'aghrum' the bread), compound form (ⵀⵗⵂⵄⵗⵛ 'buEari' the forest keeper, it is composed of two word 'ⵀⵗ' and 'ⵂⵄⵗⵛ') or derived one (ⵄⵀⵏⵛⵄⵠ 'aslmad' the teacher). The nouns in Amazigh language could be masculine or feminine, and singular, or plural in term of numbers. Nouns are presented in two cases: a free case or a construct case. The construct case is morphological variation that affects the first syllable of noun when certain conditions are met. In free case, the first syllable of the noun doesn't undergo any change.

The second category is Verb which can be divided in two categories: basic verbs and derived verbs. The basic verb is formed of a root and a radical. In the other hand, the formation of derived verbs is obtained from the basic verbs with the prefixation of one of these morphemes: ⵀ 's'/ⵀⵀ 'ss' (to indicate the factitive from), ⵜⵜ 'tt' (to indicate the passive form), ⵛ 'm'/ⵛⵛ 'mm' (to indicate the reciprocal form). The

factive from expresses the idea of doing something to someone or something ('ⵓⵙⵢ', bring out). The passive form is less employed, it expresses that the action is exercised to the subject itself ('ⵜⵜⵢⵙⵓⴰ', to be stolen). The reciprocal form is used to express the simultaneity of an action exercised and undergone by two or more participants ('ⵎⵎⴰⵖⵏ', they quarreled). In conjugation, verbs have four aspects: perfect, negative perfect, aorist and imperfective.

The last syntactic category is particle which plays a function word that can't designate noun or verb. It includes conjunctions, pronouns, aspectual, prepositions, orientation, negative particles, subordinates, and adverbs. Usually, particles don't undergo inflection words except for the demonstrative and possessive pronouns ('ⵡⴰ' ("wa") means this and 'ⵡⵉⴷ' ('win') which means these)

1.2 Constraints of Amazigh POS tagging and our contribution

If the rich resource languages have multiples POS taggers, that isn't the case in very low resource language, especially, Amazigh language. POS tagging Amazigh is a challenging task because of the scarcity of annotated corpora, lack of lexicon and morphological analyzer. Also, there are few research works in POS tagging Amazigh. All these constraints motivate researchers to enrich Amazigh language with multiple NLP tasks starting from POS tagging. Recently, the deep learning approaches seem to be delivering high performance in term of accuracy in different NLP applications and provide features extractions of text by word-embedding.

In the aim of contributing to the development of the Amazigh language processing, we present in this paper a new deep learning approach for Amazigh POS tagging based on recurrent neural networks (RNN). To elaborate this approach, a data preparation step was necessary because the available Amazigh dataset isn't normalized (just a CSV file with two columns of word and tag), so we realized a new dataset in form of tagged sentences. Our work contributes also, for the first time, with the word embedding of Amazigh language. The results obtained in this works reaches 97% in terms of accuracy and outperform the existing Amazigh POS taggers.

1.3 Organization of this paper

The rest of this paper is organized as follows: Sect.2 provides a description of the advances in the POS tagging of different language starting from well-known languages to very low resources languages mainly Amazigh language. Sect.3 presents the proposed approach of using deep learning techniques for Amazigh language. In Sect.4, we detail the material used in the experiment and discuss the results of the proposed model. Finally, in Sect. 5, we conclude our works and presents our perspectives in the future to contribute to the development of the Amazigh language.

2. RELATED WORKS

POS tagging is an essential step for other natural language processing (NLP) tasks like sentiments analysis, translation, etc. First POS tagging technics were heuristics and rule based, but recently the most used approaches are machine learning. Starting from statistical and probabilistic approaches like the hidden Markov model (HMM) [5] and the famous conditional

random fields (CRF) [6] to neural networks like recurrent neural network (RNN) and long-short-term memory (LSTM). In this section we present the application of neural network in natural language processing, precisely POS tagging.

2.1 Neural networks in POS tagging Arabic and Chinese as well studied languages

In the study [7], they developed a POS tagger for modern standard Arabic and Gulf tweets using the bidirectional LSTM and the results were promising compared to those obtained in CRF. They elaborated a dataset of 75000 tokens and used the MADARi interface to extract feature for the CRF model. The results of both models achieve good results compared to baseline, but the Bi-LSTM outperforms all the models reaching 96%. BERT has been used by Saidi et al. [8] to design an Arabic POS tagger trained on corpus said normalized. They combined multiple corpus and conceived a model based on BERT, dense and classification layer. They fine-tuned the model by freezing all the architecture and adding untrained layers of neurons and train the model. The achieved results are 91%. In the study [9], they realized two taggers for Arabic Gulf dialect based on support vector machine (SVM) [10] and Bi-LSTM. The dataset used is a Gulf Arabic tweets of 6800 tokens. The features chosen for the SVM models are clitics, probabilistic and binary and the best results is just 86%. As for the Bi-LSTM model tested, the result reaches 91%.

In Chinese language, Shao et al. [11] present a bidirectional RNN-CRF based model for joint segmentation and POS tagging. They used three different Chinese corpus CTB5, CTB9 and UD Chinese. As for the chosen features, they used multiples representation of Chinese character by concatenating different n-grams in both convolutional and recurrent networks. The results of the experiments, in terms of accuracy, were state-of-art reaching 98% in segmentation and 94% in joined segmentation and tagging. A POS tagging based on recurrent neural network proposed by Qin [12] for Chinese Buddhist language. In the contrary of the modern Chinese, the Buddhist Chinese suffers from the lack of corpus, he used a 40000 token. He implemented three RNN models that all outperform the classical HMM model. Deep graph neural network has been used in the study [13] to develop a model for Chinese dependency parsing realizing both segmentation and POS tagging tasks. This model implements three different layers, where the first is a character representation using embedding and BERT, the second layer is Bi-LSTM/Transformer based to produce features for the joint tasks and finally a decoding layer for each task. They tested in three corpus CBT5, CBT6, CBT9 and obtained 96% of accuracy in POS tagging. Generally, Chinese language is well studied language on term of natural language processing using deep learning techniques, starting from parsing [14, 15] to POS tagging [16-18].

2.2 Literature review in POS tagging low resourced language using neural networks

In low resourced language, POS tagging is also a crucial step in NLP, so the increase of the number of research recently. In the Malayalam language, Akhil et al. [19] proposes a deep learning model for POS tagging. They used the available Malayalam corpus of 280000 tokens and trained different RNN models including GRU, LSTM and Bi-LSTM that shows better results than the available POS taggers reaching 98% in

terms of F-measure. Also, a bidirectional LSTM POS tagger for Shahmukhi is proposed in the study [20]. They first collected millions of tokens and manually annotated 130000 words to be tested in the study. They have made multiple experimentations on the effect of words representation on the accuracy of the Bi-LSTM model and conclude that combining word 2vec and Elmo representations achieve the best results with 96% in terms of accuracy. In the study [21], they used the bidirectional LSTM networks to POS multiple languages. The model takes as input the concatenation of both word embedding and characters or bytes representation. The model is not trained just for tagging but also on predicting the log frequency to decrease sharing representation between common and rare words. The result on the Danish corpus of 100000 tokens reaches 96% of accuracy. Bidirectional LSTMs process a sequence of inputs on both senses forward and backward then passing on to the next layer, for more details shown in the studies [22, 23].

The first research on Amazigh POS tagging was done by Outahajala et al. [24] using the first annotated corpus [25] of just 20k token, it was based on the traditional machine learning, that are CRF and SVM with an accuracy of 88.66% and 88.26% respectively. To achieve these results, they added lexical features such as the first and last i-character n-gram and performed 10-fold cross validation of the dataset. Later, in the study [26], they used semi-supervising tagging to enlarge the tagged corpus by using a previous supervised model for POS tagging and combine it with confidence measure to tag the unlabeled data. That enhanced the accuracy to reach 91% on CRF. Later, with the elaboration of another corpus of 60k token [27], the CRF and SVM algorithm are evaluated again in the study [28] giving an accuracy of 89% and 88% respectively. In addition, tree tagger as an independent language tagger has been evaluated in Amazigh language in the study [29]. A five cross validation was done to the corpus of 60000 tokens accuracy of 89%. The tag set used in all those Amazigh models is a set of 28 tags.

Table 1. Summary of the discusses state-of-art approaches in Related works section, NS (not specified)

	Language	Reference	Dataset	Tag-set	Proposed Model	Accuracy %	Baseline Comparison
Well studied languages	Arabic	[7]	75000 tokens of Arabic and Gulf tweets	44	Bi-LSTM	96	CRF tagger: 91% of accuracy
		[8]	151701 tokens of standard Arabic	47	BERT	91.68	HMM: 77.5% Brill: 83% Of accuracy
		[9]	6844 tokens of golf tweets	18	Bi-LSTM	91.2	SVM: 86% of accuracy
	Chinese	[11]	CTB5:508000 CTB7: 2075000 UD Chinese: 123200	NS	Bi RNN-CRF	94% F1score on joint segmentation and tagging	ZPar: 94% of accuracy
		[12]	40000 words of Chinese Buddhist	30	Bi-RNN	85.26	HMM: 42% of accuracy
		[13]	Chinese Penn Treebank	NS	Framework including three layers of BERT and BILSTM/Transformer and decoder	96.46 in POS tagging	Compared with other neural joint models
Low resourced languages	Malayalam	[19]	287000 tokens	36	Bi-LSTM	98% F1 measure	TnT: 80% CRF: 86% of accuracy
	Shahmukhi	[20]	130000 tokens	36	Bi-LSTM	96.12	Tree tagger:85.77% Stanford Tagger: 94.43% of accuracy
	Danish	[21]	100000 tokens	NS	Bi-LSTM	96.35	TnT: 94.24% CRF: 93.83% of accuracy
	Amazigh	[24]	20000 tokens	28	SVM CRF	SVM: 88.26 CRF: 88.66	NS
		[26]	20000 tokens	28	Semi supervised CRF	91	Supervised CRF: 91% of accuracy
		[28]	60000 tokens	28	SVM CRF	SVM: 88 CRF: 89	SVM: 88.26% CRF: 88.66% of accuracy
		[29]	60000 tokens	28	Tree Tagger	89	SVM: 88% CRF: 89% of accuracy

As we can see in Table 1, the neural networks approaches achieve better results than the classical machine learning models in all languages. In the case of Amazigh language, even with the extending of the size of the corpus didn't add a significant change on the performance of the SVM and CRF Amazigh taggers. So, in this paper, we propose the first neural networks approach for Amazigh POS tagging that achieved remarkable results of 97% in terms of accuracy.

3. AMAZIGH POS TAGGING USING DEEP LEARNING

In this section, we present our approach of Amazigh POS tagging using deep learning techniques. Deep learning has become the center of research recently, it has the advantage of generating the features directly from the input compared with the traditional machine learning techniques which use the handcrafted features. Our approach proposes different deep

learning models such as simple recurrent neural network (RNN), the gated recurrent unit (GRU), long-short-term-memory (LSTM) and finally bidirectional LSTM.

Recurrent neural networks (RNN) [30] are a class of conventional feedforward neural network that make possible the use of previous output as input while having hidden state, which makes them primordial for time series data problems. With an input of sequence $x_1, x_2, \ldots, x_n$, an RNN model computes the output vector y_t of each word x_t using these following equations:

$$h_t = H(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \qquad (1)$$

$$y_t = W_{hy}h_t + b_y \qquad (2)$$

where, W represents weight matrix connecting two layers (like W_{hy} is the weights between output and hidden layer), h_t is the vector of hidden states, b represents the bias vector (e.g., b_y is the bias vector of output layer) and H is the activation function of hidden layer. h_t holds information from previous step's hidden state h_{t-1}, so that RNN can exploit of all input history. However, the size of input history is limited due to the problem of vanishing gradient [31]. To solve this problem, Long-Short-Term Memory (LSTM) [32] has been created.

An LSTM network makes it possible for the standard RNN to keep memory of inputs for a long-time using gating techniques. LSTM network has three gates: an input gate, a forget gate which decides whether to keep or not a memory of precedent cell and an output gate. Nowadays, LSTM is used in multiple fields depending on machine learning. Bidirectional Long-Short-Term-Memory (Bi-LSTM) has the advantage of reading a text or sequence of input in both directions, forward and backward which supplies preceding and succeeding context. Gated neural network (GRU) can be considered a variation of LSTM with just two gates which are the update gate and the reset gate. The reset gate always uses the earlier hidden state and the current input then applicates a sigmoid function as activation.
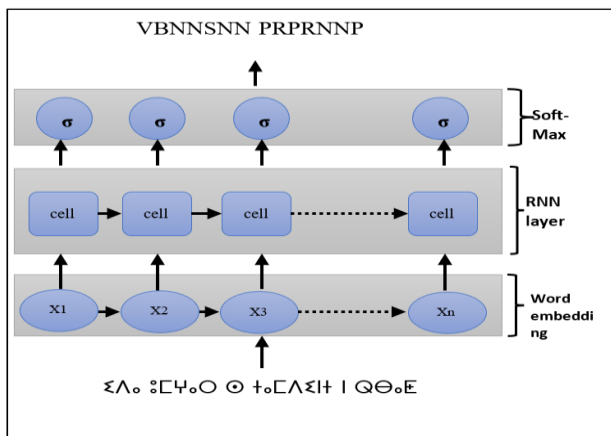

Figure 1. The workflow of our models

In this paper, we propose the implementation of the four types of neural network presented above such as simple RNN, LSTM, GRU and Bi-LSTM in the POS tagging of Amazigh language. The proposed model shown in Figure 1, includes three layers:

- A word embedding [33] layer.

- A neural layer with cells (RNN or LSTM or GRU or Bi-LSTM).
- A SoftMax layer.

Starting with a sentence of n words having tags: $y_1, y_2, \ldots y_n$, the tagger first trained to predict the tags. The sentence is transformed to $[w_1, w_2, \ldots w_n]$ where w_i is the code number given to the word at index i according to the existing vocabulary. This sequence is first transformed to vectors through word embedding layer then fed to the next deep learning layer that will be changing to test the different model (RNN, GRU, LSTM and Bi-LSTM) then finally, we realize a SoftMax on the results to obtain the predictions. As for the word embedding output, the number of neural cells, the parameters are chosen in such a way that the model performs better.

4. EXPERIMENT AND RESULTS

4.1 Dataset

Table 2. Amazigh dataset statistics

Dataset	Size	Type	Number of Sentences	Number of Tokens
Monolingual corpus	700MB	Utf-8	3231	60000

Table 3. The Amazigh tag-set

Grammatical Class	POS Tag
Noun said common	NN
Noun disignating kinship	NNK
Proper noun for person and place	NNP
Nouns of quality or Adjective	ADJ
Based verb	VB
Participle	VBP
Adverb	ADV
Determiner	DT
Conjunction	C
Focalizer	FOC
Particle indicating negation	NEG
Interjection	IN
Vocative	VOC
Preverbal particle	PRPR
Particle for predication	PRED
Particle for orientation	PROR
Other Particle	PROT
Personal pronoun	PP
Demonstrative pronoun	PDEM
Possessive pronoun	PPOS
Relative	REL
Interrogative	INT
Preposition	S
Strange word, Foreign	FW
Numeric values	NUM
Date	DATE
Residual and other	ROT
Punctuation	PUNC

In this experiment we used the dataset elaborated by study [27] which is a collection of Amazigh texts in form of tagged words inserted in CSV file of the order of 60K token. Table 2 shows the characteristics of this used dataset. To prepare this dataset for use in deep learning models, we realized a preprocessing step. We started by transforming the dataset into sentence where we assign an ID to each word, then, we

transform it into list of sentences of words with its tags: [(word1, tag1), (word2, tag2) ……., (wordN, tagN)] then realize the next processing steps such as encoding words into indices and tags to one hot encoding.

As known, the tagging process depends on a set of tags called tag-set which is a collection of part-of-speech tag(labels) to show the pos tag or other grammatical categories of each token in text. Each language has its tag-set, as for Amazigh language has its specific tag-set represented in Table 3, it is the same used in previous Amazigh natural language processing works like [27].

4.2 Results and evaluation

In this section, we present the results obtained using the models shown above and we realize multiple experiments by changing the size of the cells to find the best model in accuracy. The results are shown in Table 4.

Table 4. Performance of different models using multiple size of cells

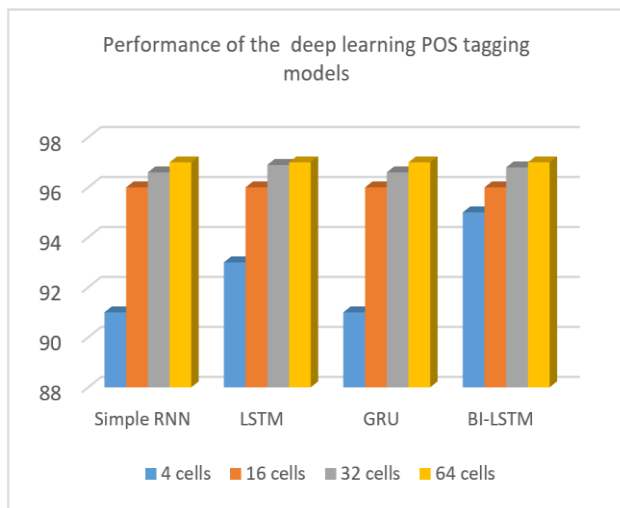| Model	Number of Cells	Accuracy
RNN	4	91
	16	96
	32	96.6
	64	97
LSTM	4	93
	16	96
	32	96.9
	64	97
GRU	4	91
	16	96
	32	96.6
	64	97
Bi-LSTM	4	95
	16	96
	32	96.8
	64	97.1

Figure 2. Performance of the proposed deep learning models on different size of cells

As shown in Table 4 and Figure 2, the accuracy of all the models is improving when we increase the number of cells. In RNN and GRU models, the accuracy starts from 91% in case of 4 cells to reach 96.9% in case of 64 cells. As for LSTM the accuracy is 93% when using 4 cells then reaches 97% in 64

cells, however in Bi-LSTM the results are better even when using just 4 cells (95% of accuracy) and the accuracy in case of 64 cells reaches 97.1%.

4.3 Baseline comparison

In this experiment, we have chosen the baseline as the existent approaches in part-of-speech tagging Amazigh. Those existent approaches were based on classical machine learning models such as Support vector machine learning (SVM) and conditional random fields (CRF) in addition of tree-tagger as a language independent tagger [28, 29]. The results presented in Table 4 and Figure 3 show that our proposed approach outperforms the baseline approaches even with the size of the dataset used which can be considered small.

To highlight the effectiveness of bidirectional long-short-term-memory on POS tagging regardless of the language, we dress Table 5. This table presents the comparison of the accuracy of Bi-LSTM tagger on cited languages from the literature review. As shown in this table, the Bi-LSTM achieves good results in all languages even with different sizes of dataset. Despite the small size of the Amazigh dataset in comparison with the other languages like Danish and Shahmukhi, our tagger exceeds their taggers.
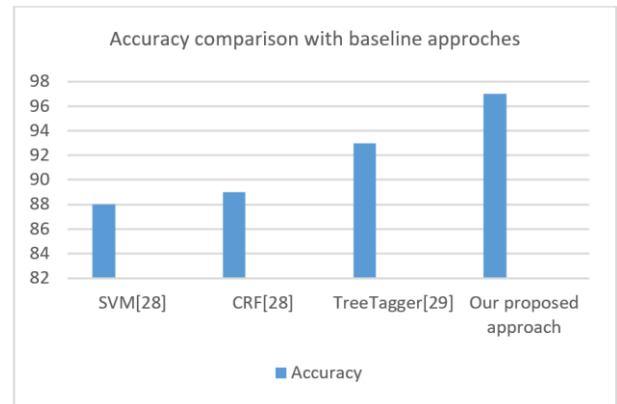

Figure 3. Accuracy comparison between our proposed approach and the baseline

Table 5. Performance of Bi-LSTM model on different languages

	Language	Dataset	Accuracy %
[7]	Arabic	75000 tokens of Gulf Arabic tweets	96
[19]	Malayalam	287000 tokens	98% F-measure
[20]	Danish	100000 tokens	96.35
[21]	Shahmukhi	130000 tokens	96.12
Our work	Amazigh	60000 tokens	97.1

5. CONCLUSIONS

5.1 Significance of this work

Part-of-speech tagging is an essential step in natural language processing which means that developing a performant POS tagger will enhance the performance of other applications depending on it. Amazigh language is a very low resource language and suffers from lack of research in this field. The existing POS taggers achieve less than 91%

accuracy which isn't comparable with the well-known language reaching more than 97% of accuracy. Our work in this paper presented a new approach of tagging Amazigh language using deep neural networks especially Bi-LSTM model. The achieved results reach 97% which is considered state-of-art results in POS tagging Amazigh language.

5.2 Future works

Natural language processing has become the most field interesting research in the world including big firms of social media and communication, etc. Languages called well-known or well-studied like English, Arabic experience a rapid development in contrary of languages said low resourced like Amazigh. For this reason, we will continue to contribute to the development of the NLP tools, not just POS tagging but also Named entity recognition. Amazigh language still suffers from the lack of this basic tools but essential in any NLP project. As well as we will enlarge the annotated dataset using this developed POS tagger and make it public for other researchers.

5.3 Mean takeaways

In this Paper, we presented advances in using deep learning models in morphological problematics such as parsing and especially POS tagging. It's known that classical machine learning approaches like CRF and TnT have good results in well-known languages. However, those approaches don't achieve comparable results in low resources languages like Malayalam, Shahmukhi or Amazigh as presented in this paper. With the use of deep neural networks especially RNN, the accuracy of the tagger is significantly enhanced. The performance of those taggers reaches the best results when using the bidirectionality on the LSTM cells in all languages. This structure has the advantage of injecting the preceding and succeeding context in each timestep.

ACKNOWLEDGMENT

REFERENCES

[1] Greenberg. (1966). The languages of Africa. The Hague.

[2] Ouakrim, O. (1995). Fonética y fonología del Bereber. Servei de Publicacions de la Universitat Autònoma de Barcelona, Vol. 3.

[3] Boukhris, F., Boumalk, A., Elmoujahid, E., Souifi, H. (2008). La nouvelle grammaire de l'amazighe. IRCAM. https://www.ircam.ma/sites/default/files/2021-02/nouvel-gram-amazigh_0.pdf.

[4] Laabdelaoui, R., Boumalk, A., Iazzi, E.M., Souifi, H., Ansar, K. (2012). Manuel de conjugaison de l'amazighe. IRCAM, Rabat, Morocco. https://tal.ircam.ma/conjugueur/manuel.pdf.

[5] Rabiner, L., Juang, B. (1986). An introduction to hidden markov models. IEEE Assp Magazine, 3(1): 4-16. https://doi.org/10.1109/MASSP.1986.1165342

[6] Lafferty, J., McCallum, A., Pereira, F.C. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. Proceedings of the Eighteenth International Conference on Machine Learning, 282-289. https://dl.acm.org/doi/10.5555/645530.655813.

[7] AlKhwiter, W., Al-Twairesh, N. (2021). Part-of-speech tagging for Arabic tweets using CRF and Bi-LSTM. Computer Speech & Language, 65: 101138. https://doi.org/10.1016/j.csl.2020.101138

[8] Saidi, R., Jarray, F., Mansour, M. (2021). A BERT based approach for Arabic POS tagging. In Advances in Computational Intelligence: 16th International Work-Conference on Artificial Neural Networks, IWANN 2021, Virtual Event, June 16-18, 2021, Springer International Publishing. Proceedings, Part I, 16: 311-321. https://doi.org/10.1007/978-3-030-85030-2_26

[9] Alharbi, R., Magdy, W., Darwish, K., AbdelAli, A., Mubarak, H. (2018). Part-of-speech tagging for Arabic Gulf dialect using Bi-LSTM. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). http://alt.qcri.org/resources/da_resources/

[10] Cortes, C., Vapnik, V. (1995). Support-vector networks. Machine Learning, 20: 273-297. https://doi.org/10.1007/BF00994018

[11] Shao, Y., Hardmeier, C., Tiedemann, J., Nivre, J. (2017). Character-based joint segmentation and POS tagging for Chinese using bidirectional RNN-CRF. arXiv Preprint arXiv: 1704.01314. https://doi.org/10.48550/arXiv.1704.01314

[12] Qin, L. (2015). POS tagging of Chinese Buddhist texts using recurrent neural networks. Technical Report, Stanford University.

[13] Wu, L., Zhang, M. (2021). Deep graph-based character-level chinese dependency parsing. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 29: 1329-1339. https://doi.org/10.1109/TASLP.2021.3067212

[14] Li, H., Zhang, Z., Ju, Y., Zhao, H. (2018). Neural character-level dependency parsing for Chinese. In Proceedings of the AAAI Conference on Artificial Intelligence, 32(1). https://doi.org/10.1609/aaai.v32i1.12002

[15] Yan, H., Qiu, X., Huang, X. (2020). A graph-based model for joint chinese word segmentation and dependency parsing. Transactions of the Association for Computational Linguistics, 8: 78-92. https://doi.org/10.1162/tacl_a_00301

[16] Wang, P., Qian, Y., Soong, F.K., He, L., Zhao, H. (2015). Part-of-speech tagging with bidirectional long short-term memory recurrent neural network. arXiv Preprint arXiv: 1510.06168. https://doi.org/10.48550/arXiv.1510.06168

[17] Tian, Y., Song, Y., Ao, X., Xia, F., Quan, X., Zhang, T., Wang, Y. (2020). Joint Chinese word segmentation and part-of-speech tagging via two-way attentions of auto-analyzed knowledge. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 8286-8296. http://dx.doi.org/10.18653/v1/2020.acl-main.735

[18] Gui, T., Zhang, Q., Huang, H., Peng, M., Huang, X.J. (2017). Part-of-speech tagging for twitter with adversarial neural networks. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 2411-2420. http://dx.doi.org/10.18653/v1/D17-1256

[19] Akhil, K.K., Rajimol, R., Anoop, V.S. (2020). Parts-of-Speech tagging for malayalam using deep learning techniques. International Journal of Information Technology, 12: 741-748. https://doi.org/10.1007/s41870-020-00491-z

[20] Tehseen, A., Ehsan, T., Liaqat, H.B., Ali, A., Al-Fuqaha, A. (2023). Neural POS tagging of shahmukhi by using contextualized word representations. Journal of King Saud University-Computer and Information Sciences, 35(1): 335-356. https://doi.org/10.1016/j.jksuci.2022.12.004

[21] Plank, B., Søgaard, A., Goldberg, Y. (2016). Multilingual part-of-speech tagging with bidirectional long short-term memory models and auxiliary loss. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2: 412-418. http://dx.doi.org/10.18653/v1/P16-2067

[22] Goldberg, Y. (2016). A primer on neural network models for natural language processing. Journal of Artificial Intelligence Research, 57: 345-420. https://doi.org/10.1613/jair.4992

[23] Cho, K. (2015). Natural language understanding with distributed representation. arXiv Preprint arXiv: 1511.07916. https://doi.org/10.48550/arXiv.1511.07916

[24] Outahajala, M., Benajiba, Y., Rosso, P., Zenkouar, L. (2011). Pos tagging in amazighe using support vector machines and conditional random fields. In Natural Language Processing and Information Systems: 16th International Conference on Applications of Natural Language to Information Systems, NLDB 2011, Alicante, Spain, June 28-30, 2011. Springer Berlin Heidelberg. Proceedings, 16: 238-241. https://doi.org/10.1007/978-3-642-22327-3_28

[25] Outahajala, M., Zenkouar, L., Rosso, P. (2011). Building an annotated corpus for Amazighe. In Proceedings of 4th International Conference on Amazigh and ICT, pp. 1-10.

[26] Outahajala, M., Benajiba, Y., Rosso, P., Zenkouar, L. (2015). Using confidence and informativeness criteria to improve POS-tagging in amazigh. Journal of Intelligent & Fuzzy Systems, 28(3): 1319-1330. https://doi.org/10.3233/IFS-141417

[27] Amri, S., Zenkouar, L., Outahajala, M. (2017). Build a morphosyntaxically annotated amazigh corpus. In Proceedings of the 2nd International Conference on Big Data, Cloud and Applications, pp. 1-7. https://doi.org/10.1145/3090354.3090362

[28] Samir, A., Lahbib, Z., Mohamed, O. (2018). Amazigh PoS tagging using machine learning techniques. In Innovations in Smart Cities and Applications: Proceedings of the 2nd Mediterranean Symposium on Smart City Applications. Springer International Publishing, 2: 551-562. https://doi.org/10.1007/978-3-319-74500-8_51

[29] Samir, A., Lahbib, Z. (2019). Training and evaluation of treetagger on amazigh corpus. International Journal of Intelligent Enterprise, 6(2-4): 230-241. https://doi.org/10.1504/IJIE.2019.101130

[30] Elman, J.L. (1990). Finding structure in time. Cognitive Science, 14(2): 179-211. https://doi.org/10.1207/s15516709cog1402_1

[31] Hochreiter, S., Bengio, Y., Frasconi, P., Schmidhuber, J. (2001). Gradient flow in recurrent nets: the difficulty of learning long-term dependencies. In A Field Guide to Dynamical Recurrent Neural Networks, IEEE Press. https://doi.org/10.1109/9780470544037.ch14

[32] Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8): 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

[33] Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J. (2013). Distributed representations of words and phrases and their compositionality. Advances in Neural Information Processing Systems, NIPS. 26. https://arxiv.org/pdf/1310.4546.pdf