ISSN: 0067-2904

Survey For Arabic Part of Speech Tagging based on Machine Learning

Zahra K. Mohammed, Nada A. Z. Abdullah*
Department Computer Science, College of Science, University of Baghdad , Baghdad, Iraq

Received: 19/5/2021 Accepted: 6/8/2021 Published: 30/6/2022

Abstract

The Arabic Language is the native tongue of more than 400 million people around the world, it is also a language that carries an important religious and international weight. The Arabic language has taken its share of the huge technological explosion that has swept the world, and therefore it needs to be addressed with natural language processing applications and tasks.

This paper aims to survey and gather the most recent research related to Arabic Part of Speech (APoS), pointing to tagger methods used for the Arabic language, which ought to aim to constructing corpus for Arabic tongue. Many AI investigators and researchers have worked and performed POS utilizing various machine-learning methods, such as Hidden-Markov-Model (HMM), Brill, Maximum-Match (MM), decision tree, bee colony, Neural-Network (NN), and other hybrid methods.

This survey groups a number of published papers based on the Arabic Language Applications (ALP) towards tagging related problems utilized and approaches with the difference between types of tags used. It addresses and tries to identify the gaps in the current studies putting a foundation for future studies in this field.

Keywords: POS tagging, Arabic language, Machine learning, NLP.

استبيان تمييز اجزاء كلام اللغة العربية بناءً على التعلم الآلي

زهراء كاظم محمد، ندا عبد الزهرة عبدالله*
جامعة بغداد, كلية العلوم, قسم علوم الحاسوب بغداد- العراق

الخلاصة

اللغة العربية هي اللغة الأم لأكثر من 400 مليون شخص حول العالم ، وهي أيضًا لغة لها وزن ديني ودولي مهم. لقد استحوذت اللغة العربية على نصيبها من الانفجار التكنولوجي الهائل الذي اجتاح العالم ، وبالتالي فهي بحاجة إلى معالجة تطبيقات ومهام معالجة اللغة الطبيعية.

تهدف هذه الورقة إلى مسح وجمع أحدث الأبحاث المتعلقة بجزء الكلام العربي .(APoS) الإشارة إلى أساليب الوسم المستخدمة في اللغة العربية والتي يجب أن تهدف إلى بناء مدون للغة العربية. تم عمل العديد من بحوث الذكاء الاصطناعي في مجال اجزاء الكلام العربي باستخدام طرق مختلفة للتعلم الآلي مثل–Hidden Markov–Model (HMM)و Brill و Maximum–Match (MM) وشجرة القرار ومستعمرة النحل والشبكة العصبية (NN) و طرق هجينة أخرى.

*Email: Nada.abdullah@sc.uobaghdad.edu.iq

يجمع هذا الاستبيان عددًا من الأوراق المنشورة بناءً على تطبيقات اللغة العربية (ALP) من أجل وضع علامات على المشكلات والأساليب ذات الصلة مع الاختلاف بين أنواع العلامات المستخدمة. يعالج ويحاول تحديد الثغرات في الدراسات الحالية ووضع أساس للدراسات المستقبلية في هذا المجال.

1. Introduction

Computer linguistic experts have worked on many mathematical methods to explain natural language and develop how machines deal with it. Scientists have noted that to reach human brain intelligence in Natural Language Processing (NLP) , they must try to build a machine that knows how to deal with languages. A machine that can think, have dialogue with people, and understand different languages and dialects.[1] [2] Thus lately, the fields under discourse and grammatical marking have made tremendous progress. Machine-learning methods that deal with text Corpora are considered a valuable part of NLP applications and programs. The Arabic language (Arabic language) is considered one of the most influential six official languages in the world. On the other hand, Arabic NLP is the most difficult and least researched. Holy Qur'an and divine miracle have been written in Arabic language and the number of people who speak the language is also large.. More than 330 million people speak in Arabic, but there are still problems encountered in processing and investigation [3].

The great growth in Arabic language texts on the internet required more work to translate the information accurately. Until now, despite the great global development in the field of translation and textual applications, the parts of speech Tagging have not been completed clearly for the Arabic language. Consequently, Arabic Computational Linguistics scientists focus on NLP (ALP) development in their research. POS Tagging means extracting all parts of speech from the text (written or spoken) and then classifying each word into the part that it represents (Noun, Verb, Adjective, etc...). Natural language processing (NLP) involves a set of phases or stages. It is can be summarized in five stages that include various linguistic treatments, starting from phonetics and progressing to the sciences of grammatical and morphological analysis (as shown in Figure 1)

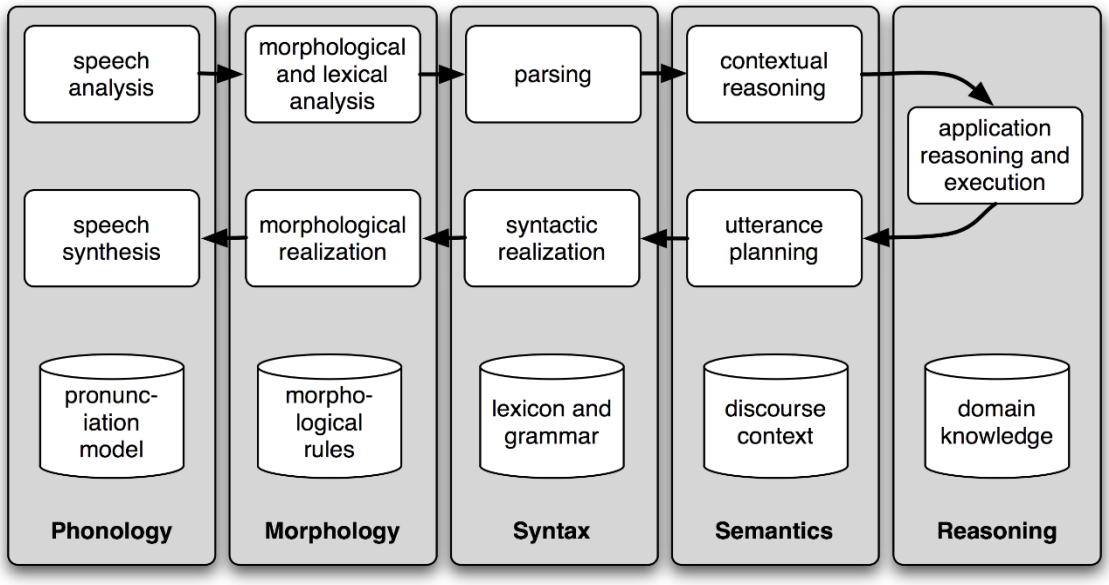

Figure 1- the main phases of NLP

POS task takes place in the Syntactic Analysis stage, which is the process of specifying the words into their appropriate tag. The POS tagging is the first logical level of annotation.[1] The POS process is an important primitive stage for numerous NLP applications, such as

Parsing Systems, Information Retrieval, Building of Dictionaries, Speech Synthesis Systems, and Word Processing [4]. Supervised and unsupervised methodologies are the two prime ways used for POS Tagger tasks. As mentioned on the well-known programmatic website MathWorks, machine learning techniques mentioned in (Figure 2) confirming the research's use of smart methods in word processing and POS under the two previously mentioned methods.

These two branches have many methods that can be used in tagging texts or corpus of the Arabic language. Moreover, the process of building huge structures of texts in various dialects has attracted a lot of attention for building the main repository for language processing operations that are known as corpora.

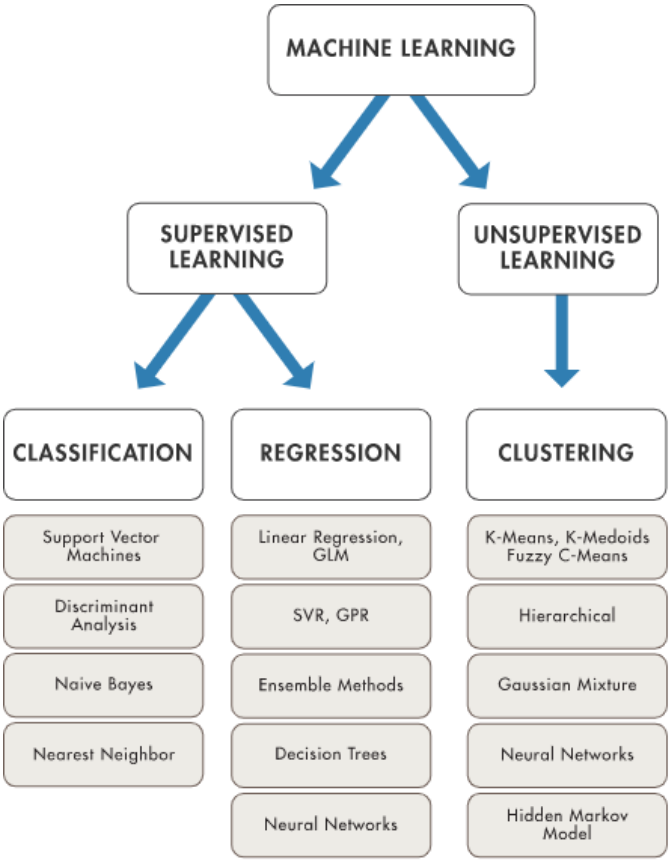

Figure 2-Machine Learning Methods

NLP involves a set of tasks inclusive in many applications; the POS is considered one of the important areas. POS signing is a vast discussion area in computational linguistics. The POS method depends on choosing the correct grammatical feature for which it belongs. Grammatical countenance and structure belong to the syntax category, for example, thing, verb, or adjective [5]. In the semantics testing/analysis, the corpus is a large and structured set of machine-readable texts that have been produced in a natural communicative setting. It is a very big staff of linguistic from the intended language that is used for testing programs designed to process the NLP or designing NLP application. It is considered the main template for extracting semantics and linguistic properties, characteristics, and features.

Corpus makes grammatical merit tagger frameworks with the desired semantic learning that aid fix the uncertainty in the tongue without the need for rigid phonetic qualification. For the English language/dialect, colossal overwork has been made to build many corpora that were separated from composed or written spoken words.[1] This Transformation began with

constructing Brown corpus in 1961 with one million English dialect words, motivating little corpora that assumed an important part in a gathering of English dialect dictionaries.[6] Arabic tongue contradicts more than Indo-European-tongues, is viewed as more complicated than English dialect in diverse perspectives. The lack of free bases/corpora for apps that deal with the Arabic language and the demand for labeled Arabic corpus that can be utilized as a part of preparing and testing Arabic grammatical form tagger are two perspectives that are the significance of our investigation. This research is concerned with Ar-POS and will attempt to present a new method for dealing with parts of speech in the Arabic language after examining and comparing a set of methods used in this field.

2. Most Recently Arabic POS Tagging (APoS) Research

In this section, a summary of the latest works in APoS tagging is presented

Two survey articles are recommended to build a complete idea of the tags used before 2016. [1]in their review surveys wrote scientific summaries about the NLP task studied at the time. It is good to review the two studies to form an complete idea about the subject. However, there are several types of research chosen in this field of research to be able to reach a result or rather a good method that we can rely on for the sober language.

[7] employed three tagger approaches for APoS. Besides common taggers Hidden-Markov-Model (HMM) and Brill-tagger, he used maximum match (MM) tagger. Two different contexts for the last model MM were employed in a way of using the master-slave technique. He used these methods to attempt a new direction in the Arabic tagging process, which is the Master-Slave technique. His idea was using automated training of machine learning methods and employing them in the Master-Slave technique. In (45) k chosen words from Al-Watan corpus-data as a Dataset used to work in two tagging levels. Tag-set applied was taken from wealthy morphosyntactic tag set that is especially constructed by him for Arabic with 2500 tag options. The first level of tagging consists of five main tags and each of them has several branches where the second level comes. The second level can be with different tag numbers, as each part has its appropriate number of secondary tag branches. He displayed the result of accuracy on an Ar corpus when he used Brill approach as master, which was 86.43%. When MM approach became the master it was 83.26%, and when the HMM tagger the accuracy was 88.81%. In the other two cases, he appointed the master with HMM and took the slave Brill for the first time and Brill+MM in the second time and found that the accuracy percentages were 89.40% and 90.05%. The highest percentage of accuracy was achieved in his methods by using HMM as the master and Brill+MM as slaves with 90.05%.

)[8] used grammatical classification techniques depending on the decision tree-based tagger. This tagger was used in many languages, so they used it in the Arabic language. It is a probabilistic teaching method that the researchers used to define the word for which POS follows. They apply tree tagger on MSA (Modern-Standard-Arabic). In other words, they traded Ar and classical Ar text. This means that they dealt with the language from two sides, the original side which is the official Ar from the Al-Mus'haf dataset, and the other side which is considered a dialect taken from the NEMLAR Dataset. Nevertheless, the researchers adopted one method in the tagging process, the tagset that they constructed has 11 basic tags and it isn't general to all languages. So,11 main tagsets are used for the pair of corpora that is especially for the Arabic language. As an accuracy result tagger achieved a 99.43% accuracy percentage with Al-Mus'haf vowel and a 92.60% accuracy percentage un vowel for the same corpus with a total number of (78,121) words. NEMLAR Arabic set with (500,000) words had achieved 81.94% with Training corpus 450,000 and the rest as test corpus.

)[9]in their research presented a method to create a standardization of all tags for the whole language, this is because of the absent specific path for the tag-sets in such difficult language. In their research, they made a useful summary of all the previous tagsets previously found in all research for the Arabic language and its branches. They explained that the process can

produce better results in the case of standardization of the tags. In their research, they studied all previously used existing APoS tag-sets after dividing the methods used. The researchers selected a group of tag-sets they considered good according to their opinion. Moreover, they referred to the tags taken from other foreign languages and explained that it is not correct to adopt tags made for other languages, as they must be specific to the Arabic language to obtain desired results. The proposed APoS tag-sets is built of three corners the first for Noun referred to Ar by (اسم) the second for Verb referred to (فعل) and the third for Particle (حرف). Each one in these three pillars, determined by the number of branches and leaves until the total number of APoS tag-sets rises to 110. As a final result of the work, the percentage studied in previous searches [5] increased significantly, this shows that the importance of dividing the Arabic language into appropriate tagsets gives better results in all cases. Finally, it should be noted for formal Arabic with each level, the number of branches increases, and the reliability percentage decreases. Al-Mus'had started with 4 main tag-sets and 97.18% percentage and ended in level 4 with 95 tag-sets and 91.65% reliability. On the other hand, NEMLAR started with 5 main tag-sets and 97.15% percentage and ended in level 4 with 107 tag-sets 97.55% reliability.

[10] in their research used the bee colony AI method as the AI algorithms to mimic nature for grammatical Ar tag. It can be said that the researchers reported using a new method of research that was freed from the tree tagger. Compared with previous studies, the reliability ratio was more than the highest score by two percent, but there could be no comparison because this research used the KALIMAT corpus that contains more than 18million Ar words with different categories.

[11] innovated (innovate) methods to develop hybrid methods previously used, so they can achieve better results than previous ones. The researchers' idea was to make the neural network key to combine the rule-based with machine learning technique. Using Word2vector after the translation was the way to activate the role of the NN algorithm. Since dealing with English words seems easier than Ar and the machine learning methods recognize numbers better. They used the Penn-Tree-Bank tagset, which consists of 21 Ar-tag-set, to tag the words of the ALKALIMAT corpus. The search result improved two percent from the hybrid methods, but the highest result from all trials was 92%.

[12] modulated and merged well-known probability and statistical learning methods. What is notable about his research is that he used his corpus of 19,322 words containing 10.300-Nouns, 5,70-Verbs,1620-particles, and 1701-Punctuation. The researcher combined probability Stanford POS with statistical Khoja POS. The study focused on the indicated tag between nouns and verbs, However, in both cases, (nouns or verbs) the combined method achieved a higher percentage than each method separately.

[13] developed a model that seeks to apply an HM method on all Arabic language Datasets only. His research showed that the highest rate of use of this model was achieved with the Qur'an by 99.4%. It was also a common method of the studied model with the Viterbi algorithm. The percentages offered for Arabic were also compared with percentages for other languages, including English, Spanish, and others. On the other hand, the researcher suggested that this statistical method can be adopted as a basis for other linguistic operations that depend on POS. He indicated that its use exceeded other methods, including the Neural Network.

)[14] according to their research published in 2019, have chosen a new method in the field of artificial intelligence and employed it to process the Arabic language tags. While the in-depth training became the focus of public and researchers' attention, the researchers used in-depth methods in choosing the tags to be trained and tested. They proved that LSTM and Word2Vec can be effective in the Arabic language field, as they achieved rates of 99.72% as the highest percentage and 97.33% as the lowest percentage. However, it seems that there are some points

to analyze the idea of using the two deep learning algorithms, do the researchers consider the Word2Vector as a method or as a model. From the research, we expect to solve the concept well but there is no mention of the topic, as the title refers to the algorithm itself and the method indicates that they used Word2Vector as a model.

After the interruption of research for a short period -As there was no specialized Arabic research in this field in 2019 except Yousif (2019)[13] - The researchers Al-Khwiter and Al-Twairesh published in (2020)[15] featured search. Their database was based on tweets from a social media site. Unlike other research, the data was gathered from Twitter media and then suspended and offered the sorted tagset. Nowadays, there are a few projects that have used Twitter's system rather than other media programs as a database instead of using corpus, most of it was about the Arabic dialects. The idea of taking data from Twitter and conducting research on it started by [16] for English tweets. The researchers later exploited the idea for the Arabic dialects, and one of them is the research in which three researchers collaborated ([17], [18]), but the research was interested in the dialect instead of the formal language. What we are interested in is that the most recent research was by researchers in [15], who are concerned with the official Arabic. They built the corpus from Tweets after several stages of the procedure and then tested it for one statistical method and the other for deep learning methods. The corpus consisted of three types: the official modern language MSA, Gulf dialect, and a combination of the two types. The number of tweets assigned to modern Arabic was 1983 tweets, 1017 tweets for GLF, and the total number of tokens was 75,677. In addition, 40 tag types were used for tagging process with 4 new tags specialized for Twitter. They experimented with CRF on three identical sets: the first for unigram and bigram features, second for morphological features, and the third for expanded window size by 8. The second group had the highest score for the three types of tweets Mixed 90.2 MSA 88.5 and GLF 85.4. On the other hand, the Bi-LSTM had the highest results for the whole groups type and the highest ratio was for Mixed Dataset 96.5. They compared their result to Darwish's one, they scored higher in the MSA by a percentage of 0.2. Table 1 presents a summary for the research's analysis in this section.

Table 1- Summary for Most Recently APoS Research

Year	Author/s	Methodology	Accuracy	Number of APoS Tag-sets	Dataset	Number of words
2015	Aliwy [7]	Master-Slaves Technique		Aliwy with 2500 tags not identical	Al-Watan corpus-data	45k word
		Master: Brill	86.43%			
		Master: HMM	83.26%			
		Master: HMM	88.81%			
		Master: HMM & Slave: Brill	89.40%			
		Master: HMM & Slave: Brill +MM	90.05%			
2016	Zeroual and Abdelhak [8]	decision tree	99.43%	Their tag-set with 11 basic tags.	Al-Mus'haf (vowelled)	78,121

			92.60%		Al-Mus'haf (unvowelled)	
			81.94%		NEMLAR	500,000
2017	Zeroua, Abdelhak and Rachid[9]	The same of 2016 decision tree tagger different tag-sets	97.18% 94.02% 91.35% 91.65%	Level[1]::4 Level[2]::26 Level[3]::79 Level[4]::95	Al-Mus'haf	78,121
			97.15% 93.86% 95.74% 97.55%	Level[1]::5 Level[2]::12 Level[3]::63 Level[4]::107	NEMLAR	500,000
2018	Alhasan and .Al-Taani[10]	bee colony	98.2%	All tags in corpus	KALIMAT	18,167,183
2018	Farrah, El Manssouri, Ziyati, and Ouzzif,[11]	Hybrid (Neural network & rule-based)	92%	21	KALIMAT	18,167,183
2018	Mohamed Labidi[12]	combined method(Stanford& Khoja)	nouns-90% verbs-92%	Three main tags	Their textual corpus	19,322
2018	Alrajhi and El-Affendi [14]	LSTM	99.72% for morpheme 99.18% for tagging words	37 tags for words 87 tags for combined morpheme	Quranic text corpus	77,915
		Word2Vec	99.55% for morpheme 97.33% for words.			
2020	AlKhwiter and Al-Twairesh [15]	CRF [BLmorph features]	Mixed 91.6% MSA 92.6% GLF 91.2%	48 tagset	GLF, MSA, Mixed	Mixed 75,677 3000tweets MSA 25,460 1000tweets GLF 22,474 1000tweets
		Bi-LSTM	Mixed 96.5% MSA 95.6% GLF 95.0%			

3. Discussion:

When using Arabic grammar, variations in the topic of tags and their number is predictably different from one study to the next and from one period of time to another. The differences in its hierarchical grammatical context and interweaving features are what set the Arabic language apart. Alian and Awajan (2019) offer 13 distinct types of Ar-tags in [20]. These tags are either based on the Ar language itself or were created for the English language and then utilized in some way for the Arabic language. The majority of these collections are businesses with three core tags (Verb, Noun, and Particle), from which various subsets or subclasses branch off. El-Kareh&Al-Ansary consists of 72 tags, Buckwalter with almost 500, Reduced 24, Extended 75, PATB 400, Alshamsi&Guessom 55 tag, ARBTAGS 28, etc. While noting differences in the groups, we find that some of them are complex, dedicated, and intertwined, some of the simple

As a result of studying the methods used in previous research, it is noted that not many machine methods have been used with this difficult language, and more research can be conducted, and other methods tested. All Arabic HMM methods indicated in [13] were displayed with their respective proportions, where the highest percentage was for Alhasan's research with optimization algorithm that relies on HMM. It is possible to resort to other methods or find a new way of using HMM to better serve Arabic Tagging. It can also be observed that there is little discussion of deep learning in the field of clarifying the Arabic language tag and the effectiveness of these methods for it. In [14][15] The LSTM Deep method was used and achieved very effective results 99.72% for the morpheme Quranic tag. While the second method was about Bi-LSTM and achieved 88.5 for MSA. The remaining methods (Master-Slave HMM+Brill, Hybrid NNs, and Rule-Based) presented accuracy were less than LSTM, Bilstm, and BCO. The trend around new techniques in the field of computer science towards deep structured learning invites us to keep pace with this development.

On the Datasets side, Al-Watan and NEYMAR corpora in [7][19] are considered as one of the relatively older datasets that have been extensively researched by many machine learning algorithms. The researchers in [10][11] used ALKALIMAT corpus, which contains a large number of words with around 18 million word. All million words are MSA collected from news, articles, and magazines. Moreover, we saw that earlier research [15] choose to take tokens from the social web application and build its corpus instead of having a ready one. The idea in choosing the appropriate dataset for the search is to select the data that represents the Arabic language in an ideal and typical way. There is no provenance for a precise study like Arabic linguistic that derived from the Holy Quran. Since it is a miracle of the language itself, it means that achieving ratios, even if they are few in the study, will mean a lot of work because the Qur'an summarizes the whole language. Moreover, the language itself derives from it. It is also considered a text that cannot be altered or tampered with, and it is a static and eloquent linguistically text.

4. Conclusion and Future Work

To find the strength and weakness points of the previous work, to be used for future work, two problems were identified to be avoid.

4.1- Drawbacks in Datasets:

Specific resources must be available to start any study, and what cannot be doubted is the choice when starting to choose the required rules and resources. Through the presented narration, many Arab databases (databases) have been repeated, but some of them are considered old and poor. In addition, most of the databases did not consider the language's small speech sections. It also sometimes studies and displays the language in a prosaic way. The trend is increasing towards dialects while databases are still scarce and need attention. However, we can exclud from this criticism the Holy Qur'an Corpus [21], because it is effective, connected to details, and is accurate in the fewest parts

4.2- Drawbacks in Tag-Sets:

The morphological system of the classical Arabic language can be placed in the form of a table whose vertical dimension is a hierarchy of division buildings, and the horizontal dimension of this table is the structure of inflection. The buildings of division are subdivided into absolute formulas and forms, and that the conjugative buildings branch into affixes and appendices. The study of the classical Ar hierarchy requires a fixed approach in the Tag-set. It cannot be said that there is a fixed tag with numbers and details arranged for the whole language and this makes the matter more difficult. Also, increasing the number of tags is not a good approach if it is not consistent with the text used.

As a summary of the practical narrative on the classification of APoS, previously there were only two main methods, statistical and probabilistic. It is worth noting that most of the research and surveys were compared in this way. But nowadays, new pathways have opened, such as deep learning. At present, new and varied methods have been introduced and new approaches have been taken, such as hybrid and complex techniques and deep learning, etc. Most of the methodologies were for different languages, while Arabic got a smaller share compared to its popularity. Most of the systems and tools were not fully suitable for Arabic, while it was classified as poor in research in this aspect. The Arabic literature hasn't been developed, modernized, and an adequate amount of research in the field of APoS. Tagging with ANNs and Genetic algorithms and Deep learning is a new approach in Arabic NLP, but it is absent from Arab uses as it was not given much opportunity. The POS is an important rule for conducting other Arabic applications. Therefore, the basis must be developed pave the way to help in other applications.

5. References:

[1] S. Salameh, "A Review of Part of Speech Tagger for Arabic Language," International Journal of Computation and Applied Sciences IJOCAAS, Volume 4, Issue 3, June 2018, ISSN: 2399-4509no. July, 2018.

[2] K. Almeman and M. Lee, "Automatic building of Arabic multi dialect text corpora by bootstrapping dialect words," *2013 1st Int. Conf. Commun. Signal Process. Their Appl. ICCSPA 2013*, no. February, 2013, doi: 10.1109/ICCSPA.2013.6487247.

[3] A. Farghaly, A. Farghaly, K. Shaalan, and Khaled, "Arabic Natural Language Processing: Challenges and Solutions," *ACM Trans. Asian Lang. Inf. Process.*, vol. 8, pp. 1-, 2009.

[4] R. A. Abumalloh, H. M. Al-Sarhan, O. Bin Ibrahim, and W. Abu-Ulbeh, "Journal of Soft Computing and Decision Support Systems Arabic Part-of-Speech Tagging," vol. 3, no. 2, pp. 45–52, 2016.

[5] J. Yousif, "Real Time Arabic Character Classifying System," *Irbid J. Res.*, vol. 7, pp. 15–32, 2004.

[6] M. Marcus, M. Marcinkiewicz, and B. Santorini, "Building a Large Annotated Corpus of English: The Penn Treebank," *Comput. Linguist.*, vol. 19, pp. 313–330, 2002.

[7] A. H. Aliwy, "Combining POS taggers in master-slaves technique for highly inflected languages as Arabic," *Proc. - 2015 Int. Conf. Cogn. Comput. Inf. Process. CCIP 2015*, pp. 1–5, 2015, doi: 10.1109/CCIP.2015.7100682.

[8] I. Zeroual and L. Abdelhak, "Adapting a decision tree based tagger for Arabic," *2016 Int. Conf. Inf. Technol. Organ. Dev. IT4OD 2016*, pp. 2–7, 2016, doi: 10.1109/IT4OD.2016.7479306.

[9] I. Zeroual, A. Lakhouaja, and R. Belahbib, "Towards a standard Part of Speech tagset for the Arabic language," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 29, no. 2, pp. 171–178, 2017, doi: 10.1016/j.jksuci.2017.01.006.

[10] A. Alhasan and A. T. Al-Taani, "POS Tagging for Arabic Text Using Bee Colony Algorithm," *Procedia Comput. Sci.*, vol. 142, pp. 158–165, 2018, doi: 10.1016/j.procs.2018.10.471.

[11] S. Farrah, H. El Manssouri, E. H. Ziyati, and M. Ouzzif, "An hybrid approach to improve part of speech tagging system," *2018 Int. Conf. Intell. Syst. Comput. Vision, ISCV 2018*, vol. 2018-May, pp. 1–6, 2018, doi: 10.1109/ISACV.2018.8354032.

[12] M. Labidi, "New Combined Method to Improve Arabic POS Tagging," *J. Auton. Intell.*, vol. 1,

no. 2, p. 23, 2019, doi: 10.32629/jai.v1i2.30.

[13] J. Yousif, "Hidden Markov Model Tagger for Applications Based Arabic Text: A Review," *SSRN Electron. J.*, vol. 7, no. 1, pp. 466–475, 2019, doi: 10.2139/ssrn.3451440.

[14] K. Alrajhi and M. A. ELAffendi, "Automatic Arabic part-of-speech tagging: Deep learning neural LSTM versus Word2Vec," *Int. J. Comput. Digit. Syst.*, vol. 8, no. 3, pp. 308–315, 2019, doi: 10.12785/ijcds/080310.

[15] W. AlKhwiter and N. Al-Twairesh, "Part-of-speech tagging for Arabic tweets using CRF and Bi-LSTM," *Comput. Speech Lang.*, vol. 65, p. 101138, 2021, doi: https://doi.org/10.1016 /j.csl.2020.101138.

[16] Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A. Smith, "Part-of-Speech Tagging for Twitter: Annotation, Features, and Experiments," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics, HLT '11*, 2011, vol. 2, pp. 42–47.

[17] K. Darwish ,Hamdy Mubarak, Mohamed Eldesouki, Ahmed Abdelali, Younes Samihy, Randah Alharbiz, Mohammed Attia, Walid Magdyz, Laura Kallmeyer,, "Multi-Dialect Arabic POS Tagging: A CRF Approach," 2018.

[18] Younes Samih, Mohamed Eldesouki, Mohammed Attia, Kareem Darwish,

[19] Ahmed Abdelali, Hamdy Mubarak, and Laura Kallmeyer,, "Learning from Relatives: Unified Dialectal Arabic Segmentation," 2017, pp. 432–441, doi: 10.18653/v1/K17-1043.

[20] A. H. Aliwy, "Arabic Morphosyntactic Raw Text Part of Speech Tagging System," no. January, pp. 1–139, 2013, [Online]. Available: http://portal.mimuw.edu.pl/wiadomosci /aktualnosci/doktoraty/pliki/ahmed_hussein_aliwy/aa-dok.pdf.

[21] M. Alian and A. Awajan, *Arabic tag sets: Review*, vol. 868. Springer International Publishing, 2018.

[22] K. Dukes and N. Habash, "Morphological annotation of quranic Arabic," *Proc. 7th Int. Conf. Lang. Resour. Eval. Lr. 2010*, pp. 2530–2536, 2010.