

A Systematic Literature Review of Hausa Natural Language Processing

Rufai Yusuf Zakari
Information Technology Department
I.T Dept. Skyline University Nigeria, SUN
Kano, Nigeria.
Email: rufai6 [AT] gmail.com

Zaharaddeen Karami Lawal
Information Technology Department
Federal University Dutse, Jigawa
Jigawa, Nigeria
Email: Deenklawal [AT] gmail.com

Idris Abdulmumin
Computer Science Department
Ahmadu Bello University, Zaria
Kaduna, Nigeria.
Email: iabdulmumin[AT] abu.edu.ng

Abstract— The processing of natural languages is an area of computer science that has gained growing attention recently. NLP helps computers recognize, in other words, the ways in which people use their language. NLP research, however, has been performed predominantly on languages with abundant quantities of annotated data, such as English, French, German and Arabic. While the Hausa Language is Africa's second most commonly used language, only a few studies have so far focused on Hausa Natural Language Processing (HNLP). In this research paper, using a keyword index and article title search, we present a systematic analysis of the current literature applicable to HNLP in the Google Scholar database from 2015 to June 2020. A very few research papers on HNLP research, especially in areas such as part-of-speech tagging (POS), Name Entity Recognition (NER), Words Embedding, Speech Recognition and Machine Translation, have just recently been released. This is due to the fact that for training intelligent models, NLP depends on a huge amount of human-annotated data. HNLP is now attracting researchers' attention after extensive research on NLP in English and other languages has been performed. The key objectives of this paper are to promote research, to define likely areas for future studies in the HNLP, and to assist in the creation of further examinations by researchers for relevant studies.

Keywords- Machine Learning, Hausa Language Processing, Natural Language Processing, Artificial Intelligence, Speech Recognition

I. INTRODUCTION

Natural language processing (NLP) has become one of the most competitive research areas in Artificial Intelligence through the vast output of user-generated content on social media, blogs and websites [1]. NLP is a field of computer science that seeks to establish concepts, discover strategies,

and create software that can understand, learn, and generate natural human languages through writing and speech to promote human interaction with computers [2]. NLP helps computers describe, in simple words, how humans use their language. When compared to other languages such as English, French, German, Chinese (Mandarin) and Arabic, Hausa NLP, thus referred to as HNLP, is a virgin area of study. This is because NLP relies on huge amount of human-annotated data for training intelligent models. HNLP is now attracting attention of researchers after significant research has been performed on English and other languages NLP. Several research papers on HNLP research, especially in areas such as part-of-speech tagging (POS), Name Entity Recognition (NER), Words Embedding, Speech Recognition and Machine Translation, have recently been released. A systematic analysis of current research on HNLP-based ML systems is given in this paper. This analysis focuses on the following:

- To review papers to understand the current state of science.
- Deliberate on the studied literature on the key problems affecting NLP in Hausa.
- Present the HNLP features and challenges.
- Show the complexity of the Hausa language to deliver brief related awareness of HNLP to new researchers in the area.

The definition of supervised ML techniques used to advance current HNLP-based applications/tools and recognize well-known techniques is discussed.

The remainder of this paper is composed as follows: - Section 2 provides the method used for the review. Section 3 provides HNLP Challenges. The contribution of ML algorithms in HNLP and precisely in HNLP applications are discussed in

section 4. Findings are demonstrated in Section 5. Section 6 presents the investigation and discussion of the current research to identify the research gaps and make recommendations for future research. At last, Section 7 draws up the conclusion

II. SYSTEMATIC REVIEW METHODOLOGY

The Systematic Literature Review (SLR) was developed as a tool for the discovery and evaluation of NLP in Hausa and other languages. Using a structured, unambiguous and laborious standard, a systematic literature review is accepted, aimed not only at reviewing up-to-date research on the subject, but also at providing an element of logical critique [1]. This SLR introduces four key phases that are central to being systematically thorough in every review [3]. We followed the guidelines of [3-5] in our selection and review process in this paper. The specifics are provided in the following subsections below.

A. Research questions

Detecting the study questions is the initial phase of a systematic review [6]. This phase has to be short and clear. In the context of this study, the research questions are itemized as follows:

RQ1: What are the key challenges in HNLP?

RQ2: What is the existing state of research in the field of HNLP? Who has published? And when?

RQ3: What are the most important gaps and limits in the reviewed studies?

RQ4: What are the ways of future research on HNLP?

B. Searching the Literature

The targeted approach of this Systematic review encompassed choosing resources, developing search strings, and the inclusion and exclusion conditions. The reviewed articles of this study are obtained through querying Google Scholar database using the following sets of keywords: (Hausa + Natural + Language + processing), (Hausa + Language + processing), (Hausa + NLP), and (Hausa + Natural + Language + resources). Using the above queries, the search yielded a total number of 31 publication from Google scholar database consisting of conference papers and journal articles. The articles returned are listed in Table 1. Using the same search strings, we queried journals plus conferences dealing with NLP which are indexed in the journal index in Scopus, such as Springer, Elsevier, and IGI Global, but we could not find any related to Hausa Language.

C. The Selection/Inclusion and Exclusion Criteria

The selected papers for this review are dated from 2015 to 2020. Attempting to give a quantitatively overview of HNLP literature since this period, we present in Table 1 the studied HNLP articles by Title, where it was published, year and Research area. Upon a visual inspection, we note that only a few papers were published and all articles were recently

published. The selection principles below were used to select the articles. Papers that didn't meet the selection criteria, were excluded.

The motive behind selecting this time period is that research in this area of HNLP is new and only a few papers have been published both in journals and conferences.

This paper covers only journal papers and conference articles on Google Scholar. Other publication forms (unpublished working papers, master's and doctoral dissertations, newspapers and books, etc.) were not included, due to the numerous articles having contents of the word "Hausa Language". The reason for choosing Google scholar database is that both technology practitioners and researchers from academic arena use it frequently to obtain knowledge and spread their study findings. However, there are some local journals which are not indexed on Google Scholar and may have publish related articles, but we could not access those journals because they are not publicly visible.

D. Data Collection

The information gathered from each article to perform the HNLP analysis was described as follows:

The source whether they are conference or journal

The title of the paper, publishing year, and publisher

The sort of HNLP tasks carried out

The dataset used

The approach/Method

This SLR bringing together approximately 29 articles. Figure 1 indicates the phases of the pre-screening process of the papers that culminated in the analysis being carried out. Second, a preliminary analysis was performed by looking for journals and conferences. The purpose of this preliminary study was to test the parameters of the hunt. 21 articles resulted in this phase. Secondly, by selecting the associated articles, the initial group of articles was reduced to 16. By reviewing titles and abstracts, this phase involved eliminating replicated papers and filtering unrelated articles. Finally, with full text reading, the filtering and selection of relevant articles was done, only 10 relevant articles were identified across the journals and conferences. The review of the selected final articles is shown in Table 2.

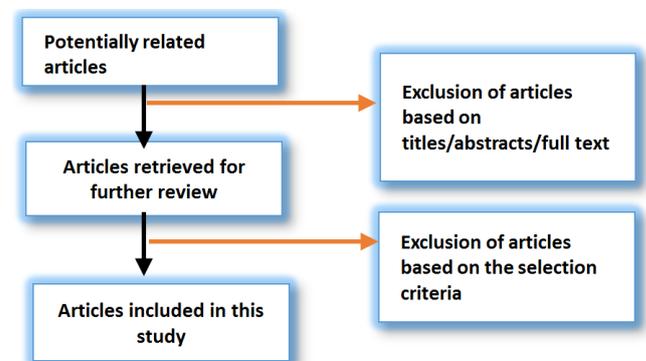


Figure 1. The procedures to select the articles.

Table 1. Summary of the relevant Literatures.

SN	Title	Published in	Year	Research Direction
1	houWE: Hausa Words Embedding for Natural Language Processing	International Conference of the IEEE Nigeria	2019	Natural Language Processing
2	Masakhane - machine translation for Africa	Published as a conference paper at ICLR 2020	2020	Machine Translation
3	HausaMT v1.0: Towards English–Hausa Neural Machine Translation Parts-of-Speech Tagging of Hausa-Based Texts Using Hidden Markov Model	Association for Computational Linguistics, ACL 2020	2020	Machine Translation
4	A Word Stemming Algorithm for Hausa Language	Dutse Journal of Pure and Applied Sciences (DUJOPAS)	2019	Parts-of-Speech Tagging
5	DNN-Based Text-To-Speech System for Hausa as an Under-Resourced Language Taxonomy, Review and Research Challenges	IOSR Journal of Computer Engineering (IOSR-JCE)	2015	Word Stemming
6	Automatic Hausa Language Text Summarization Based on Feature Extraction using Naïve Bayes Model For Hausa Text Parts of Speech Tagging, A Corpus Based Transformation-Based Learning	International Journal of Scientific & Engineering	2019	Speech recognition
7	Intelligent system speech recognition Voice and Speech Recognition for Hausa Words and Numerals	World Applied Sciences Journal	2017	Feature Extraction
8	Quantitative methods in African Linguistics - Predicting plurals in Hausa	International Journal of Computing and Digital Systems	2020	Parts-of-Speech Tagging
9		2 nd International Conference on recent innovation in management and Engineering	2017	Voice and Speech Recognition
10		ACAL 48, April 2017, Indiana, USA	2017	Natural Language

Table 2. Review of Selected papers.

Paper	Year	Datasets	Methods	Results
[7]	2019	Tanzil, News Websites	Bojanowski, hauWE CBOW & hauWE SG	houWE CBOW= 88.7%, hauWE SG=9.3%, Bojanowski= 22.3%
[8]	2020	JW300, Tanzil,	Standard word-level tokenization & Byte Pair	-

		Tatoeba & Wikimedia Freedom Radio & AfriHausa	Encoding (BPE)	
[9]	2019	Hausa online newspapers	Hidden Markov Model	Accuracy of 76.75%
[10]	2020	Hausa Leadership newspaper	Hidden Markov Model Tagging, N-Gram Model Tagging	Scored 64%, 52%, and 53%
[11]	2015	-	Porter’s algorithm(Modified)	Accuracy of 73.8%
[12]	2019	-	Hidden Markov Models (HMMs) and Deep Neural Networks (DNNs)	
[13]	2017	Aminiya and Leadership Hausa Newspapers	Naive Bayes model	Average F-score of 78%
[14]	2017	Bargery (1951) dictionary	Neural network	Accuracy : 0.5425 No Information Rate : 0.2082
[15]	2017	-	Mel-Frequency Cepstral Coefficients (MFCC) and Dynamic Time Warping (DTW). Hausa	-
[16]	2020	Hausa online newspapers	Hidden Markov Models (HMMs) and Deep Neural Networks (DNNs)	Scored 65%, 57% and 41%

III. HAUSA NATURAL LANGUAGE PROCESSING CHALLENGES

Voluminous natural language processing (NLP) work has been done on the key languages across the globe. Nevertheless, little to no research work has been done on over 1800 African languages (The little research work that has been done are on the foremost languages like Afrikaans, Zulu, Yoruba, Igbo, Hausa and Swahili [16] There is a large population with the right to use digital devices and the internet, as well as digital text, for languages such as Hausa and Yoruba, but the natural language processing (NLP) tools do not support them[17]. Despite the large population of native and non-native Hausa speakers, Hausa Natural Language Processing (HNLP) received very few publications when compared with other languages in developing countries. Hausa NLP is facing unprecedented challenges that need to be addressed by the researchers and technology practitioners. The Hausa Language lacks corpora and sentiment lexica. The dialects and morphology of Hausa language make it very difficult for Google Translate and Babylon to give exact or similar translation from other languages to Hausa (cite the source).

A. Hausa Morphology

African languages are of great dialectal difficulty and range, with diverse morphologies and phonologies, including lexical and grammatical tone shapes, and numerous are trained within

multilingual cultures [18].

Hausa like other African languages has a diverse morphology (cite the source). “Hausa has been used as a written language since the 19th century, when Islamic reformers began to write religious poetry and other religious texts using Arabic script (cite the source). A Roman orthography was developed by British colonial officials in Nigeria in the early twentieth century for use in official documents and schools” [19]. Hausa language is written in either Arabic Alphabet or Amended Latin Alphabet. Therefore, the scripts of Hausa Language can be divided into two. These are: - Ajami (Arabic Alphabet) and Boko (Modified Latin Alphabet). The figure below shows the two Hausa scripts [20].

The Modified Latin or Boko scripts used to write in Hausa language was made an official Hausa Script in 1930 by the colonial masters [20]. From early 1950s to date, Boko scripts has been adopted as the main Hausa alphabets, and Ajami script are mostly used in Islamic schools and school of languages of tertiary institutions that offer Hausa language studies [19]. Ajami script has become obsolete among Hausa native and non-native speakers in their writings [18]. The table below shows the Modified Latin Alphabets or Boko scripts alone.

(a) Ajami		(b) Boko	
ب [b]	م [m]	A a [a], [æ]	M m [m]
ب̣ [ḅ]	ن [n]	B b [b]	N n [n]
ث [tʃ]	ر [r], [r]	Ɓ ɓ [ɓ]	O o [o]
د [d]	س [s]	C c [tʃ]	R r [r], [r]
ط [ɗ]	ش [ʃ]	D d [d]	S s [s]
ف [β]	ت [t]	Ɗ ɗ [ɗ]	Sh sh [ʃ]
غ [g]	ظ [tsʼ]	F f [β]	T t [t]
ه [h]	و [w]	G g [g]	Ts ts [tsʼ]
ج [dʒ]	ي [j]	H h [h]	U u [u], [u:]
ك [k]	ع [ʔ]	I i [i]	W w [w]
ق [kʼ]	ز [z]	J j [dʒ]	Y y [j]
ل [l]		K k [k]	Y y [ʔ]
		Ƙ ƙ [kʼ]	Z z [z]
		L l [l]	ʼ [ʔ]

Figure 2. Writing Systems for Hausa Language [20].

Table 3. Modified Latin Alphabets or Boko scripts.

Hausa Alphabet	English Sound	Pronunciation Example
A a	/a/	short ‘a’ as in rat, long ‘a’ as in father
B b	/b/	‘b’ as in bed
Ɓ ɓ	/ɓ/	an implosive ‘b’
C c	/tʃ/	‘ch’ as in child
D d	/d/	‘d’ as in dog
Ɗ ɗ	/ɗ/	representing an implosive ‘d’
E e	/e/	short ‘e’ as in get, long ‘a’ as in that
F f	/f/	‘f’ as in fat
G g	/g/	‘g’ as in go
H h	/h/	‘h’ as in hat
I i	/i/	short ‘i’ as in hit, long ‘e’ as in elephant
J j	/(d)ʒ/	‘j’ as in joke
K k	/k/	‘k’ as in kit
Ƙ ƙ	/kʼ/	an adjective ‘k’

L l	/l/	‘l’ as in leg
M m	/m/	‘m’ as in man
N n	/n/	‘n’ as in nice
O o	/o/	short ‘o’ as in hot, long ‘oo’ as in tool
R r	/r/, /r̄/	flapped r’ as in Scots English
R ‘r	/r/, /r̄/	flapped r’ or ‘trilled r’ as in Scots English
S s	/s/	‘s’ as in smile
Sh	/ʃ/	‘sh’ as in shout
T t	/t/	‘t’ as in ten
Ts ts	/(t)s’/	‘t’ as in tutor
U u	/u/	short ‘u’ as in put, long ‘oo’ as in look
W w	/w/	‘w’ as in win
Y y	/j/	‘y’ as in you
(Y y)	/ʔ/	y’ as in yes
Z z	/z/	‘z’ as in zebra
ʼ	/ʔ/	glottal stop

Some of the Hausa words that contain letters like Ɓ ɓ, Ɗ ɗ, Ƙ ƙ, or Y y are difficult to be translated using NLP. Although effort had been made in producing Hausa Keyboard for computer users and virtual keypad for smart phone users respectively [20]. Word processors like Microsoft also includes Hausa Alphabet. The aforementioned letters are sometimes interchange with alphabets like ‘B ‘b, ‘D d’, ‘K ‘k’, and ‘Y ‘y’ or even with Bb, Dd, Kk, and Yy respectively. When writing for translation how will the machine knows that the writer is using ‘B or B in place of Ɓ. Some writers are now using Qq instead of Ƙ ƙ. The translation of Hausa to English on social media platforms is inaccurate in many instances. For example, the two Hausa words Bawo (shell) and Bawo (bowl) are distinct in their meanings, but when writing someone may write ‘Bawo or Bawo to mean Ɓawo. The same with the word Karami (Small) to be written as Karami or Qarami, and these instances are many in Hausa writings. Most of the Hausa writers and bloggers on social media platforms and even on Hausa news websites are not using the aforementioned four Hausa letters regularly. This will result in poor translation that may be far from the meaning of the word, sentence or even an entire paragraph. Goggle Translate has been improved in its Hausa translation recently and some instances similar to the ones that have been given as an example can be handled.

B. Lack of Corpora

Despite the availability of electronic resources, African languages have received relatively little attention in the corpus linguistics [14]. The inadequacy of labeled training data has limited the development of natural language processing tools for several languages spoken in developing countries, such as named entity recognition [17]. In several low-resourced languages, they have, at least, a Wikipedia corpus [16], but it’s not the same case with Hausa Language. As of now, there is no any available Hausa dataset(s) on GitHub. According to [7] The only openly available dataset for machine translation is the Tanzil dataset, consisting of a meager 127k parallel sentences (updated literature like that of [8] will show you that there are more parallel data available for MT). The dataset is

also made of less than 10k unique Hausa sentences [8]. This dataset also does not signify the various dialects of Hausa language [14]. Apart from the Tanzil dataset and other publicly available data, Web Crawled articles from the Hausa local and international news websites has been a source of monolingual corpora for conducting HNLP research. The richness of the language data set will encourage and help HNLP researchers to conduct more work on it which in turn will also help the technology practitioners to improve the language translation tools. There is need for sufficient datasets that will be available to public. This will pave a way for more many researchers and make their work easier.

C. Use of dialectal Hausa

Dialect refers to as district diversity of language with variances in lexis, grammars, and pronunciation as well [19]. In alternative word, dialect is a kind of dissimilarity that exists in the practice of a particular language from one part of the district to the other and that they shares the same language [21]. The early Hausa writers observed several dialect variation; they never concerned to pay serious attention to it let alone endeavored to write a grammar for each dialect nor did they attempt to point out the differences in any detail [22]. Researchers and industry experts understand that the major Hausa dialects can be categorized into the dialects of Eastern, Western, Northern, and Southern Hausa [21]. Kananci, Dauranci, Gudduranci, Bausanci, and Hadejanci, among others, are said to be Eastern dialects. Kutebanci, Katsinanci, Arewanci, Sakkwatanci, and Kurhwayanci are among the western Hausa dialects. Katsina is the intermediary between the dialects of the East and the West [19]. For Arawa, the northern Hausa dialects are Arewanci. The major Southern dialect is Zazzaganci in Zaria [21].

Within the Eastern dialect, standard Hausa is created. It's based on Kananci the dialect of Kano, the largest Hausa Commercial center located in Northwestern part of Nigeria [22]. The standard Hausa has been adopted as the standard for writing language as contained in books and newspapers and also for broadcast in Radio and Television. The Multilingual Radio stations like BBC, DW, RFI, VOA, Radio Beijing, Russia and Tehran also uses Kananci dialect to broadcast their programs in Hausa[21]. Despite the adaptation of Kananci dialect as the Standard Hausa, the speakers of other dialects are still using it in their formal and informal writings [19]. The use of social media platforms gave them an independent to write in their own dialect. Since this is the case, two people may write a simple sentence that has the same meaning using their respective dialects. The issue that may arise is that how machine will translate the two sentences that mean the same using two distinct dialects. For example, the English sentence. They have touched their faces can be translated as Sun taba

fuskokin su in Kananci dialect and Sun taba huskokin su” in Katsinanci dialect. Another example is “I am Coming” in Kananci “Ina Zuwa” and Katsinanci is “Ina Zowwa”. The diversity of Hausa dialects is making HNLP a bit tough.

IV. HOW HNLP IS EVOLVE IN MACHINE LEARNING

The HNLP journal articles we have described before use machine learning methods (predominantly supervised) and have recently attracted the attention of researchers. The development of these research often includes numerous phases as shown in Figure 3 and presented in detail in the following subsections.

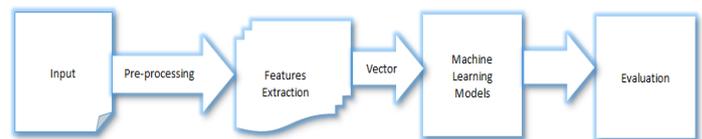


Figure 3. The NLP development phases.

A. Corpora

Data gathering is the first stage in HNLP-based machine learning work. There are few freely available Hausa corpora, such as the Tanzil dataset, which is a multilingual text aimed at producing a highly verified multi-text Quran text [8] and JW300 dataset, as well as a parallel corpus of more than 300 languages with an average of approximately 100,000 parallel phrases for each language pair on average among which is Hausa. These corpora have been used in Hausa processing studies, especially in text Machine Translation and words embedding. Data is not readily available for other NLP tasks, such as text classification, sentiment analysis, question answering, text summarization and others.

B. Pre-processing

Data pre-processing is an essential step to clean the text by removing unwanted data such as removal of preposition, pronouns, conjunctions words, auxiliary verbs, digits, numbers, stop words and formatting tags [23]. Preprocessing method plays a very important role in Natural Language and applications. It is the first phase in the text mining process [24].

C. Feature Selection

Feature extraction is a critical task in the domain of Natural Language Processing that aims to clean text documents, includes stemming and presenting text in a clear format [23].

V. FINDINGS

This segment discusses HNLP's results and contributes to answering the research questions above. The findings obtained from our analysis are applicable to this section below. During the period from 2015 until the end of June 2020, a total of 29 papers were found. The number of HNLP study studies has certainly increased in the last three years from Figure 4 below;

i.e., in both 2019 and 2020, it reached 3 papers each. In both 2016 and 2018, there were no posts, with only one paper in 2015. It is clear that HNLP has recently become a hot research subject. Henceforth, researchers will observe a rapid growth and an increased interest. It is now, therefore, an attractive research area.

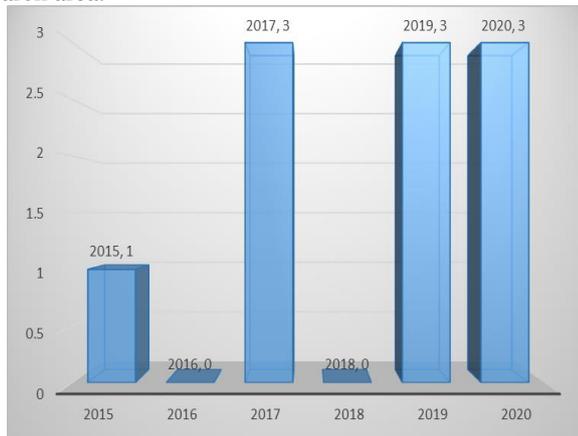


Figure 4. Article from 2015 until the end of June 2020

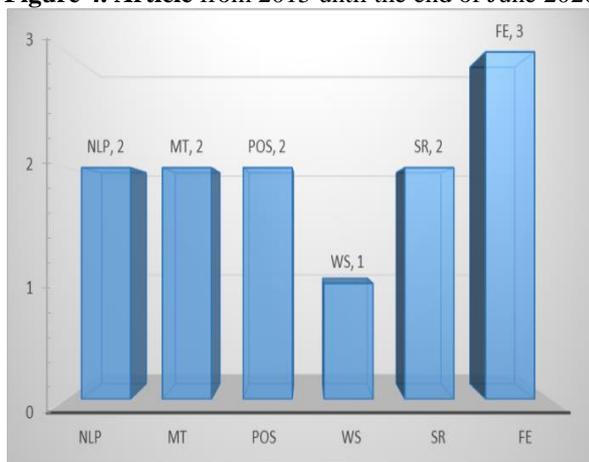


Figure 5. Most of the task areas used in HNLP

In addition, Figure 6 provides evidence for most of the task areas used in HNLP. As shown in Figure 5 below, the tasks are broken down by returning to the examined studies. In addition, it is evident from the figure that, relative to the other tasks, the task of Feature Extraction (FE) is the most common task. Indeed, the language of Hausa still lacks tools and instruments that can be used to support these tasks.

ML-based methodologies for HNLP typically include a number of stages in the literature, including: data preprocessing, generation of features, word embedding, selection, and methods of ML.

Hausa is an ironic language that has an inspiring linguistic field and needs forward thinking with morphological complexities and dialectal diversity. Another complexity of Hausa Language reflects the feeling in the dialect.

The preprocessing and analysis of raw Hausa texts greatly moderates noise and improves performance. A successful preprocessing results in a suitable function being selected. A

semantic symbol is included in the function representation, which in NLP is still a difficult task.

The outcomes of the HNLP tasks can be enhanced by integrating word embedding with a mixture of N-gram models. Most of this studies went through the preprocessing phases, which involve text washing, normalization, tokenization, stop words elimination, and stemming. In conclusion, little work has been done with regards to HNLP in virtually every domain. However, there was no intended background from researchers to study the domain in particular. Besides, the collected literature was small as the findings shows, this is as a result of lack of research towards the direction of HNLP.

VI. DISCUSSION AND FUTURE RESEARCH

In this section, we examine the gained outcomes from SLR and to offer the answer to the third and fourth research questions. 29 articles have been reviewed in HNLP to capture the current state and accomplish the research aims. The goals were brief and the most active methods used in HNLP, close-fitting the gaps and limitations in the reviewed studies, and highlighting the directions of future research on HNLP. As presented in the literature, there are still some challenges in HNLP that have to be solved. These challenges varies into preprocessing strategies, feature selection, and the targeted domain. The sentences in Hausa Language are seriously involve noisy, missing, and unreliable data that need to be preprocessed. Not utilizing preprocessing, for example, wiping out inconsequential remarks and redundancy of letters, may lead pay no attention to significant words? Applying a wide-range set of preprocessing approaches, such as the normalization, tokenization, word removal, and stemming, will improve HNLP.

6.1 Implications for Future Research

As SLR concentrated on contributions from current HNLP-related literature, the following are almost insinuations for future research:

Applying deep learning methods to HNLP

To achieve the appropriate processes that comply with the characteristics of the Hausa language, a widespread prototype must be developed that expresses all the details of the preprocessing process in different situations.

It is an open area to develop the Hausa language vocabulary. Many scientists have developed several lexicons that are either partial or not freely accessible in size. As a consequence, as there are very few publicly available Hausa corpora and HNLP lexicons, focused lexicons should be designed for each domain.

The majority of current Hausa language demonstrations are borrowed from other languages, such as Arabic and English. For that reason, it will help to improve the outcomes by developing new feature representation that matches Hausa language characteristics.

VII. CONCLUSION

Based on the works published in the HNLP, the systematic

review was structured. In this report, we review the literature and principles and implementations and analyze them using dimensions linked to current and emerging HNLP issues. This paper is based on an HNLP literature review from 2015 to 2020 using a search on Google Scholar for a keyword index and article title. Generally speaking, we have found that HNLP-related behavior tends to be a few and, as a result, provides some literature for prospective researchers in the field of HNLP. We hope that this study will encourage researchers, businesses, and experts to explore the problems and applications of HNLP. It is clear that HNLP still needs more and more comprehensive research through this SLR. The study shows that there is insufficient research on the creation of structured datasets and the implementation of auspicious algorithms. In addition, the analysis also shows a lack of study to create a new feature representation that matches the characteristics of the Hausa Language. In addition, in creating reference structures in many fields and enriching the work for HNLP in various domains, prospects for future research also exist. Academics and companies are invited to enter this new research sector.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

REFERENCES

- [1] O. Oueslati, E. Cambria, M. Ben, and H. Ounelli, "A review of sentiment analysis research in Arabic language," *Futur. Gener. Comput. Syst.*, vol. 112, pp. 408–430, 2020, doi: 10.1016/j.future.2020.05.034.
- [2] S. L. Marie-sainte, N. Alalyani, S. Alotaibi, S. Ghoulali, and I. Abunadi, "Arabic Natural Language Processing and Machine Learning-Based Systems," *IEEE Access*, vol. 7, pp. 7011–7020, 2019, doi: 10.1109/ACCESS.2018.2890076.
- [3] C. Okoli and K. Schabram, "Working Papers on Information Systems A Guide to Conducting a Systematic Literature Review of Information Systems Research," vol. 10, no. 2010.
- [4] S. E. Group, "Guidelines for performing Systematic Literature Reviews in Software Engineering," 2007.
- [5] D. Tranfield, D. Denyer, and P. Smart, "Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review *," vol. 14, 2003.
- [6] A. Ghallab, A. Mohsen, and Y. Ali, "Arabic Sentiment Analysis : A Systematic Literature Review," vol. 2020, 2020.
- [7] I. Abdulmumin, "hauWE: Hausa Words Embedding for Natural Language Processing."
- [8] A. Akinfaderin and D. Duality, "HausaMT v1.0: Towards English–Hausa Neural Machine Translation," pp. 2017–2020, 2020.
- [9] A. Tukur, K. Umar, and A. Sa, "Parts-of-Speech Tagging of Hausa-Based Texts Using Hidden Markov Model," vol. 6, no. 2, pp. 303–313, 2020.
- [10] J. Awwalu, S. E. Abdullahi, and A. E. Ewwiekpaefe, "A Corpus Based Transformation-Based Learning for Hausa Text Parts of Speech Tagging."
- [11] M. Bashir, A. B. Rozaimée, W. Malini, and B. Wan, "A Word Stemming Algorithm for Hausa Language," no. June, 2015, doi: 10.9790/0661-17362531.
- [12] A. A. Aliero, D. Muhammed, and A. Ibrahim, "Taxonomy, Review and Research Challenges Of DNN-Based Text-To-Speech System for Hausa as Under-Resourced Language," vol. 10, no. 7, pp. 548–560, 2019.
- [13] M. Bashir, A. Rozaimée, W. Malini, and W. Isa, "Automatic Hausa LanguageText Summarization Based on Feature Extraction using Naïve Bayes Model," vol. 35, no. 9, pp. 2074–2080, 2017, doi: 10.5829/idosi.wasj.2017.2074.2080.
- [14] M. G. Naranjo and L. Becker, "Quantitative methods in African Linguistics - Predicting plurals in Hausa," no. April, p. 2017, 2017.
- [15] A. T. Hamid and S. M. Tahir, "Intelligent system speech recognition Voice and Speech Recognition for Hausa Words and Numerals," pp. 133–146.
- [16] K. Ogueji, "PidginUNMT: Unsupervised Neural Machine Translation from West African Pidgin to English," pp. 0–5, 2019.
- [17] D. I. Adelani, M. A. Hedderich, D. Zhu, D. Klakow, and S. I. Campus, "D INSTANT S UPERVISION AND N OISY L ABEL L EARNING FOR L OW R ESOURCE N AMED E NTITY R ECOGNITION :," pp. 1–9, 2020.
- [18] I. Orife *et al.*, "Masakhane -- Machine Translation For Africa," pp. 1–4, 2020, [Online]. Available: <http://arxiv.org/abs/2003.11529>.
- [19] R. G. Schuh, "No Title."
- [20] L. Salifou, "Design of A Spell Corrector For Hausa Language," pp. 14–26, 2014.
- [21] A. S. Gulbi and U. Ahmed, "Languages and National Identity: Relevance of Dialect in Hausa Regional Identity," vol. 8, no. 9, pp. 211–215, 2018, doi: 10.29322/IJSRP.8.9.2018.p8128.
- [22] M. Approach, "Hausa Substantives : A Natural Semantic," vol. 1, no. 1, pp. 51–62, 2020.
- [23] A. H. Mohammad, O. Al-momani, and T. Alwada, "Arabic Text Categorization using k-nearest neighbour , Decision Trees (C4 . 5) and Rocchio Classifier : A Comparative Study," vol. 6, no. 2, pp. 477–482, 2016.
- [24] S. Vijayarani, M. J. Ilamathi, and M. Nithya, "Preprocessing Techniques for Text Mining - An Overview," vol. 5, no. 1, pp. 7–16.
- [25] Z. F. Hussain, H. R. Ibraheem, M. Alsjari, and A. H. Ali, "A new model for iris data set classification based on linear support vector machine parameter ' s optimization," vol. 10, no. 1, pp. 1079–1084, 2020, doi: 10.11591/ijec.v10i1.pp1079-1084.
- [26] A. M. F. Al Sbou, "A survey of Arabic text classification models," vol. 8, no. 1, pp. 25–28, 2019, doi: 10.11591/ijict.v8i1.pp25-28.
- [27] D. Khurana, A. Koli, K. Khatter, S. Singh, and M. Rachna, "Natural Language Processing : State of The Art , Current Trends and Challenges Department of Computer Science and Engineering Accendere Knowledge Management Services Pvt . Ltd ., India Abstract," no. Figure 1
- [28] H. H. Patel and P. Prajapati, "Study and Analysis of Decision Tree Based Classification Algorithms," pp. 74–78, 2018.
- [29] S. B. Imandoust and M. Bolandraftar, "Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background," vol. 3, no. 5, pp. 605–610, 2013