Al-Noor Journal for Humanities, Septembre (2024); 3 (3): 283- 307
DOI: https://doi.org/10.69513/jnfh.v2n3.en12
IRAQI Academic Scientific Journals
ISSN: 3005-5091
AL-NOOR JOURNAL FOR HUMANITIES
Available online at : http://www.jnfh.alnoor.edu.iq
JNFH Al-Noor Journal for Humanities

Artificial Intelligence (AI) Challenges and Opportunities in Translation : An African Experience

Dr. Ahmed Mohammed Bedu,
Department of Languages and Linguistics
University of Maiduguri,
P.M.B. 1069 Maiduguri,
Borno State
Nigeria.
ahmedbedu@unimaid.edu.ng
+2348039542870

Abstract

The Artificial Intelligence (AI) revolution has become a reality in today's world and its importance for linguistics was recognized very early. Despite its unprecedent surge and integration into various academic fields including language teaching and translation, surprisingly, little work has been done by scholars in advancing discussions on the profound impact of the AI on the diversity of widely available languages in both developed and developing world.

Africa is linguistically diverse continent with about one third of the world's languages that are vastly underrepresented in the

global digital data pool. AI translation machine is supported in only 25 languages out of over 2000 languages in the continent. The paper deploys homomorphism model of AI theory to interrogate the natural language data drawn the African languages to present the current and future challenges, opportunities and potential for developing AI algorithms that could fit neatly into the translation of the African languages. Most of the discussions in the paper focuses on the seven patterns of the AI, the usage and implementation of AI algorithms in the translation science. The research findings show some of the complexities of the African languages in which their syntactic categories have multiple corresponding semantic objects. Unlike English, the findings also reveal that syntactic operation in the African languages do not always have one corresponding semantic operation as postulated by the homomorphism model of AI theory. the study contributes to scholarly literature by stressing the limits and opportunities that relate to using AI in translation science and supplying input from NLP algorithms practitioners to expand the AI applicability operation in the translation science.

Keywords : Artificial Intelligence (AI), translation, African languages, homomorphism, NLP algorithms

التحديات والفرص التي تواجه الذكاء الاصطناعي (AI) في الترجمة: تجربة أفريقية

الدكتور أحمد محمد بدو
قسم اللغات واللسانيات
جامعة مايدوغوري
P.M.B. 1069مايدوغوري، ولاية بورنو ، نيجيريا
ahmedbedu@unimaid.edu.ng
+2348039542870

الملخص
أصـــبحت ثورة الذكاء الاصـــطناعي (AI) واقعاً في عالم اليوم، وتم الاعتراف بأهميته في مجال اللســـانيات في وقت مبكر. على الرغم من الانتشـــار غير المســـبوق والتكامل في مجالات أكاديمية مختلفة، بما في ذلك تعليم اللغات والترجمة، إلا أنه من المدهش أن هناك عملاً قليلاً قد تم من قبل

الباحثين لتوسيع المناقشات حول التأثير العميق للذكاء الاصطناعي على تنوع اللغات المتاحة في كل من العالم المتقدم والنامي.
تُعد إفريقيا قارة متنوعة لغويًا، حيث تضم حوالي ثلث لغات العالم، والتي تمثل جزءًا صغيرًا في البيانات الرقمية العالمية. تدعم آلات الترجمة بالذكاء الاصطناعي 25 لغة فقط من بين أكثر من 2000 لغة في القارة. تستخدم الورقة نموذج التماثل في نظرية الذكاء الاصطناعي لاستجواب بيانات اللغة الطبيعية المستمدة من اللغات الأفريقية لعرض التحديات الحالية والمستقبلية، والفرص والإمكانات لتطوير خوارزميات الذكاء الاصطناعي التي قد تتناسب مع ترجمة اللغات الأفريقية. تركز معظم المناقشات في الورقة على الأنماط السبعة للذكاء الاصطناعي، واستخدام وتنفيذ خوارزميات الذكاء الاصطناعي في علم الترجمة. تظهر نتائج البحث بعض التعقيدات في اللغات الأفريقية، حيث أن فئاتها التركيبية لديها العديد من الكائنات الدلالية المقابلة. على عكس اللغة الإنجليزية، تكشف النتائج أيضًا أن العمليات التركيبية في اللغات الأفريقية لا تتوافق دائمًا مع عملية دلالية واحدة كما يفترض نموذج التماثل في نظرية الذكاء الاصطناعي. تساهم الدراسة في الأدب الأكاديمي من خلال التأكيد على الحدود والفرص المتعلقة باستخدام الذكاء الاصطناعي في علم الترجمة وتقديم مدخلات من ممارسي خوارزميات معالجة اللغات الطبيعية (NLP) لتوسيع قابلية تطبيق الذكاء الاصطناعي في علم الترجمة.
الكلمات المفتاحية :الذكاء الاصطناعي(AI) ، الترجمة، اللغات الأفريقية، التماثل، خوارزميات معالجة اللغات الطبيعية(NLP)

Introduction

The Artificial Intelligence (AI) refers to the ability of the computer or machine to process all forms of data including human languages to reach results that are similar to human thinking, learning and decision making. In fact, the Artificial Intelligence (AI) has brought a revolution and reality in today's world that their importance for linguistic science was recognized very early (Hauseer, 1989). The unprecedent surge of AI and its integration into various aspects of language teaching and translation triggers the development of many applications such as speech recognition, concordance software, neural machine translation among others that aid many linguistic researches (Slocum,1984). While scholars are focused on using AI to meet challenges of developed languages (e.g. Arabic, Chinese, English Russian, Turkish), the incorporation of the undeveloped languages in Africa into of AI space in Africa and developing their algorithms have largely been ignored. To ensure that African languages benefit from the attendant gains of AI, the natural

language programme NLP algorithms of AI need to be robustly considered and accommodated African languages.

It is not an exaggeration that no serious research investment was made on harnessing the potentialities of African languages to add value to the AI translation capabilities for achieving greater efficiency especially in translation research. In recent years for instance, many major technology companies have recognized the value of languages as well as the increasing demand for localized their content and services and yet African languages were lagging behind in enjoying this patronage of the giant tech companies. For instance, Google actively supports Natural Language Processing (NLP) – an AI platform that enables computers to understand, interpret, and generate human language for research on language data with an aim at making online content more accessible and inclusive but only 25 of the 2000 languages in the continent that are presently recognized in its digital datasets.

The goal of AI in all academic fields is to develop application systems to easily deal with complicated problems in a similar way to human operations. However, the field of translation is one of the various academic fields that still defies the prowess of the AI from 1954 to date, when the first machine translation (MT), was invented at Georgetown IBM experiment and become operational in 1964 (Slocum 1985; Winder 2023).

As explained by Newmark (1988), interlingual translation which involves rendering the meaning of a text into another language in the similar way that the author's intended message produced in source language; the transformation processes of the same message into receptor language have continued to be one of the herculean tasks for AI driven-translation. In this regard, this paper presents not only the hurdles and the excited prospects that AI brings to

translation science but showcases the huge potentialities of African languages in solving many of the puzzles that limit the AI effectiveness in the translation science.

African Languages and AI Translation

Reader (1998) says The first thing that impresses anyone considering the Africa continent as a whole is the incredible diversity of its linguistic landscape. Africa boasted a repertoire of more than over 2000 living languages that constitute one third of the world's living languages and each of the languages has unique grammar, vocabularies and linguistic resources (Diamond, 1997; Wolff, 2019). With over 520 languages, Nigeria accounted for around fourth of the total languages spoken in Africa among which only few of these languages are developed with standard orthography, grammar books and dictionaries (Blench 1998). Cameroon and the Democratic Republic of Congo followed, each with over 200 living languages with three-quarter of them are undeveloped. This makes Africa as the most linguistically unexploited continent with about one third of the world's languages that are vastly underrepresented in the global digital data pool. While Africa is home to one third of the world's languages, AI is also not yet available for many of its languages due to the scarcity of their developed literature, even though their linguistic diversity is truly remarkable and accessible.

As million people speak these 2,000 living languages in Africa – roughly one third of all languages spoken in the world, there are no serious efforts to preserve this rich linguistic and cultural heritage in the digital space to help in making these languages relevant in the AI applications and operations. Progress in AI tools, which are now capable of handling extensive datasets and enabling software

to interact in different languages, has unleashed new market possibilities.

At the present, AI translation machine is supported in only 25 languages out of over 2000 languages in the continent. These include Afrikaans, Amharic, Arabic, Bambara, Chichewa**,** Ewe, Hausa, Igbo, Kinyarwanda, Krio, Lingala, Luganda, Malagasy, Oromo, Sepedi, Swahili, Sesotho, Shona, Somali, Tigrinya, Tsonga, Twi, Xhosa, Yoruba, and Zulu. Though, there are several other languages that are spoken across borders like Kanuri, Fulfulde, Masai Nama among others that are unattended.

Right now, it is only twelve African languages are available on the Google Translate app on iOS and Android, including Hausa, Yoruba and Igbo, three of West Africa's most spoken languages. The web browser firm Mozilla has recently incorporated Ghana's Twi language into its open-source linguistic repository, Common Voice, which collects input from real-life language speakers. This part of the initiative aims to improve speech recognition technology and to promote a broader range of local languages on the internet, challenging the dominance of European languages as the main – or sole – online communication method. Indeed, embedding linguistic cultural diversity of African languages into the development of universal AI algorithms will facilitate the contextual understanding of reality of the AI and improve acceptability in the translation. AI development and use in Africa needs to be sensitive to African languages, their cultural values and beliefs which are currently lacking in the global discussion on AI development and its operations.

This important challenge is what this paper wants to advance for linguists to tackle because every language, regardless of its population size, is important to human communication (Bedu,

2022). In fact, the accessible mean of easing the translation of these languages in Africa which is the core goal of AI translation machine, can facilitate unhindered connection of African languages with the world around us since we speak different languages. If we are committed to make sure all people, regardless of their sociolinguistic status can understand and be understood, it's a significant and technical challenge for the modern linguistics to strive and make the dream of AI unifying the world a reality.

In AI development, European languages along with other languages in Asian continent are considered to be favorable for the AI model towards the development of Universal **AI Grammar**. The current perception of non-African linguists is that many of the African languages are lexically poor to enrich the AI model to attain the goal of **Universal AI Grammar.** In this paper, the study intends to galvanize academics to take a step further in making the African languages relevance to the goal-oriented AI translation systems and other digital spaces.

Literature Review

The potential and challenge of, as well as prospects for human languages including African languages with regard to AI and translation science are not a subject about which scholars feel neutral (Arakpogu et. al. 2021; Khalati and Al-Romany 2020;). Artificial Intelligence (AI) has many applications such as speech recognition, cognitive automation, translation machine and application for predictive analytic that attracted the attention of numerous scholars across different academic fields. Regardless of how AI is applied each of these applications has something in common which is known as seven patterns of AI as shown in figure (1) below:

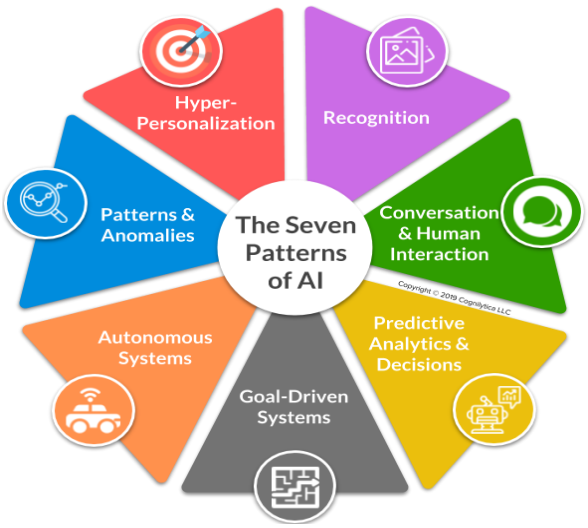

Figure 1: Seven patterns of AI (adopted from Bin Rashid, A. et. al. 2023:2)

From the figure (1) , the seven patterns of artificial intelligence are goal-driven systems, autonomous systems, conversational/human interactions, predictive analytics, hyperpersonalization, and decision-making support that some of these patterns are critical to language translation. Since these patterns of AI are what makes AI machine smart, the AI has, without fear of contradiction, contributed in revolutionizing translation science in the recent years especially by providing tools that automate not only certain aspects of the translation process but carrying out tasks relevant to human linguistics in Natural Language Processing (NLP), predictive analytics and other forms of decision-making.

However, it is well established across many literatures that AI techniques work with big data and the recent studies on AI and its related translation machine on African languages show that AI related translation machines and their applications are inadequate to provide accurate pragmatic interpretation of meaning related to an utterance context of African languages especially regarding to surface-meaning mapping of some cultural related words. As the

NLP makes the computer to understand the human language i.e., words, sentences, and paragraphs, for the analysis and synthesis of text, the lack of written grammar and documentation for most of the African languages makes the NLP inadequate to accommodate these languages in AI space. In view of this, many of scholars advocate for transforming African languages regardless of their sociolinguistic status into NPL so that AI especially its translation tool will effectively support the languages in the continent as stated thus:

> *"AI related machines rely on a field called natural language processing, a technology that enables computers to understand human languages. In fact, the NLP is the catalyst for the computers to master a language through training on how to pick up the patterns in speech and text data. However, AI applications can fail when data in a particular language is scarce in NLP algorithms, as presently happening to many African languages."*

Recognizing African languages in automatic machine translation will help to revolutionize communication between people around the world especially when there is NLP model for African languages. However, there is no comprehensive NLP model that captures African languages to significantly reverse the situation in which African languages' speakers are often left out of, because their languages are not yet taken into account in mainstream NLP research. This is the research gap that need to be filled so that AI translation machine will support and translate more languages accurately over and above the 25 recognized African languages that AI translation machine is inaccurately translating.

The Nigerian linguist, Aremu Adeola, presents an interesting example on why AI translation machine is failing to achieve accurate translation of many *African* languages especially Yoruba:

> *"Most translations done by machines render some words wrong, especially words that are culturally nuanced. For example, Yorùbá words ayaba and obabìnrin have their meanings situated in a cultural context. Most machines translate both words as 'queen.' However, from a traditional-cum-cultural vantage point, it is essential to note that the meanings of ayaba and obabìnrin are different: Ọbabìnrin means 'queen' in English while ayaba is 'wife of the king.'"*

The problem that Aremu highlights above is the input-output problem with regard to machine translation of African languages which demands the linguists to develop AI algorithms that will comprehensively focus beyond the area of synchronic syntax, semantics and pragmatics but to include cultural diversity of the African languages to explain the holistic functioning of all these language and culture resources in communication.

The highlighted input-output problem in the above is similar to what Roland Hausser (1989:25) highlights in his book *Computation of Language* when a speaker utters a meaningful sentence, and the hearer fails to/understand(s) it, can't be different to instances of AI translating languages inaccurately especially African language that majority of them are aliens to AI world. Because its language inputs are only available in high resource languages like English, French, Chinese etc. He then explains that input-output problem of AI is based on the non-verbal context is similar to the speaker-hearer context as explained below:

> The context of the speaker is defined as an internal representation of what he or she perceives and remembers at the moment (or time internal) of the utterance. Similarly, the context of the hearer is a representation of what he or she perceives and remembers at the moment of interpretation of the utterance (Hausser 1989:26).

Context problem as problem in AI translation is not one way interlingual form of translation as in African languages (input) to English (output), we can also envisage same input-output problem in English (input)-African languages (output) when AI translation machine confronts with ambiguous words at the different levels of a translation as you can see below:

i. Word senses: bank (finance or river?)
ii. Part of speech: chair (noun or verb?)
iii. Syntactic structure: I saw a man with a telescope
iv. Quantifier scope: Every child loves some movie
v. Multiple: I saw her duck

In integrating African languages into AI space, we therefore need to model the ambiguous words of the African languages as well as their cultural diversity by taking care of them in the AI algorithms in order provide their correct interpretation in context in the translation. Since NLP algorithms are designed for computer implementation the algorithms themselves and Homomorphism model of algorithms can help to handle African languages.

Homomorphism Model of AI Theory

Homomorphism is defined as the relationship between the level of the language and the level of the referents (Montague 1974). It has two levels of principle (both with input-output relationship) which explain the homomorphism relations between syntax and semantic

and at the same time explain universality of the different languages. These levels of principle are:

1. For each syntactic category there is a corresponding semantic object.
2. For each syntactic operation there is a corresponding semantic operation.

Let Σ_1 and Σ_2 be two alphabets. A function $h: \Sigma_1^* \rightarrow \Sigma_2^*$ is homomorphism if it is semigroup homomorphism from semigroup Σ_1^* to Σ_2^*. This means that

- h preserves the empty word: $h(\lambda) = \lambda$, and
- h preserves concatenation: $h(\alpha\beta) = h(\alpha)h(\beta)$ for any word $\alpha, \beta \in \Sigma_1^*$

Since the alphabet Σ_1 freely generates Σ_1^*, h is uniquely determined by its restriction to Σ_1. Conversely, any function from Σ_1 extend to a unique homomorphism from Σ_1^* to Σ_2^*. In other words, it enough to know what h (α) is for each symbol α in Σ_1. Since every word ɯ over Σ is just a concatenation of symbol in Σ, $h(w)$ can be computed using second condition above. The first condition takes care of the case when ɯ is the empty word.

Suppose $h: \Sigma_1^* \rightarrow \Sigma_2^*$ is homomorphism, $L_1 \subseteq \Sigma_1^*$ and, $L_2 \subseteq \Sigma_2^*$. Define

If L_1, L_2 belong to certain family of languages, one is often interested to know if $h(L_1)$ or $h^{-1}(L_2)$ belongs to that same family. We have the following result

1. If L_1 and L_2 are regular, so are $h(L_1)$ and $h^{-1}(L_2)$.
2. If L_1 and L_2 are context free, so are $h(L_1)$ and $h^{-1}(L_2)$.
3. If L_1 and L_2 are type -0, so are $h(L_1)$ and $h^{-1}(L_2)$.

However, the family **f** of context-sensitive languages is not closed under homomorphism, nor inverse homomorphism. Nevertheless, it can be shown that **f** is closed under a restricted class of

homomorphism, namely λ-free homomorphism. A homomorphism is said to be λ-free or non-erasing if $h(a) \neq \lambda$ for any $a \in \Sigma_1$.

Remarks

- Every homomorphism induces a substitution in a trivial way: if $h: \Sigma_1^* \rightarrow \Sigma_2^*$ is homomorphism, then $h_s(a) = \{h(a)\}$ is a substitution.
- One can likewise introduce the notion of antihomomorphism of languages. A map $g: \Sigma_1^* \rightarrow \Sigma_2^*$ is antihomomorphism if $g(\alpha\beta) = g(\beta)\, g(\alpha)$, for any word α, β over Σ_1^*.

  It is easy to see that is an antihomomorphism ***iff*** $g \circ rev$ is a homomorphism, where ○ is the reversal operator. Closure under antihomomorphisms for a family of languages follows the closure under homomorphisms, provided that the family is closed under reversal.

Source of Data

Simiyun et.al (2022) say African languages lack datasets and this discourages many NLP practitioners and AI researchers in the continent to start from scratch in developing both NLP and AI translation applications that will appreciate the potentialities of all living languages in the continent. The dearth of literature on AI and African languages couples with the low economic interest of the top companies that are driving changes in AI space contribute to lack of datasets for African languages. In view of this, the present study utilizes open access source to generate data from various secondary source of data to discuss the linguistic nature of the African and their resources to enrich AI learning and research with regards to the translation science. The data are restricted to three different language family groups with one language each from the language

phyla. These languages are Hausa (Afroasiatic), Kanuri (Nilo-Sahara) and Fulfulde from Niger-Congo language family.

All of the three languages spoken by million speakers either as first or second language across different international boundaries in the West and Central Africa.

Complexity of the Linguistic Typology of African Languages to Translation

Unlike European languages that generally exhibit uniform Subject-Verb-Object word-order in their sentence construction, African languages vary in both syntactic word-order and morphological structures (Bedu, 2010). The complexity of their linguistic typology especially syntactic word-order are not only crucial but important factors for consideration in developing comprehensive AI model that will accommodate African languages for achieving accurate translation of these languages by the AI and other digital machines. Consider the following table (1) below:

Table (1): Syntactic word-order of African languages

Language	**Sentence Structure**
Hausa	Subject Verb Object Ali ya-a ci abinci Ali PRVPrn-Past eat food 'Ali ate the food' Abinci ne Ali ya-a ci Food Foc Ali PRVPrn-Past eat
Kanuri	Subject Object Verb Musa-ye kasuwu-ro lewono Musa-AG market-DT go/3s PSTs 'Musa went to the market' kasuwu-ro lewono market-DT go/3s PSTs

	'(He/she) went to the market'
Fulfulde	Subject Verb Object Innawuro loot -ii binngel muudum Innawuro wash – VAP child-of her (VAP is Voice Aspect Polarity) 'Innawuro has bathed her child' Binngel Innawuro loot -aama Child of Innawuro wash -VAP Innawuro's child has been bathed Binngel Innawuro loot -ake Child of Innawuro wash -VAP 'Innawuro's child has taken bathed (self)'

From the above table, the three languages exhibit different forms of linguistic typology in the sentence constructions with regard to their respective deep structures. Hausa is a S-V-O with tense marker attached to preverbal pronoun that precedes the verb and followed by the object (accusative). While on the other hand, Kanuri is S-O-V language similar to Japanese language with the agential marker attached to subject and dative marker to object. The verb of Kanuri is agglutinating with third person pronoun and aspectual tense marker. And the Fulfulde is S-V-O similar to English and its aspectual tense marker is attached to the verb while genitive (possessive) pronoun precedes the possessed referent.

These are some of the complexities in the typology of the African language sentences that can serve as best input for the development of good AI models for the AI to effectively translate all African languages. As the individual functions of natural language process are dependent on and influenced by the input-output condition, AI models cannot arrive at valid linguistic generalization if there is no

unified theory of grammar that will incorporate the grammar of all African languages and makes it compatible with the input-output condition of the natural language processing (NLP).

Cultural Diversity and AI Translation

Although, the intimate relations between language and culture have long been the hot point of discussions by scholars and experts in many fields and disciplines, AI as the independent field will not be in exception since goal-driven is one of its seven patterns that focuses on intercultural communication or exchange. African languages are rich in in culture and their cultural diversity cannot be ignored in ensuring that the target audience can understand the text produced through AI in a culturally appropriate context. It has been established for long that translation is no longer a cross-linguistic affair but essentially cross-cultural communication (Nida 1981). Therefore, cross-cultural translation is another challenge for AI in translation science since this cultural diversity constitutes the most serious challenge to human translation more than linguistic differences.

For instance, one of the unique characteristics of the African languages is that it is uncommon in translating an English proverb to convey same referential and pragmatic meaning because they are always cultural-specific in both terms and expressions. Consider following examples in the table (2) below:

Table (2): Cultural diversity in African Language Proverbs

Language	Proverbs
English	Let sleeping dogs lie
Hausa	A bar kaza a ciki gashinta Literally ‘leave the chicken with its feather’
Kanuri	Wande kam letcin sangəmi Literally ‘Don’t wake up the sleeping person’
Fulfulde	Taa ɓoltu waandu haa wicco

	Literally ‘Don’t skinning the monkey to its tail’

From the above table, one can understand how cultural diversity of the African languages is enormous from the simple English proverb which generates many striking differences of culture materials among the languages involved. Since the NLP aims to power the computers or machines with the ability to do tasks involving human language, it could be uncompleted system when AI programmers to fail to resolve such issues of cultural diversity of African language in the NLP. So also, when the NLP system is unable to establish universal culture especially the (dis)similarities among living languages, it will be difficult for the AI to facilitate accurate and good translation output on languages that contain different cultural resources for the easy understanding of the different people in their communication. If AI is to process all forms of data including human languages to reach optimal results similar to human thinking, learning and decision making in translation, we must work towards promoting universal cultural to properly capture both cultural similarities and dissimilarities among the languages so that the AI program system may look as robust in the translation science.

For AI translation to be sensitive to linguistic and cultural contexts of all living languages in the world, it is important to consider African perspectives. AI systems are designed to solve problems within contexts. The values, interests and culture and linguistic diversity of the African languages need to be factored into the design and deployment of any AI technology. So also, for the AI systems and application to achieve, there is greater need to ensuring that NLP algorithms programmed in line with the bag-of-words (BoW) and Continuous-bag-of-words (CBoW) models as the solutions to the present inadequacy of AI translation tools.

Bag-of-words (BoW)

For comprehensive AI model to universally cater for all linguistic and cultural needs of living languages, the present paper advocates the consideration of African languages in remodeling of bag-of-words model to provide solutions to the present inadequacy of AI translation tools. The bag-of-words model is a way to represent all the words of a text without regard for their context, grammar or semantic relationships such as synonyms. Depending on how the method is implemented, the model can be robustly case sensitive. This means that the model will treat even words with different capitalization as different words. The model represents the text as the frequency of each unique word. It is a very easy algorithm, that simply counts how many times each word appears in a text.

The bag-of-words concept is built upon the distributional hypothesis, which was introduced in the Harris' (1954) paper Distributional Structure. The distributional hypothesis is the assumption that words that appear in similar contexts tend to have similar meanings. This is based on the concept that words can be defined, to some extent, by the words surrounding them.

For example, the words "eat" and "drink" are often used in similar contexts, like "You eat too much candy and drink too much soda", by the hypothesis this would suggest that the words have a similar meaning. This hypothesis has since been a very important concept in NLP tasks. Harris actually mentions the phrase bag of words in the paper, but not in the form as we know of it today, it is however clear that the paper has been a big inspiration for bag-of-word models.

Bag-of-words representation of words have been a very common way of word representation in NLP tasks, and was almost exclusively used up until the introduction of word embeddings.

Algorithm

Let

$$T = \{t_1, t_2, \ldots, t_n\}$$

be a set of n. The count of unique words in T is denoted m. Here a 'word' refers to an individual token generated from the text by a tokenization process, which typically separates the text into tokens based on spaces and punctuation. For each text t_i define a vocabulary W_i that is a set of m key-value pairs. Each key-value pair in the set contains a unique word w_j and its corresponding count k_j , representing how many times the word appeared in the given text.

$$W_i = \{(w_1, k_1), (w_2, k_2), \ldots, (w_m, k_m)\}$$

For each text the model outputs a set of key-value pairs, that for each unique word of all the texts has a number associated with it that represents how many times that word appeared in the given text.

Example of using bag-of-words: Let:

t_1 = ("This is an example")

t_2 = ("Is the model good or is it not")

Be the texts, the vocabulary would then be:

W_1 = {(this, 1), (is, 1), (an, 1), (example, 1), (the, 0), (model, 0), (or, 0), (it, 0)}

W_2 = {(is, 2), (the, 1), (model, 1), (or, 1), (it, 1), (not, 1), (this, 0), (an, 0), (example, 0)}

This can then be used to calculate the frequency of the words and the probability that a specific word will appear. Such as the probability of word w_i appearing in t_j:

$$P(w_i|t_j) = \frac{f(w_i, t_j)}{\sum_{w \in W_j} f(w, tj)}$$

Where the function f (wi, tj) is the frequency of word wi in text tj. This model assumes that each word occurrence is independent of

the others. Meaning that the probability of a word occurring, is not affected by the appearance or absence of another word thus disregarding any kind of context. The algorithm is very easy and efficient and can be used on very large datasets. The problem with the algorithm, is that completely ignores the words order and context, it also considers each different word unique even if the words could be considered synonyms. This makes the model unable to capture context of words and nuances of languages that depend on word order. This can lead to a decrease in performance in tasks that may require a deeper understanding of the meaning of words, such as question answering. The frequency of the words in a given text can then be used to train a classifier. The bag-of-words model can also be combined with N-grams, to create a "bag-of-N-grams" which in can help increase performance in many tasks. So also, the consideration of African languages by Continuous-bag-of-words (CBoW) model that operates in form of given a set of words that the model tries to predict a target word based on the input words context will facilitate more inclusion of African languages in the AI operations especially its translation tools; and at the same time, minimize the technological gaps between African languages and other world languages. Similar, the development of comprehensive NLP that embrace African languages will go a long way to minimize the challenges of low availability of input data for African languages as well as the poor discoverability of resources that do exist, thus hindering the ability of researchers to do machine translation.

For clear understanding of CBoW, Continuous Bag of Words (CBoW) is a popular natural language processing technique used to generate word embeddings. Word embeddings are important for many NLP tasks because they capture semantic and syntactic relationships between words in a language. CBoW is a neural

network-based algorithm that predicts a target word given its surrounding context words. It is a type of "unsupervised" learning, meaning that it can learn from unlabeled data, and it is often used to pre-train word embeddings that can be used for various NLP tasks such as sentiment analysis, text classification and machine translation. Consider the working of CBoW in figure (2) below:

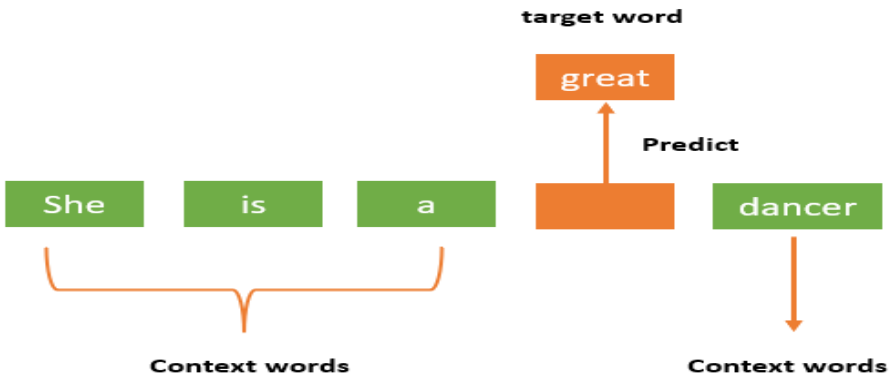

Figure (2): Prediction of the target words

From the discussion and illustration above, both the Bag-of-Words model and the Continuous Bag-of-Words model are techniques used in natural language processing to represent text in a computer-readable format, but they differ in how they capture context.

As indicated above, BoW model represents text as a collection of words and their frequency in a given document or corpus. It does not consider the order or context in which the words appear. And even though, it may not capture the full meaning of the text but the BoW model is simple and easy to implement on African languages despite its limitations in capturing the meaning of language. In contrast, the CBoW model is a neural network-based approach that captures the context of words. It learns to predict the target word based on the words that appear before and after it in a given

context window. By considering the surrounding words, the CBoW model can better capture the meaning of a word in a given context.

Opportunities of AI on the African Languages

While the African languages present unique challenges to AI development in translation that must be addressed, African languages will bring to AI immense potential to revolutionize linguistic science as the living languages in the continent have rich linguistic and cultural resources that will create excited prospects of AI for the language and communication research and teaching as well as the documentation of these languages. The inclusiveness all African languages in the digital datasets and ability of the AI to translate these languages will facilitate the following opportunities:

1. **Personalized Learning**: AI translation can adapt learning materials to individual student needs, pacing, and preferences in their individual mother tongue. This personalized approach can help students grasp concepts more effectively, regardless of their learning pace.
2. **Increased Access to Quality Education**: AI-driven translation tools can provide access to high-quality educational resources, even in remote areas. This can bridge educational gaps and empower learners in underserved communities.
3. **Efficient Administrative Processes**: Educational institutions can use AI translation to streamline administrative tasks, from admissions to resource allocation. This efficiency allows educators to focus more on teaching and mentoring students.
4. **Data-Driven Insights**: AI translation can provide valuable insights into student performance, helping educators tailor

interventions to support struggling students and challenge high achievers.

5. **Language Diversity**: Africa's linguistic diversity is a unique challenge to AI. AI translation can harness the diversity of the African languages to develop multilingual education tools, making learning more accessible to diverse language groups.

In fact, AI translation holds immense promise for transforming not only communication but education in Africa and beyond if it will ensure the inclusion of African languages in its datasets. However, apart from addressing the challenges related to access, content quality, teacher training, and data security, AI will successfully break the linguistic barriers between different nations when all African languages become accessible in the digital space. By embracing AI with a strategic and inclusive approach, African languages can unlock the full potential of AI in education, creating a brighter future for the first and second language learners across the continent and beyond.

Conclusion

Overall, the study is able to highlight the different factors that can limit the ambition of AI and its operation in the translation science especially when the linguistic and cultural potential of African languages are not properly harnessed and utilized in the NLP. It then advocates for the optimization of NLP algorithm model to ensure for AI system that will produce target hypotheses will be correspond to the source sentence with output as grammatical and fluent as possible in translation.

The paper insists on domesticating African languages through BoW model of NLP algorithm that is only concerned with whether known words occur in the document, not where in the document so that the characteristic and diversity of African languages will be

taken care by the AI translation applications. It similarly urges for extending CBoW model that tries to predict the target word by trying to understand the context of the surrounding words to African languages for achieving robust AI translation operations. Furthermore, the study contributes to scholarly literature by interrogating the limits and various opportunities that are related to the use of AI in translation science and supply input for NLP algorithm practitioners to expand the AI applicability operations in the translation science.

References :

Arakpogun, E et. al (2021) Artificial Intelligence in Africa: Challenge and Opportunities. Cham:

Springer Pp 375-388. Retrived from: 10.1007/978-3-030-62796-6_22

Bedu, A.M. (2010). Remarks on Hausa Definite Article and its Categorical Features from

Minimalist Perspective. Liwuram Journal of Humanities, Vol. 20: 184-203

Bin Rashid, A. et. al. (2023) Artificial Intelligence in the Military Africa: An Overview of the

Capabilities, Application and Challenge. International Journal of Intelligent Systems Volume 2023:1-31 pages https://doi.org/10.1155/2023/8676366

Broussard, Meredith. 2018. Artificial Unintelligence: How Computers Misunderstand the World,

1st ed. Cambridge: MIT Press.

Blench, Roger, 1998. The status of the languages of Central Nigeria. In Endangered Languages

of Africa, ed. Matthias Brenzinger, 187–205. Koln: Rudiger Koppe Verlag.

Deloitte. 2014. Demystifying Artificial Intelligence. New York: Deloitte.

Diamond, Jared. 1997. Guns, Germs, and Steel: The Fates of Human Societies. New York and

London:W.W. Norton & Co.

Eke, D. O. et. al. (2023). Responsible AI in Africa: Challenges and Opportunities. Cham: Palgrave

Macmillan.

Goutte C. et. al. (2009) Learning Machine Translation. Cambridge: The MIT Press

Hausser, R. (1989) Computation of Language: An Essay on Syntax and Pragmatic in Natural Man-

Machine Communication. Berlin: Springer- Verlag

Jonathan Slocum. 1984. *Machine Translation: its History, Current Status, and Future Prospects*.

In *10th International Conference on Computational Linguistics and 22nd Annual Meeting of the Association for Computational Linguistics*, pages 546–561, Stanford, California, USA. Association for Computational Linguistics. Retrived from: 10.3115/980491.980607

Jason Whittaker (2019). *Tech Giants, Artificial Intelligence, and the Future of Journalism*. New

York: Routledge.

Khalaf, A. E. (2017).“Metaphorical Meaning and its Effect on Interaction”. *YOBE Journal of*

Language literature and Culture. Damaturu. Yobe State, Nigeria.Vol. (5). pp. 33-47.

Khalati, M. M, & Al-Romany, T. A. H (2020). Artificial Intelligence Development and Challenges

(Araic language as a model). *International Journal of Innovation, Creativity and Change*, 13(5):916-926.

Reader, John. 1998. *Africa*, *a Biography of the Continent*. New York: Alfred A. Knopf, Inc.

Wolff, H. Ekkehard. 2019. *A Grammar of the Lamang Language*. Gluckstadt: Augustin.

Goutte C. et. al. (2009) Learning Machine Translation. Cambridge: The MIT Press