



[library@chuka.ac.ke](mailto:library@chuka.ac.ke); [www.chuka.ac.ke](http://www.chuka.ac.ke)

## DEEP LEARNING IN AFRICAN LANGUAGES TRANSLATION-A REVIEW

*Gitari T. M., Kinyori S., Mwathi G., Too E. C. and Ileri H. K.*

*Department of Computer Science, Faculty of Science, Engineering and Technology, Chuka University, Kenya*

[tmunene@chuka.ac.ke](mailto:tmunene@chuka.ac.ke); [dgmwathi@chuka.ac.ke](mailto:dgmwathi@chuka.ac.ke); [hkirimi@chuka.ac.ke](mailto:hkirimi@chuka.ac.ke)

Citation:

**Gitari T. M., Kinyori S., Mwathi G., Too E. C. And Ileri H. K. (2023). *Deep Learning in African Languages Translation-A Review*. In: Isutsa, D. K. (Ed.). *Proceedings Of the Chuka University 9th Annual International Research Conference Held in Chuka University, Chuka, Kenya From 24th To 25th November, 2022*. 396-408 Pp.**

### ABSTRACT

Machine translation (MT) is the use of computers to automatically translate one language to another. Africa has approximately 2,000 spoken languages, however, only 30 African languages have been machine translated. The main technical factor for the low-rate adoption of MT in Africa is the poor translation accuracy of existing machine translators from one African language into English. Currently, there are two approaches to MT in Africa. The first approach is the classical approach; this approach utilizes the direct mapping of input texts to produce a translated output. Examples of classical MT approaches include: statistical-based machine translators (SBMT), rule-based machine translators (RBMT) and hybridized machine translators (HMT). Classical approaches are the most widely adopted MT approach for African languages, the main reason for the wide adoption is the low cost of computing power in utilizing classical approach. However, classical approach has high-levels of inaccuracy due to language structures differentiation. The second approach is the use of Deep learning (DL) MT. Deep learning MT is a field in artificial intelligence concerned with the application of artificial neural networks to mimic the human brain learning process in language translation. Deep learning MT has the advantage of understanding phrases, complex sentence structures, and even slang when compared to classical MT approach. Deep learning has produced results 60-90% more accurate than the classical approach in translating structured languages such as French into English. However, DL has shortcomings in MT, including, high-costs of training and evaluating models, and, DL is data intensive.

These review aims to analyze the current status of machine translation approaches in Africa and provide an output recommendation for universalizing applicable MT in African languages translation. The results of these review will be in both graphical and tabular format.

**Keywords:** Machine Translation, Classical Approach, Deep Learning, Artificial Neural

## INTRODUCTION

Human language has evolved over the last 200,000 years (Baronchelli, 2012). This evolution originated with gesture basis from early humans, then primitive vocalizations and songs from Homo habilis humans, and eventually, the use of complex grammatical rules by modern humans. The modern world has over 7,117 languages (Eberhard, 2020) spoken by 7.8 billion people (Worldometer, 2021). The existence of so many spoken languages has led to the rise of a large machine translation (MT) services market valued at \$800.0million (Statista Research Department, 2021) and projected to grow to \$7.5billion by 2030. However, the vast majority of MT technology is focused on seven languages, namely English, Chinese, Urdu, Farsi, Arabic, French, and Spanish (Siavoshi, 2020). Africa's MT in language translation is underdeveloped, with a size of only \$6.0million as of 2015 (Statista Research Department, 2016) despite Africa having 16.72% of the global population (Division, 2019). The major reasons for Africa's low adoption of MT in languages translation are, low accuracy in machine translation using classical approaches (Buliva, 2017) and the lack of adequate digitized African language datasets (Knowledge 4 All Foundation Ltd., 2020).

The large language translation services market has attracted a lot of attention from the information technology (IT) industry. This has led to many breakthroughs in the application of IT into human language translations (Junyoung, 2014). The latest breakthroughs are through the application of artificial intelligence in language MT called Deep learning (DL) MT technology, this technology utilizes artificial neural networks to perform translation, Deep learning MT has shown great promise in solving the challenges of the classical approach by applying machine learning (John, 2019b). Examples of DL MT include: Simple recurrent neural networks frameworks, Gated recurrent Unit frameworks and the Long Short-Term Memory unit frameworks (Rodriguez., & Fonollosa., 2020). However, DL MT adoption in Africa has been slow compared to the rest of the world, this is due to the fact that DL MT are complex to implement and expensive due to the large computing resources required, in addition, DL MT technologies are data-hungry (Mittal, 2019)..

Thus, in this review, will seek to understand the current status of MT in Africa and analyze the classical and DL MT approaches adoption in the African languages market. The analysis will be based on verified secondary sources of data gathered from the internet open-sources. The conclusion of the analysis will provide insights into MT trends in Africa and assist researchers interested in pursuing DL-MT technologies research in the African languages translation sector.

## METHODOLOGY

1) *Data collection*: The data gathered for this review is from 25 reputable online peer-reviewed publications, published research articles, open-source research papers that are peer-reviewed, corporate articles (2016-2022). All these papers have

different data applications that has been studied and analyzed in this review. The attributes compared were MT used technologies, benefits and challenges in current MT approach, DL MT adoption approach and African continent area MT analysis.

2) *Data inclusion criteria*: To evaluate the data inclusion criteria a comparison table was drawn to include as the following attributes: Author, Sub vertical, Data collection measurements, Technologies, Benefits, Challenges, Solutions and Drivers of MT in Africa. Articles were excluded when selected attributes were not present. In our analysis, the number of classical MT approach, amount of DL MT approach, underlying technologies adoption was included since no information can find with all the peer-reviewed publications (2016-2022).

3) *Data analysis*: Oversampling method is utilized to analyze the 25 peer-reviewed articles, Oversampling is applicable due to the small size of available data on African MT. A higher sample selection of DL MT based data will be utilized.

The two metrics for analysis are:

- i. BLEU (bilingual Evaluation understudy)  
BLEU score is a metric for automatically evaluating

machine translated texts. Table 1 is of the BLEU score

interpretation guideline.

Table 1: BLEU score interpretation from 0-100

BLEU Score	Machine Translation accuracy interpretation
< 10	Poor translation
10 – 19	Low quality translation
20 – 29	The translation is clear, but has significant grammatical errors
30 – 40	Understandable to good translations
40 – 50	High accuracy translations
50 – 60	Very high accuracy, adequate, and fluent translations
> 60	Machine translation accuracy is better.

ii. Conclusion from the peer-reviewed articles that provide analytical data terms such as adoption rates of MT technologies (high, low and medium).

The article sources are arranged in Table 2 with 8 columns, including:

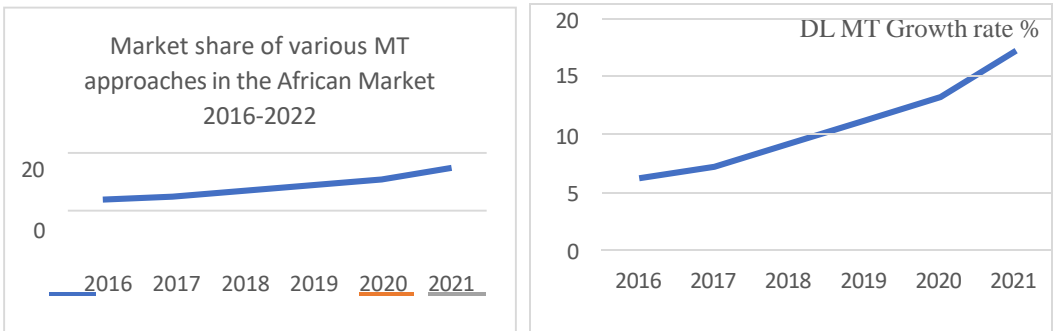
- Column 1-article number arranged alphabetically
- Column 2-article citation (APA sixth edition)
- Column 3-MT approach-classical or DL MT technology
- Column 4-Application area of MT approach e.g. Yoruba translation
- Column 5-Status of MT technology deployment e.g. prototype, widely used
- Column 6-Benefits of MT approach
- Column 7-Challenges of MT approach
- Column 8-Future Potential e.g. shutdown project, unknown

4) *Data Results:*

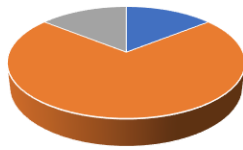
Results of analysis will be presented in Graphs and charts using the following parameters:

- Market share of various MT approaches in the African Market 2016-2022
- DL MT growth rate in the African MT market 2016-2022
- Adoption level of DL MT vs Classical MT in Africa in 2021/22
- Translation accuracy of DL MT vs Classical MT in African languages translation (using BLEU mean score of 0-100)
- Future application preference between DL MT and Classical MT

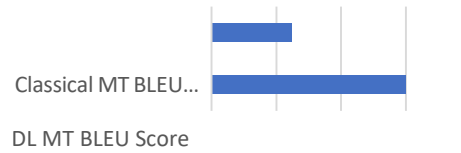
**Analysis of Results:**



Adoption level of DL MT vs Classical MT in Africa in 2021/22



Translation accuracy of DL MT vs Classical MT in African languages translation (using BLEU mean score of 0-100)



Future application preference between DL MT and Classical MT



**TABLE 2 MACHINE TRANSLATION ANALYSIS FROM 25 PEER REVIEWED SCIENTIFIC ARTICLES OF 2016-2022**

No.	Article citation	MT Technological approach	MT technology Application Geographic Area	MT technology Current status	Benefits of MT Technology	Challenges of MT Technology Approach	Future interest of Approach

1	(Research, 2022)	MT market size, share and trends	Global	<ul style="list-style-type: none"> <li>-Classical MT- 70% world market share</li> <li>-DL NMT- 15% world market share</li> <li>-Other MT approaches- 15%</li> </ul>	<p>Benefits of Classical MT:</p> <ul style="list-style-type: none"> <li>-Easy to implement</li> <li>-Cheap to maintain</li> <li>-Open-source resources</li> <li>-Well tested Technology</li> </ul> <p>Benefits of DL MT:</p> <ul style="list-style-type: none"> <li>-Higher accuracy in translation</li> <li>-Ability to capture language differentiation nuances</li> <li>-Ability to self-learn and improve, hence easier to update with higher automation over classical (Siminyu, 2018)MT</li> </ul>	<p>Challenges of Classical MT:</p> <ul style="list-style-type: none"> <li>-Low-accuracy in translation compared to DL MT</li> <li>-Labor intensive in maintenance and updates</li> <li>-Low language structures differentiation capacity</li> </ul> <p>Challenges of DL MT:</p> <ul style="list-style-type: none"> <li>-High costs of resources such as cloud computing</li> <li>-Complex technical requirements in developing frameworks</li> <li>-Data hungry</li> </ul>	<ul style="list-style-type: none"> <li>-DL MT has higher investments than classical MT, thus, higher adoption rate in the future</li> </ul>
---	------------------	----------------------------------	--------	---	--	--	---

2	(Ben Goertzel, 2021)	DL MT in Africa	-Sub-Sahara Africa	-Low-adoption	-Higher accuracy in Sub-Saharan African Languages Translation	-High input costs of cloud computing and training frameworks -Lack of adequate Digitized datasets- Most Sub-Saharan languages (80%) have no written form	-DL MT has High potential for adoption
3	(Siminyu, 2018)	DL MT	Sub-Sahara Africa	-Bilingual DL MT development for ChiChewa and Kiswahili translation to English	-Higher accuracy	-High costs of framework -Inadequate digital dataset	-DL MT has High adoption for other African languages translation
4	(Agbolo, 2022)	-DL MT called OBTranslate version 1.9.6 released in 2022 -Classical MT called OBTranslate 1.1.1 released in 2020	-Sub-Sahara Africa	-OBTranslate 1.1.1 cancelled BLEU score of 25 -OBTranslate 1.9.6 in early adoption BLEU score of 51	-Version 1.9.6 is 60% more accurate in translating Igbo compared to version 1.1.0	-Version 1.9.6 is 200% more expensive in cloud computing costs compared to version 1.1.0 -Version 1.9.6 is taking 12 months longer to train, test and validate	-DL MT has High adoption potential with 20% greater market participation

						compared to version 1.1.0	
5	(Masakha ne open source project, 2019)	DL MT using Transformer's neural network	Sub-Saharan Africa	-Pidgin dialect to English DL MT translator under development BLEU score of 60	-High accuracy compared to Classical MT translators of Pidgin dialect in Nigeria	-Lack of funding -Inadequate sized digital datasets	-DL MT has High adoption rate
6	(Knowledge 4 All Foundation Ltd., 2020)	DL MT vs Classical MT in Africa	Africa	-Classical MT 90% adoption -DL MT less than 9% adoption -Other MT 1% adoption	Classical MT has lower barriers of entry such as: -Low-cost -Easy to implement -open-source resources DL MT has higher accuracy	-Classical MT has low accuracy and is labour intensive -DL MT is resource expensive and has high data consumption rate	-High interest in DL MT in Africa
7	(Adepetun, 2022)	DL-MT	Sub-Saharan Africa	-High adoption in 55 African languages translation	-High accuracy -Digitized datasets availability	-Non-open source	DL MT has High adoption through Meta social media platforms
8	(Odoje, 2013)	-Classical MT -DL MT	-Nigeria	-Classical MT dominant adoption -DL MT nil adoption	Classical MT: -Low cost -Easy to implement DL MT: -High accuracy	Classical MT: -Low accuracy -Labour intensive DL MT: -Untested technology -High costs	DL MT has high possibility of future adoption

9	(Albarino, 2020)	DL-MT	Sub-Sahara Africa	-DL MT adoption	-High accuracy -Digitized datasets availability of 16 languages	-High costs of input resources	DL MT has High adoption
10	(Marivate, 2020)	DL-MT	Africa	-DL MT adoption	-High accuracy -Digitized datasets availability	-High costs of input resources	Resolving high costs to lead to high adoption of DL-MT in Africa
11	(INDABA X, 2021)	DL MT in Africa	Sub-Sahara Africa	-Low-adoption	-Higher accuracy in Sub-Saharan African Languages Translation -	-High input costs of cloud computing and training frameworks -Lack of adequate Digitized datasets- Most Sub-Saharan languages (80%) have no written form	-DL MT has High potential for adoption
12	(Burg Translations, 2022)	-DL MT vs Classical MT in Africa	-Africa	-Classical MT 90% adoption -DL MT less than 9% adoption -Other MT 1% adoption	Classical MT has lower barriers of entry such as: -Low-cost -Easy to implement -open-source resources DL MT has higher accuracy	-Classical MT has low accuracy and is labour intensive -DL MT is resource expensive and has high data consumption rate	-High interest in DL MT in Africa

13	(Quartz Africa, 2022)	DL-MT	-Sub-Sahara Africa	-High adoption in 65 African languages translation BLEU score 65	-High accuracy -Digitized datasets availability	-Non-open source	High adoption through Meta social media platforms
14	(LionBridge, 2022)	MT market size, share and trends	Africa	-Classical MT- 90% Africa market adoption -DL NMT-9% Africa market adoption -Other MT approaches- 1% adoption	Benefits of Classical MT: -Easy to implement -Cheap to maintain -Open-source resources -Well tested Technology Benefits of DL MT: -Higher accuracy in translation -Ability to capture language differentiation nuances -Ability to self-learn and improve, hence easier to update with higher automation over classical	Challenges of Classical MT: -Low-accuracy in translation compared to DL MT -Labor intensive in maintenance and updates -Low language structures differentiation capacity Challenges of DL MT: -High costs of resources such as cloud computing -Complex technical requirements in developing frameworks -Data hungry	-DL MT has higher investments than classical MT, thus, higher adoption rate in the future

15	(Lim, 2022)	MT market size, share and trends	Global	<ul style="list-style-type: none"> <li>-Classical MT- 65% world market adoption</li> <li>-DL NMT- 25% world market adoption</li> <li>-Other MT approaches- 10% adoption</li> </ul>	<p>Benefits of Classical MT:</p> <ul style="list-style-type: none"> <li>-Easy to implement</li> <li>-Cheap to maintain</li> <li>-Open-source resources</li> <li>-Well tested Technology</li> </ul> <p>Benefits of DL MT:</p> <ul style="list-style-type: none"> <li>-Higher accuracy in translation</li> <li>-language differentiation nuances capacity</li> <li>-self-learning capacity and improve, hence easier to update with higher automation over classical MT</li> </ul>	<p>Challenges of Classical MT:</p> <ul style="list-style-type: none"> <li>-Low-accuracy in translation compared to DL MT</li> <li>-Labor intensive in maintenance and updates</li> <li>-Low language structures differentiation capacity</li> </ul> <p>Challenges of DL MT:</p> <ul style="list-style-type: none"> <li>-High costs of resources such as cloud computing</li> <li>-Complex technical requirements in developing frameworks</li> <li>-Data hungry</li> </ul>	<ul style="list-style-type: none"> <li>-DL MT has higher investments than classical MT, thus, higher adoption rate in the future</li> </ul>
----	-------------	----------------------------------	--------	--	--	--	---

16	(Microsoft News Center, 2022)	DL MT	South Africa	-Launched in April 1,2022 for Zulu language translation BLEU score 58	-Higher translation accuracy compared to classical MT for Zulu language -Higher language differentiation fluency compared to classical MT	-Complexity in technology -Inadequate digitized Zulu data for model training, validation and testing	-High adoption
17	(Tomedes, 2022)	DL MT vs Classical MT	Africa	Analysis of 2022 Africa MT -Adoption rate 11% for DL MT And 88% for classical MT	DL MT advantages over classical MT in Africa: -High accuracy -Cost Efficiency -Scalability -Flexibility	Challenges of DL MT over classical MT in Africa: -Cost -Complexity -Lack of open-source framework shells	-High adoption rate
18	(Diden, 2022)	DL MT	West Africa	10 West African languages under Google translate DL MT for translation to French BLEU score 53	-Higher translation accuracy -Ability to capture language differentiation nuances	- Lack of open-sources -Cost	-High adoption rate

19	(Kevin Duh, 2020)	DL MT vs classical MT(SBMT)	Africa	-Swahili and Somali BLEU test scores show DL MT to be 75% and 50% more accurate than classical MT of the SBMT variant	--Higher translation accuracy -Better algorithm tuning,i.e.,DL MT is better at updates to souce code efficiency than SBMT	-Digitized Data inadequacy for African languages -	-High Adoption rate
20	(Allahsera Auguste Tapo, 2020)	DL MT	-Sahel region -DL MT in translating Bambara language in	-DL MT applied to Bambara into French translation BLEU score 45	-Higher translation accuracy	-Training data scarcity -Training data pre-processing is costly	-Low adoption rate
			Mali into French			-Lack of Smartphones to utilize the framework in Bambara	
21	(Muscleh, 2016)	DL MT	Sahara region	-Applied DL MT in medical translation for migrants from the Sahel heading to EU	-High accuracy -Flexibility	-Scarcity in training data -Low adoption due to low skills in utilizing technology by the migrants	-Medium adoption

22	(Liu, 2018)	DL MT vs Classical MT in 3 African languages	Sub-Saharan Africa	DL MT and Classical MT(SBMT) applied to Shona, Oromo and Burj translation to English BLEU Score of DL MT is 55,57 and 59 respectively, compared to SBMT at BLEU score of 17,27,40	-Higher accuracy of DL MT	Challenges of DL MT over classical MT in sub-Saharan Africa: -Cost -Complexity -Lack of open-source framework shells	-High adoption rate
23	(Maiei, 2020)	DL MT in Afrikaans,isi Zulu,N.Sotho,Setswana and Xitsonga	South Africa	DL MT applications in five South African languages to English translation BLEU score of 33.3 for DL	-Higher accuracy		Low adoption
				MT isiZulu translation			

24	(Martinus, 2020)	DL MT	South Africa	BLEU Scores for 10 South Africa languages translated by a DL MT platform to English: -Afrikaans-59 -Ndebele-24 -Xhosa-37 -Zulu-44 -Northern Sotho- 45 -Sesotho-42 -Setswana-47 -Swati-36 -Tshivheni-52 -Xitsonga-46	-Higher accuracy of DL MT	-High costs of implementation	-High adoption
25	(Z.Abbott, 2022)	DL MT	South Africa	DL MT application to translating English-to- Setswana BLEU score higher by 5 points above previous classical MT(SBMT) approach	DL MT has higher accuracy	DL MT is resource expensive and has high data consumption rate	-High interest in DL MT in South Africa

Note: -All BLEU scores are rounded off to whole numbers.

### CONCLUSION

From our analysis of the 25 peer-reviewed publications (2016–2022) in reviewing MT approaches in the African automated translation industry, it was found that classical MT approach the most widely utilized, however, DL MT has the fastest growth rate. In addition, DL MT approach demonstrates a much higher translation accuracy of 60% plus for African languages. However, classical approach MT remains dominant in application. Finally, this

analysis indicates DL MT approach has the highest future application preference to classical MT.

## REFERENCES

- Agbolo, S. (2022). *A Machine Translation Software for African Languages By An African*. Accra: Kabod Group.
- Albarino, S. (2020). *The Masakhane Project Puts Africa on the Machine Translation Map*. Witwatersrand: Slator.
- Allahsera Auguste Tapo, e. a. (2020). *Neural Machine Translation for extremely Low-Resource African languages: A case study on Bambara*. Casablanca: Deepai.org.
- Ben Goertzel. (2021). *Speech-To-Speech Machine Translation: The key to Uplifting Africa's Economy*. Seattle: Forbes Technology Council.
- Buliva, A. (2017, Summer). Machine Natural Language Translation Using Wikipedia As A Parallel Corpus: A Focus On Swahili. *Machine Natural Language Translation Using Wikipedia*. Nairobi, Nairobi Metropolitan Area, Kenya: United States International University - Africa.
- Burg Translations. (2022). *Are you overlooking translation opportunities in the African Market*. Chicago: BURG Translations, Inc.
- Diden, A. (2022). *Google: 10 West African languages under Neural Machine translation training*. Dakar: Le Journal de l'Afrique.
- Eberhard, D. M. (2020, September 14). *How many languages are there in the world?* Retrieved from Ethnologue Languages of the World: <https://www.ethnologue.com/guides/how-many-languages>
- INDABAX. (2021). *Strengthening African Machine Learning*. Durban: Deep Learning Indaba 2017-2022.
- Kevin Duh, P. M. (2020). *Benchmarking Neural and Statistical Machine Translation on Low-resource African Languages*.  
Cupertino: MetaAI.
- Knowledge 4 All Foundation Ltd. (2020, 9 20). *Cracking the Language Barrier for a Multilingual Africa*. Retrieved from Knowledge 4 All Foundation Ltd.: <https://www.k4all.org/project/language-dataset-fellowship/>
- Lim, S. N. (2022). *2022 Translation Industry Trends and Stats*. Montebelluma: Redokun Srls.
- LionBridge. (2022). *Embrace the Online Opportunity of African Languages*. Waltham: Lionbridge Technologies, LLC.
- Liu, A. (2018). *Context Models for OOV Word Translation in Low-Resource Languages in Africa*. Seattle: University of Washington.
- Maiei. (2020). *Research Summary: A focus on Neural Machine Translation for African Languages*. Montreal: Montreal AI Ethics Institute.
- Marivate, V. (2020). Sauti Yetu: African NLP Uconference . *Unconference 2020* (pp. 5-9). Johannesburg: Sauti Yetu Foundation.
- Martinus, L. (2020). *Neural Machine Translation for South Africa's Official Languages*. Cape Town: ResearchGate.
- Masakhane open source project. (2019). *The Masakhane project wats machine translation and AI to transform Africa*. Addis ababa: VentureBeat.

Microsoft News Center. (2022). *Microsoft Translator launches Zulu as next African language to be translated.*

Johannesburg: Microsoft 2022.

Muscleh, A. (2016). *Enabling Medical Translation for Low-resource Language.* Tunis: arXiv.

Odoje, C. (2013). Language Inequality: Machine Translation as the Bridging Bridge for African languages. *University of Ibadan International Research Conference* (pp. 5-12). Ibadan: Researchgate.

Quartz Africa. (2022). *Meta's AI can now translate more African languages, helping create a more inclusive internet.*

Kampala: East Africa 4IR Foundation.

Research, G. V. (2022). *Machine Translation(MT) Market Size, Share & Trends Analysis Report By Application.* San Francisco: Grand View Research.

Siavoshi, M. (2020). *The Importance of Natural Language Processing for Non-English Languages.* Boston: LionBridge AI. Siminyu, K. (2018). *Using Neural Machine Translation to Create African Languages Corpora.* Johannesburg: Africa's

Talking.

Statista Research Department. (2016). *Natural language processing market revenue in Africa 2015-2024.* Hamburg: Statista, Inc.

Statista Research Department. (2021, November 24). *Market size of the global language services industry from 2009 to 2021.* Retrieved from Statista: <https://www.statista.com/statistics/257656/size-of-the-global-language-services-market/>

Tomedes. (2022). *What is eural Machine Translation: Your complete guide.* Atlanta: Tomedes Ltd.

Worldometer. (2021, January 21). *Current World Population.* Retrieved from Worldometers.info: <https://www.worldometers.info/world-population/>

Z. Abbott, J. (2022). *Towards Neural Machine Translation for African Languages.* Durban: CSIR.