Automatic Amharic Part of Speech Tagging (AAPOST): A Comparative Approach Using Bidirectional LSTM and Conditional Random Fields (CRF) Methods

Worku Kelemework Birhanie[1](✉) and Miriam Butt[2]

[1] Faculty of Computing, Bahir Dar Institute of Technology, Bahir Dar University, Bahir Dar, Ethiopia
workukelem@gmail.com

[2] Department of Linguistics (Computational Linguistics), University of Konstanz, Constance, Germany
miriam.butt@uni-konstanz.de

Abstract. Part of speech (POS) tagging is an initial task for many natural language applications. POS tagging for Amharic is in its infancy. This study contributes towards the improvement of Amharic POS tagging by experimenting using Deep Learning and Conditional Random Fields (CRF) approaches. Word embedding is integrated into the system to enhance performance. The model was applied to an Amharic news corpus tagged into 11 major part of speeches and achieved accuracies of 91.12% and 90% for the Bidirectional LSTM and CRF methods respectively. The result shows that the Bidirectional LSTM approach performance is better than the CRF method. More enhancement is expected in the future by increasing the size and diversity of Amharic corpus.

Keywords: Amharic · POS · BI-LSTM · CRF

1 Introduction

Amharic is the working language of Ethiopian Federal Government. There are also some states which use the language at a regional level. The language has 34 million speakers according to Meshesha and Jawahar (2008), which makes Amharic the second most spoken Semitic language in the world after Arabic (Gambäck et al. 2009; Gezmu et al. 2018). Amharic is written left to right unlike other Semitic languages such as Arabic and Hebrew. It uses the Ge'ez script as its orthography (Seid and Gambäck 2005; Gambäck et al. 2009). Amharic has thirty-three core letters. Each of the letters has six additional letter versions called orders (Bender et al. 1976).

This paper contributes towards the enrichment of Amharic with technological resources. The study explores automated Amharic POS tagging (AAPOST) using a deep learning approach and Conditional Random Fields (CRF). The POS tagging task is based on news text data, in which each token has been manually tagged as to its POS.

N. G. Habtu et al. (Eds.): ICAST 2019, LNICST 308, pp. 512–521, 2020.
https://doi.org/10.1007/978-3-030-43690-2_37

Amharic follows the Subject Object Verb (SOV) arrangement in a sentence structure; nouns often come at the start of a sentence and verbs tend to appear at the end of a sentence. Sometimes, objects and subjects may switch (Yimam 2017). We have used the sentence level as our input for tagging.

Amharic is a morphologically rich language like Arabic and other Semitic languages. The language is rich in inflectional and derivational morphology. This greatly affects POS tagging, as the POS of a word changes when the word form changes derivationally. Low performance is recorded in POS tagging when word forms are considered as an input in POS tagging (Tachbelie et al. 2011).

In previous Amharic POS studies, statistical approaches have been used for the POS classification of words/tokens. In this study, a comparison of an established statistical method and a new deep learning method has been made to explore a potentially better approach for the Amharic POS tagging task.

The paper first discusses related works on Amharic POS tagging in Sect. 2, and the tagset used in the study, preprocessing tasks done on the data used, methods used, results found and discussion based on the results in the next sections. The paper closes with conclusion and recommendations for the improvement of Amharic POS tagging.

2 Related Work

Studies done on Amharic part of speech tagging so far are: Getachew (2001), Adafre (2005), Demeke and Getachew (2006), Gambäck et al. (2009), Tachbelie and Menzel (2009), Kebede (2009), Gebre (2010), and Tachbelie et al. (2011).

According to Adafre (2005), the first study on Amharic automatic part of speech tagging was done by Getachew (2001). Getachew has identified 25 classes of POS and used the Hidden Markov Model (HMM) method for the task.

Adafre (2005) used 1,000 tokens extracted from 5 Amharic news articles. An accuracy of 74% was registered using the conditional random fields method. Ten tagsets have been used, these are: Noun, Verb, Auxiliary, Numeral, Adjective, Adverb, Adposition, Interjection, Punctuation, and Residual. The corpus size is limited, and the accuracy is not satisfactory.

Demeke and Getachew (2006) worked on manual part of speech tagging for Amharic. They developed an Amharic tagset at the Ethiopian Language Research Center (ELRC) under the project called “The Annotation of Amharic News Documents”. The project aimed at filling the resource gap in Amharic. 1,065 Amharic news texts were collected from Walta Information Center (WIC), which is private broadcasting corporation. A total of 210,000 tokens were tagged with part of speech. This corpus has subsequently been used in most Amharic part of speech tagging studies and is also used in this study. Hereafter, this dataset is referred as “WIC dataset”.

Gambäck et al. (2009) used three algorithms and three tagset types to experiment with Amharic part of speech tagging. The dataset used was the WIC dataset and the tagset types used were: ELRC, BASIC, and SISAY. ELRC is the tagset developed at Ethiopian Language Research Center by Demeke and Getachew (2006), which has 31 tags. BASIC tagset is the major tags in ELRC tagset, which includes 11 tags. SISAY tagset contains 10 tags recommended by Adafre (2005). The average accuracy results were 85.56%, 88.30% and 87.87% for the HMM, SVM and Maximum Entropy methods, respectively.

Tachbelie and Menzel (2009) presented "Amharic Part-of-Speech Tagger for Factored Language Modeling". They have used the WIC dataset; and the methods employed were Hidden Markov Model (HMM) and Support Vector Machine (SVM), which achieved 82.57% and 84.87% average accuracies respectively.

Kebede (2009) applied a decision tree algorithm for Amharic part of speech tagging and registered an accuracy of 84.9%. Here again, the WIC dataset was used, but only 800 sentences were selected as a sample.

Gebre (2010), used the WIC dataset for an Amharic part of speech experiment. As a methodology, he employed Conditional Random fields (CRF), Support Vector Machine (SVM), Brill tagger and Hidden Markov Model (HMM). The average accuracy results found were, 90.95%, 90.43%, 87.41% and 87.09% in order.

Tachbelie et al. (2011) also studied Amharic POS tagging based on the WIC data. The study applied POS tagging on segmented and unsegmented Amharic words. The average accuracies obtained are 86.30% and 93.5% for the unsegmented and segmented words, respectively. The segmented one removes the affixes and determines the tag of a given word. However, affixes are very important for the determination of POS of a word in Amharic. In this study, affixes are not removed from words and words are used as they appear in the corpus.

All of the studies used more or less similar tagsets with minor differences. Getachew come up with 25 tagsets for the first time. Adafre (2005) used a more compact form of the tagsets, which in turn are very similar to the major classes of the WIC dataset. Gambäck et al. (2009) tested three types of tagsets.

This study tries to enhance the performance by considering different factors. The sentence level is taken to be the most essential component for deciding the part of speech of a word as it provides important distributional information about the word. Word embedding has been incorporated for the betterment of the result since it is used to draw relationship between words.

3 Tagset

As already discussed, most Amharic part of speech tagging studies used the WIC dataset prepared by ELRC. The same dataset is utilized in this study. The dataset is tagged into 11 major parts of speech, these are: Noun (N), Pronoun (PRON), Verb (V), Adjective (ADJ), Preposition (PREP), Conjunction (CONJ), Adverb (ADV), Numeral (NUM), Interjection (INT), Punctuation (PUNC), and Unclassified (UNC).

4 Preprocessing

In the WIC dataset there is informative tagging related to the date of the news article, the file name, copyright information and tags that indicate the title and body of each of the news articles, etc. All this meta-information is not relevant for the task at hand. Hence, such embedded tags are excluded, and only relevant information is extracted for the task.

In the data, there are tokens which have not been given a POS tag. Such tokens are removed from the data. The other important preprocessing step is the identification of sentences. Sentence information is not included in the data. If sentence structure is ignored, we cannot capture distributional information of words in the context of grammatical structure. Hence, further processing has been accomplished to identify sentences based on sentence punctuation markers in Amharic. Punctuation marks that signal the end of a sentence are: ጥያቄ ምልክት*Tyaqe mlkt* ‘question mark’ (?), ቃል አጋኖ*qalagano* ‘exclamation mark’ (!), and አራት ነጥብ*arat neTb* ‘period’ (።).

The format of the data in the WIC dataset is an Amharic word followed by its target tag, both in one file. Since the algorithms require separated input and output, the data is split into a list of words and their associated target tag.

The data is then converted in a way amenable to the algorithm used. Each token and tag have been integer encoded; additionally, the tags are converted into one hot vectors encoding. The vectorization is done to represent the text data and to feed it into the machine learning algorithms.

5 Methods

Two machine learning methods were applied in this study, Bidirectional Long Term Short Term Memory (BÏ-LSTM) and Conditional Random Fields (CRF). In both methods, different layers are used from the input to the output. In the layered model, Dropout has been used. Dropout helps to overcome the problem of overfitting by controlling the effect of noisy data in the training set (Srivastava et al. 2014).

A word embedding layer is also used in this study. Word embedding helps to produce linguistic relationships between words (Schnabel et al. 2015); which foster the model to produce good predictions. The word embedding layer used in this study is based on the word2vec algorithm developed by Mikolov et al. (2013). The underlying algorithms employed for the task part of speech tagging, BI-LSTM and CRF, are discussed below.

The deep learning method provides multiple layers to allow for more learning and to facilitate good data representation. This makes the overall result better (Chollet 2018). The great advantage of using Recurrent Neural Network (RNN) is the consideration of context information in a data, which is essential for modeling human language (Graves et al. 2005; Huang et al. 2015; Chollet 2018).

BI-LSTM combines the utility of Bidirectional RNN (BRNN) and LSTM. BRNN considers contexts by working both in a backward and forward manner over a sequence data. This is very helpful in a problem like POS tagging since the POS of a target word is affected by the POS of surrounding words. LSTM has the ability to consider long distance dependencies; as a result, it helps to solve problems of vanishing or decaying information exhibited by RNNs (Graves et al. 2005; Graves and Jaitly 2014; Huang et al. 2015).

Historical and future information can be considered in a BI-LSTM. The input layer in BI-LSTM is a sequence of vectors for the tokens of a sentence and the output layer is a sequence of hidden states for each token’s vector. The final hidden state is determined by the combination of the forward and backward hidden layers’ result (Lin et al. 2017).

A Time Distributed Dense layer and activation layer are parts of the model in this study. The Dense layer is a fully connected network used to output the part of speech of each Amharic token. The time distributed facility helps to create a one to one relationship between the input token and the output part of speech of a given Amharic token. The activation function helps to activate a neuron for the output layer. It calculates the probability of the 11 parts of speech, given a particular Amharic token.

CRFs are the other approach implemented in this study. They are based on probability estimation. CRF is defined as "Let G = (V, E) be a graph such that Y = (Yv)v ∈ V, so that Y is indexed by the vertices of G. Then (X, Y) is a conditional random field in case, when conditioned on X, the random variables Yv obey the Markov property with respect to the graph: p(Yv | X, Yw, w ≠ v) = p(Yv | X, Yw, w ∼ v), where w ∼ v means that w and v are neighbors in G" (Lafferty et al. 2001). In this study, X is Amharic sentences and Y is the range of POS tags given for tokens in those sentences.

The architecture of this study encompasses three major components: input, classifier, and output as illustrated in Fig. 1.

Fig. 1. Model of AAPOST.

6 Result and Discussion

6.1 Data

8041 sentences have been identified for the experiment. Table 1 shows token statistics of the corpus for the 11 parts of speech used in the study after pre-processing.

Table 1. Token statistics.

No.	Part of speech	Number of tokens	Percentage (%)
1	Noun	116470	58.322
2	Pronoun	2696	1.35
3	Verb	37392	18.724
4	Adjective	11451	5.734
5	Preposition	5534	2.771
6	Conjunction	1364	0.683
7	Adverb	2294	1.149
8	Numeral	8629	4.321
9	Interjection	2	0.001
10	Punctuation	13700	6.86
11	Unclassified	170	0.085
Total		**199702**	**100**

The statistics indicates that nouns form the largest group and that the smallest group is represented by interjections. Interjections do not occur in large numbers because the document consists of very formal reporting of news articles.

The data has been split into two sets in the ratio of 95:5. The first set is composed of 7638 sentences, which is used for training and validation. The validation set is about 10% (764) of these sentences and the remaining 90% (6874) sentences are used for the training. The validation set is essential for controlling overfitting during learning. The other 5% (403) sentences are used for the purpose of evaluation. The model is evaluated using accuracy based on the test set.

6.2 Word Embedding

The model has been trained using an embedding layer, which utilizes pretrained word embedding developed via a combination of two corpora, the WIC (Getachew and Demeke 2006) and the Contemporary Amharic Corpus (Gezmu et al. 2018).

The embedding model from the combined corpus has a dimensionality of 100 and a window size of 5. The values are selected since they are default values in most applications (Elton et al. 2019).

The final vocabulary contains 744,269 tokens. The word embedding model is able to produce semantically related words given randomly selected Amharic words. Sample outputs for a word are depicted below. The algorithm has been told to show ten most likely related words for the word ሴት*sEt* 'female', and generate the output:

እናት*enat* 'mother', አንዷት*andit* 'one female', ሚስት*mist* 'wife', እህት*eht* 'sister', ወንድ*wend* 'male', ሴትና*sEtna* 'female and', እርጉዝ*erguz* 'pregnant', መበለት*mebelet* 'widow', ነፍሰ*nefse* 'a word used in combination with other word to form a compound word, in this sense it can be used to form a compound word to refer to pregnancy', የደረሰች*yederesech* 'a female about to give birth'.

6.3 POS Experiment

Various epoch levels have been tried and a training over 15 epochs was found to be optimal. Figure 2 shows the trend in accuracy at various epochs. In the Figure, values on the vertical axis can be multiplied to determine the accuracy in terms of percentage. For example, 0.88 means 88% accuracy.

From Fig. 2 one can see that accuracy increases in the initial learning steps. After 15 epochs, the validation accuracy goes in a constant rate. Hence, epoch 15 was selected for the experiments.

The model is trained on two learning algorithms, and 91.12% and 90% accuracies are recorded for BI-LSTM and CRF methods respectively. The BI-LSTM learning method has showed an accuracy increase by 1.12% over the CRF method. Both methods consider contextual information, but BI-LSTM includes a long term memory that enables it to take into account longer distance relationships between parts of speech than what CRF allows. This may be the reason for the better performance by the BI-LSTM learning method. The highest accuracy in previous Amharic POS tagging studies was 90.95% using the unsegmented words, the detail is indicated in Sect. 2. In this study, the accuracy increased by 0.17%.

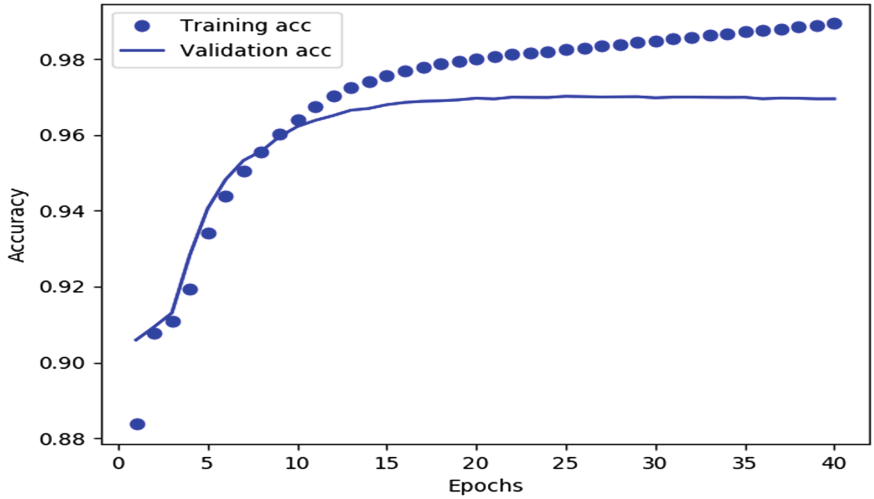

Fig. 2. Epoch vs accuracy.

The taggers Cohens Kappa results are 0.85 and 0.83 for the BI-LSTM and CRF classifiers in respectively. The Cohens Kappa result tells the quality of the classifiers (McHugh 2012; Gulzar et al. 2018). Based on McHugh (2012), if the Kappa result is above 0.9, it is almost perfect classifier and if the Kappa result lies between 0.8 and 0.9, it is a strong classifier. Hence, we can say that the models produced strong taggers. The BI-LSTM classifier is better than the CRF classifier by 0.02.

6.4 Error Analysis

Table 2 shows the confusion matrix for the BI-LSTM method. The result for CRF method has the same trend in the output except some variations.

Table 2. Amharic part of speech tagging confusion matrix.

Predicted / Actual	N	PRON	V	ADJ	PREP	CONJ	ADV	NUM	PUN	UNC
N	5584	4	90	51	19	0	16	12	0	0
PRON	25	96	3	3	2	0	0	1	0	0
V	349	0	1494	6	0	0	0	0	0	1
ADJ	130	5	21	388	1	0	2	8	0	0
PREP	13	1	2	0	272	1	1	0	0	0
CONJ	7	1	2	0	5	50	0	0	0	0
ADV	21	1	9	3	9	1	63	0	0	0
NUM	39	1	2	1	0	0	0	363	0	0
INT	0	0	0	0	1	0	0	0	0	0
PUN	3	0	0	0	0	0	0	0	685	0
UNC	1	0	3	0	0	0	0	0	0	2

From the total of 9874 tokens, 8997 (91.12%) tokens are correctly classified and 877 (8.88%) tokens are wrongly classified. The major misclassifications happened between verbs and nouns. 349 verbs are incorrectly tagged as noun and 90 nouns are misclassified as verbs. The second misclassification is between noun and adjective. 130 adjectives are incorrectly tagged as nouns and 51 nouns are wrongly classified as adjectives. This resulted in a misclassification of 620 (70.7%) tokens from a total of 877 misclassified tokens.

Most of the misclassifications in pronouns, numerals, prepositions, adverbs, and conjunctions part of speech were found to be in the noun tag. Two reasons may account for this. The first may be that nouns represent the most frequent POS in the data. Nouns constitute 58.32% of the total data in the corpus; as a result, tokens from less frequent POS may erroneously be tagged as the noun category. Secondly, nouns have a more variable distribution in Amharic sentences in comparison to the other Amharic POS. Nouns may occur at the very beginning of a sentence as a subject or may appear in the middle of a sentence as an object; hence, if there is an ambiguity in the classifier, it may incorrectly classify a token as a noun.

Adverbs classification performance is poor as compared to other part of speeches. There are 107 tokens classified as adverb. From these tokens, 63 (58.88%) are classified correctly and 44 (41.12%) are classified wrongly. Very few primary adverbs are available in Amharic. Most of the Adverbs are normally derived from other part of speeches, to clarify time, place, situation, etc. This may create confusion for the classifier and may assign a different part of speech for an adverb token.

In the CRF method, 8887 (90%) tokens are classified correctly and 987 (10%) tokens are misclassified into wrong POS out of 9874 tokens. The trend in the result is like the BI-LSTM output; similar kinds of misclassifications are recorded. The only difference here is a 1.12% decrease in accuracy.

7 Conclusion and Future Work

Part of speech tagging has been applied to a medium sized WIC corpus comprising news texts. The corpus was tagged on the basis of 11 major POS tags.

The plain text format of the corpus has been taken as an input for preprocessing. Preprocessing steps such as identifying words and tags from the corpus, finding sentence boundaries in the corpus and converting the data into one-hot vectors were accomplished.

The vector form of the data is feed into the learning algorithms, BI-LSTM and CRF. The algorithms have been further enhanced via a word embedding model. A dropout facility has also been incorporated into the model to overcome the negative impact of noise in the data. There are mistagged tokens in the corpus. The dropout helps to deal inconsistencies as a result of mistagging.

After training over 15 epochs, a 91.12% accuracy was registered for the BI-LSTM method and 90% accuracy for the CRF method. The result shows that BI-LSTM performs better than CRF and it also showed 1.12% accuracy increase as compared to previous highest Amharic POS tagger. This is likely due to the consideration of longer dependencies in the POS.

70.7% of the misclassification is between noun and verb, and noun and adjective tokens. Tokens in the other POS are misclassified to nouns in most cases. This is due to sparseness of some tag categories; that is, less frequent tokens are incorrectly tagged as the class of more frequent tokens by the model.

Error in the corpus contributed to the decrease in performance of the tagger. Hence, the corpus also needs some correction. The subsequent paragraphs discuss points to be considered for future research in the area.

Though the result produced is good, there are aspects that need improvement. The first is to increase the available Amharic corpus both in number and diversity. This study used texts from the news domain. Diverse sources and genres should be included to provide a more general training and test sets. This will also likely help to raise the representation of all parts of speech and help to minimize the misclassification problem.

The model has been tested only on the news domain. If used in a different context, it may not have similar performance. This remains to be tested. As another further step, it is desirable to scale up the domain under consideration.

The sparseness of the data contributed to the decrease in performance of the tagger. The situation is worse for adverb tokens. Hence, it is good to balance the dataset across various part of speeches.

Acknowledgement. We thank the research group of Computational Linguistics, University of Konstanz for the support and provision of required materials for the study. The appreciations is extended for Bahir Dar University, Bahir Dar Institute of Technology. We also thank Toqeer Ehsan for his constructive comments.

References

Adafre, S.: Part of Speech tagging for Amharic using conditional random fields. In: Proceedings of the ACL Workshop on Computational Approaches to Semitic Languages, pp. 47–54. ACL, Ann Arbor (2005)

Bender, L., Bowen, D., Cooper, R., Ferguson, C.: Language in Ethiopia. Oxford University Press, London (1976)

Chollet, F.: Deep Learning with Python. Manning Publications Co, Shelter Island, NY (2018)

Demeke, G., Getachew, M.: Manual annotation of Amharic news items with part of speech tags and its challenges. ELRC Working Pap. **2**(1), 1–16 (2006)

Elton, D., et al.: Using natural language processing techniques to extract information on the properties and functionalities of energetic materials from large text corpora. In: Proceedings of the 22nd International Seminar in New Trends in Research of Energetic Materials. arXiv e-prints (2019). https://arxiv.org/pdf/1903.00415.pdf

Gambäck, B., Olsson, F., Argaw, A., Asker, L.: Methods for Amharic part-of-speech tagging. In: Proceedings of the EACL Workshop on Language Technologies for African Languages – AfLaT, pp. 104–111. Association for Computational Linguistics, Athens (2009)

Gezmu, A., Seyoum, B., Gasser, M., Nürnberger, A.: Contemporary Amharic corpus: automatically morpho-syntactically tagged Amharic corpus. In: Proceedings of the First Workshop on Linguistic Resources for Natural Language Processing, pp. 65–70. Association for Computational Linguistics, Santa Fe (2018)

Gebre, B.: Part-of-Speech Tagging for Amharic. A Project submitted as part of a programme of study for the award of MA Natural Language Processing and Human Language Technology, Universite De Franche-Comt 'E Centre Lucien Tesniere' and The University of Wolverhampton School of Law, Social Sciences and Communications (2010)

Graves, A., Jaitly, N.: Towards end-to-end speech recognition with recurrent neural networks. In: Proceedings of the 31 International Conference on Machine Learning, Beijing, China, JMLR: W&CP, vol. 32 (2014)

Graves, A., Fernández, S., Schmidhuber, J.: Bidirectional LSTM networks for improved phoneme classification and recognition. In: Duch, W., Kacprzyk, J., Oja, E., Zadrożny, S. (eds.) ICANN 2005. LNCS, vol. 3697, pp. 799–804. Springer, Heidelberg (2005). https://doi.org/10.1007/11550907_126

Gulzar, M., Ali, A., Naqvi, B.: Performance evaluation and comparison of classification techniques for outcome estimation in strategic board games. Int. J. Comput. Sci. Netw. Secur. **18**(7), 103–110 (2018)

Huang, Z., Xu, W., Yu, K.: Bidirectional LSTM-CRF Models for Sequence Tagging. arXiv e-prints (2015). https://arxiv.org/pdf/1508.01991.pdf

Kebede, G.: The application of Decision Tree for part of speech (POS) tagging for Amharic. A thesis submitted to school of graduate studies of Addis Ababa university in partial fulfillment of the requirements for the degree of Master of Science in information science (2009)

Lafferty, J., McCallum, A., Pereira, F.: Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: Proceedings of the 18th International Conference on Machine Learning (ICML 2001), pp. 282–289. ACM (2001)

Lin, B., Xu, F., Luo, Z., Zhu, K.: Multi-channel BiLSTM-CRF model for emerging named entity recognition in social media. In: Proceedings of the 3rd Workshop on Noisy User-generated Text, pp. 160–165. Association for Computational Linguistics, Copenhagen (2017)

McHugh, M.: Interrater reliability: the kappa statistic. Biochem. Medica **22**(3), 276–282 (2012)

Meshesha, M., Jawahar, C.: Indigenous scripts of African languages. Indilinga: Afr. J. Indigenous Knowl. Syst. **6**(2), 132–142 (2008)

Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient Estimation of Word Representations in Vector Space. arXiv e-prints (2013). https://arxiv.org/abs/1301.3781v3

Schnabel, T., Labutov, I., Mimno, D. Evaluation methods for unsupervised word embeddings. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 298–307. Association for Computational Linguistics, Lisbon (2015)

Seid, H., Gambäck, B.: Speaker independent continuous speech recognizer for Amharic. In: Interspeech, 9th European Conference on Speech Communication and Technology, Lisbon (2005)

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**, 1929–1958 (2014)

Tachbelie, M., Menzel, W.: Amharic part-of-speech tagger for factored language modeling. In: International Conference RANLP, pp. 428–433. Association for Computational Linguistics, Borovets (2009)

Tachbelie, M., Abate, S., Besacier, L.: Part-of-speech tagging for under-resourced and morphologically rich languages – the case of Amharic. In: Conference on Human Language Technology for Development, pp. 50–55. Alexandria (2011)

Yimam, B.: Amharic Grammar, 3rd edn. Addis Ababa University B.E. Printing Press, Addis Ababa (2017). ባየ ይማም. የአማርኛ ሰዋሰው. ሶስተኛ እትም. አዲስ አበባ ዩኒቨርሲቲ ቢ.ኢ. ማተሚያ ቤት፡ አዲስ አበባ(2009)