Building Word Representations for Wolof Using Neural Networks

Alla Lo[1], Cheikh M. Bamba Dione[2], Elhadji Mamadou Nguer[3(✉)], Sileye O. Ba[4], and Moussa Lo[3]

[1] Université Gaston-Berger, Saint-Louis, Senegal
lo.alla@ugb.edu.sn
[2] University of Bergen, Bergen, Norway
dione.bamba@uib.no
[3] Université Virtuelle, Dakar, Senegal
{elhadjimamadou.nguer,moussa.lo}@uvs.edu.sn
[4] Dailymotion, Paris, France
sileye.ba@dailymotion.com

Abstract. Because a large portion of population in rural areas in sub Saharan Africa understand only local languages, they do not have access all to content available in the World Wide Web. Most content are available in English, Spanish, French, etc. Content in low-resource languages such as Wolof, which is mostly spoken in Senegal, are scarce. Automatic systems for natural language understanding such as machine translation systems that can transform information from common to low-resource languages would allow people in rural areas to access relevant scientific or health content.

Nowadays, word representation is the preliminary step of natural language understanding models. This paper presents investigations we conducted to build Wolof words representation using a corpus gathered from Internet. We applied neural word embedding models to the Wolof language corpus. These models are known to be able to capture into the embedding space semantic an syntactic relations between words. Experiments we conducted suggest that, despite a limited corpus size, our models successfully captures relations between words.

Keywords: Neural network · Word embedding · Low resource language · Wolof

1 Introduction

According to UNESCO, in 2012, only 20% of children in rural sub-Saharan African countries were enrolled in primary school [1]. Thus, a large portion of the population, specifically elderly people, only communicate in local languages.

Authors thank the CEA MITIC for funding this work.

J. P. R. Thorn et al. (Eds.): InterSol 2020, LNICST 321, pp. 274–286, 2020.
https://doi.org/10.1007/978-3-030-51051-0_20

These people cannot access all relevant information about agriculture, health, available into digitized documents on the internet.

Nowadays, machine learning models such as neural networks can be used to automatically extract information from digitized documents [3,18].

Although languages such as English, Chinese, French, Spanish, Arabic are dominant in the digitized world, digital documents are also available for low-resource languages such as Wolof, Pulaar which are among the mostly spoken languages in sub-Saharan Africa. These aforementioned low-resource languages have been investigated by linguists in term of their syntactical and grammatical structures [16]. Also, dictionaries relating these languages to French and English exist [17]. However, because of data scarcity, these low-resource languages have not been explored with machine learning methods to address natural language understanding tasks. Building machine learning models, such as neural machine translation systems, that can automatically extract information from audio and text data will have a strong impact in regions of the world where those languages are dominant. This will allow people from rural areas of sub-Saharan Africa, specifically elderly people, to access relevant information about health, they would not otherwise.

In this paper, we present investigations we have conducted on building word representations for Wolof. For efficient processing, words that are discrete tokens have to be represented in a numeric format that is amenable to algebraic calculations. The basic word representation is the so called one-hot encoding: for a vocabulary of N words, the n'th word is encoded as an N dimension vector filled with 0 except for the n'th component which is 1. The issue with this representation is that every word is at equal distance to every other word. More precisely, distances between words encode neither syntactic nor semantic information. Another possible representation is the so called bag-of-words model which consists in representing the n'th word as an N dimension vector whose m'th component is the counts of the number of time word m co-occurs with word n [19]. This model allows to capture information about word co-occurrences: two words that co-occur with the same words will have similar vector representation. A problem with this model, however, is that it generates very high dimensional word representation.

Recently, more sophisticated models to word representation based on neural networks have been proposed [4]. These models are able to predict a word using its context (i.e. words around it) as input or, vice versa, predict the context using the word as input. The internal representation of these networks allows to generate lower dimensional word embedding that captures a richer semantic information than the bag-of-words.

In our research work, we have gathered a corpus of Wolof documents from the web. The corpus is used to train neural network-based models for word representation. In this paper, we discuss how we have applied word embedding models to this low-resource language and present our results. To our knowledge, this is the first time Wolof language is the topic of such a study. In our investigations, we built three models: the continuous bag-of-word model (CBOW, [4]) that predicts a word given its local context, the skip-gram model that predicts a word

local context given the word (skip-gram, [4]), and the global vector model that predicts a word given its global context (GloVe, [6]). We have conducted various experiments to assess the effectiveness of our models. The results show that despite the limited corpus size, constructed word embedding models successfully capture semantic structure that can be recovered using distances between words in the embedding spaces.

The remainder of this paper is organized as follows. Section 2 discusses related work. Section 3 gives details about the word embedding models we have investigated so far. Section 4 presents the corpus we used to learn the models parameters. Section 5 describes the experiments we have conducted. Section 6 concludes the discussion and outlines future work.

2 Related Work

Because of the importance of language in human interaction, natural language understanding has been widely investigated in computer science. Natural language understanding can be approached either from speech or text inputs. In this paper our focus is on text inputs. Without being exhaustive, natural language understanding can have as object parts-of-speech tagging, named entity recognition, text classification, automatic translation, etc. [3,18].

Natural language understanding requires representing word tokens as elements of a vector space so that algebraic calculations can be applied to them. The simplest representation is the one-hot encoding vector which does not allow to capture relationships between words, as every word is at the same distance to every other one. The bag-of-words representation allows to capture co-occurrence relations between words and thus some form of syntactic and semantic relationships [19]. The main drawback of this representation is that it is very high dimensional. Using word-document occurrences matrix factorization allows to obtain lower dimensional representation of the bag-of-words vector while maintaining relations between words [19].

In [5], word embedding with the skip-gram and the continuous bag-of-words (CBOW) were proposed. Furthermore, experimental evidences were given about the abilities of these embedding models to capture semantic and syntactic relationships between words. These relations could then be recovered by computing the distances between word vectors in the embedding space. We may note in passing that word embeddings with neural network have been implicitly used in other works. In [3], although not explicitly stated, Collobert et *al.* used neural network generated word embeddings to jointly solve some natural language understanding tasks. In [2], neural word embedding was implicitly used for language modelling tasks. However, the effectiveness of neural word embedding to capture word semantic relationships was explicitly pointed out by Mikolov et al. in [5].

The purpose of this paper is to provide a representation of Wolof words using neural network embeddings. So far, investigations about Wolof have mostly been about building dictionaries [11], and studying the grammatical structures [8–10]

of the language. Other African languages such as Swahili (Southern Africa) and Amharic (Eastern Africa) have been subject of research about automatic statistical based machine translation [12–14]. However, to our knowledge, this work is the first to address Wolof (Western Africa) word representation.

3 Word Embedding Models

Let us introduce notations we use in the remainder of this article. We define $\mathcal{V} = \{v_1, v_2, ..., v_N\}$ as the Wolof vocabulary, more precisely the set of words occurring in the corpus. Every element of the vocabulary v_n can be represented by a one-hot encoded vector $H(v_n) = (\delta_{nm})$, $m = 1, ..., N$ where δ_{nm} is the Kronecker delta symbol which is 1 when $n = m$ and 0 otherwise. Our corpus is split into phrases. If we define the size of our context to be the integer C, from every phrase, we can extract sequence of words $w_{t-c}, w_{t-c+1}, ..., w_{t-1}, w_t, w_{t+1}, ..., w_{t+c-1}, w_{t+c}$. Given this sequence of words, the target word is the central word w_t. The other words $w_{t-c}, w_{t-c+1}, ..., w_{t-1}, w_{t+1}, ..., w_{t+c-1}, w_{t+c}$ represent the context of the target.

Given the target words and their contexts, neural word embeddings are all based on the same principle: build a neural network that predicts the target word from its context, or vice versa. Parameters obtained during training of the neural networks associated to every vocabulary words are taken as their embedding. In the following sections, we describe the continuous bag-of-words (CBOW) first, then the skip-gram, and finally the global vector (GloVe) model.

3.1 Word Embedding with CBOW

In the CBOW model, word embedding is obtained using a three-layer neural network comprised of an input, a hidden layer, and an output layer.

The input layer takes the words of the context in their one-hot encoded forms $H(w_{t-c})$. Then every one-hot encoded word is projected into the embedding space using an embedding matrix E. The embedding matrix E is of dimension $N \times D$ where N is the number of words of the vocabulary, and D is the dimension of the embedding space. The embedding dimension D is an hyper-parameter to be set. Multiplying the one-hot encoding vector of a word with the embedding matrix consists simply in selecting from the embedding matrix the row corresponding to the word. Then projected words representations are averaged. The average representation is fed into a logistic regression layer with parameter W to predict the target word. Figure 1 gives a graphical representation of the CBOW model architecture.

Learning the CBOW is achieved by optimizing the following cost:

$$J(E, W) = \frac{1}{T} \sum_{t=1}^{T} \log p(w_t | w_{t-c}, c = -C, ..., C, c \neq 0, E, W) \tag{1}$$

where T is the total number of target words in the training corpus. The model is named continuous bag-of-words, because the word one-hot encoding averaging step removes all information related to the words ordering.

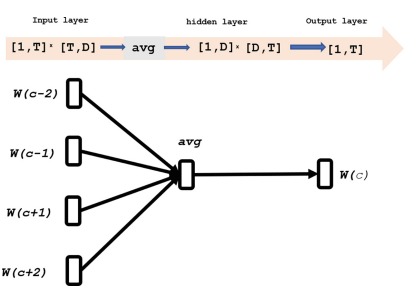

Fig. 1. CBOW model architecture.

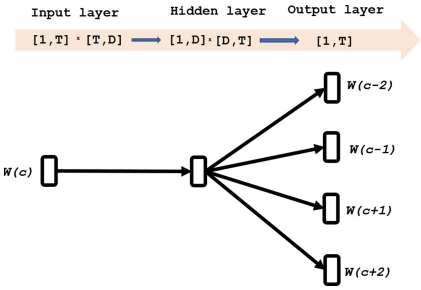

Fig. 2. Skip-gram model architecture.

3.2 Word Embedding with the Skip-Gram

The skip-gram model is essentially the inverse of the CBOW model. In skip-gram, the target word w_t is used as input, and context words as predicted outputs. Represented in the one-hot encoding format, the input word is projected into the embedding space using the embedding matrix E. As for the CBOW, projecting the word's one-hot encoding vector consists in selecting from the embedding matrix the row corresponding to the target word. Then, the projected word vector is fed into a logistic regression layer with parameter W to predict words belonging to the target word's local context. Figure 2 gives a graphical representation of the skip-gram architecture.

Learning the skip-gram from a training corpus consists in optimizing the following cost function:

$$J(E, W) = \frac{1}{T} \sum_{t=1}^{T} \sum_{c=-C,\ldots,C,c\neq 0} \log p(w_{t+c}|w_t, E, W) \tag{2}$$

with respect to the embedding matrix E and the regression parameters W. The skip-gram is named as is because words of the context are predicted from the one that is skipped, in reference to language models where n-gram is the name of model predicting a word from its n predecessors.

3.3 Word Embedding with GloVe

The global vector (GloVe) introduced in [7] combines two approaches, count-based matrix factorization and neural network-based embedding. GloVe is based on the construction of global co-occurrence matrix X between words, by processing the entire training corpus using a sliding context window. Each element X_{nm} represents the number of times word v_n of the vocabulary appears in the context of word v_m. In contrast to the CBOW and skip-gram models which are trained from target words' local contexts, GloVe takes into account all the aggregated local target words contexts. Thus, the co-occurrence matrix X contains the words' global context.

Once the co-occurrence matrix X is computed, least squares regression is used to build vector representations E_n for word v_n. It has to be noticed that vectors E_n are rows of an embedding matrix E. Learning GloVe using a training corpus is achieved by minimizing the following cost:

$$J(E, W, b) = \sum_{n,m=1}^{N} f(X_{nm})(E_n W_m^{\perp} + b_n + b_m - \log X_{nm})^2 \tag{3}$$

where $f(.)$ is a function weighting words contributions to the cost $J(E, W, b)$ according to co-occurrence counts X_{nm} and b, b_n, b_m are bias. The weight function f is defined as:

$$f(X_{nm}) = \begin{cases} \left(\frac{X_{nm}}{\tau}\right)^{\alpha} & \text{if } X_{nm} < \tau \\ 1 & \text{Otherwise} \end{cases} \tag{4}$$

where $\tau = 100$ and $\alpha = 3/4$ [7]. The main role of the weight function is to reduce the effect of very frequent words such as the stop words.

4 Wolof Corpus

Wolof is mainly spoken by 14 million people in Senegal, Gambia, and Mauritania, three countries located in West Africa. Wolof texts are written either in Arabic or Latin alphabets. In this work, we only consider texts written in Latin alphabet. To give the reader a flavour about what Wolof looks like, we provide a very short extract from a classical Senegalese literary text *Bataaxal bu gudde nii* which translates in English as *So long a letter* [20]:

"Aysatu, Jot naa sa bataaxal, fekk ma ci nattu. Ni ma lay tontoo mooy, wéttalikoo kaye bii, lu ma xalaat def ci; ndax, yàgg a déeyook yow, tax na maa xam ni, «waxtaan ay dund bi gor», day giifal naqar. Aysatu, sunu diggante, dug tandle. Sunuy maam a jélloo woon i sàkket, daan diisoo bés bu Yàlla sàkk. Sunu yaay ya, ku ca masaan a am rakk, yaak say moroom a koy xëccoo boot. Maak yow noo jàngandoo Alxuraan, daan ànd ca mbeddum xeer mooma aayoon lool ciy dàll ak sér. Bëñ sax, bu ñu ko masaan a foq, nooy bokk pax mu nu koy suul te naan: «Jinax, am bëñ bu rafet, te jox nu bëñ bu ñaaw!»Su at yi demee, nimse yi naaxsaay yit, pàttalliku yi, dara jógu fi, ñoom laay bànneexoo, ñoo tax sama àddina saf xorom. Di la fàttaliku nag; la woon lépp delsi, teewaat ci sama kanam. Ma gëmm, sama xol dekki, may yëg, ñuy dem, di ñëw: tàngaay baak leeraayu taalu matt ya, màngo xayli bu saf sàpp ak kaani, ku xàmp tàqamtiku jox sa moroom."

Figure 3 displays the corpus word cloud. This figure shows that in the corpus, stop-words such as *ci* or *ak* (meaning *in* or *with* in English) have the highest occurrences.

The corpus contains about a hundred documents covering various topics about society, religion, politics, history, agriculture, art, culture, justice, health, science, etc. The corpus contains 47457 phrases, and a total of 867951

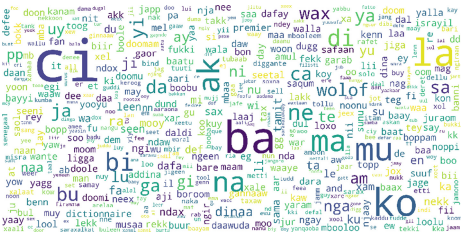

Fig. 3. Wolof corpus word cloud. Words are displayed according to their occurrence in the corpus, the more frequent, the larger the word will be displayed.

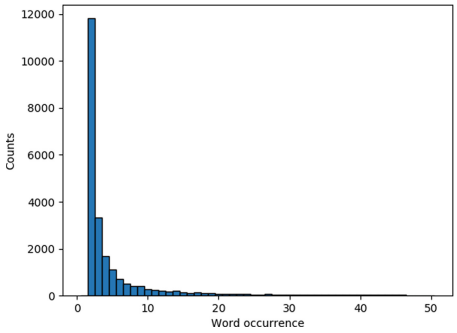

Fig. 4. Corpus word occurrence histogram depicting the number of words that have a given occurrence count.

repeated words. The vocabulary consists of 24232 unique words, with only thirty percent of the words occurring more than five times.

Figure 4 shows the histogram of the word occurrence frequencies in the corpus following the classical Zipf law: most words occur rarely, and few words, such as stop-words, occur very frequently [15].

5 Experiments

We have built three word embedding models based on the aforementioned algorithms using the Wolof corpus presented in Sect. 4. The raw texts of the corpus are preprocessed by lower-casing words, and removing accents. We restricted the vocabulary only to words occurring more than five times. We also added to the vocabulary an unknown word token to account for words in the corpus ignored in the restricted vocabulary. We then extracted phrases, and for each phrase, we build a training corpus by gathering the target words along with their contexts. Our models use the following hyper-parameters: the vocabulary size, the context size, and the embedding dimension. The vocabulary involved 7942 words comprising the unknown word token. The context size was set to 10 words. More precisely, around each target word, five words to the left and

to the right are taken to be the context. We experimented various context sizes (6, 8), however, 10 was the best experimental settings. To set the embedding space dimension we tested many values ranging from 100 to 300. Because this value did not significantly affect the final results, the embedding dimension was set to 300.

The models were implemented in TensorFlow and trained using back propagation. Training neural networks involve defining learning parameters such as the learning rate, the batch size, and the number of epochs.

For all the models, the learning rate was set to 0.001. This value ensures stable convergence of the models with a regular decrease of the optimized cost. The number of epochs was set to 10 for all the models. One reason for this is that, at around this number of epochs, the cost was no longer significantly decreasing. The batch size was set to 512 for the skip-gram and the CBOW to speed up the training: the number of samples to train was very large (i.e. million of samples) and the models were trained with local context information. For GloVe, the size of the training samples is equal to the vocabulary size. Accordingly, GloVe was trained with a smaller batch size of 32 samples.

5.1 Qualitative Analysis

Having trained the models, we qualitatively assessed the word embedding validity by selecting a sample set of words from the vocabulary and displaying their five nearest neighbours in the embedding space. Neighborhood was taken with respect to the euclidean distance. These results are provided in Tables 1, 2 and 3. In these tables, we provide Wolof words together with their English translation for indicative purposes. Our models only exploit Wolof word occurrences. The results show that the obtained neighbours are semantically related to the target words. If we consider for example the word *bànk* (meaning *bank* in English), for the CBOW model, its nearest neighbours are *leble* (to lend), *leb* (to borrow), *cfa* (acronym for financial african community francs), *kopparu* (her/his/its money), *tayle* (pledge). Using the skip-gram model, the neighbours of the word *bànk* are *dugal* (to put), *kont* (bank account), *jàngi* (to go to school), *monjaal* (worldwide), *xareñal* (to educate). For the GloVe model, neighbours of the word *bank* are *fmi* (acronym for international world fund), *kont* (bank account), *nafa* (purse), *monjaal* (worldwide), *xaalisu* (her/his/its money). Similar analysis can be conducted for other target words in Tables 1, 2 and 3. This qualitative analysis suggests that the GloVe model produces more semantically related neighbours than the other models.

We further qualitatively verified the validity of our models by training the GloVe model on a corpus composed of 350000 Wikipedia French articles. Table 4 shows that in this large-scale French corpus, words and their nearest neighbours are semantically related. For example, the first three neighbours of the word uranus (uranus in English), are *jupiter* (jupyter), *saturne* (saturn), and *pluton* (pluto). Similarly, the nearest neighbors of the word *boudhisme* (buddhism), *hindhouisme* (hindhuism), *brahmanisme* (brahmanism), *jainisme* (jainism).

Table 1. Examples of Wolof words (in bold) with their five nearest neighbours according to CBOW embedding. It should be noted that Wolof words English translation (non bold) are only given for indicative purposes. Our models only exploit Wolof words occurences.

Word	n_1	n_2	n_3	n_4	n_5
afrig	**patiriis**	**kongo**	**lumumbaa**	**reyee**	**goney**
africa	patrice	Congo	lumumba	killed	youth of
bànk	**leble**	**leb**	**cfa**	**koppar**	**tayle**
bank	to lend	borrow	cfa	money	pawn
banaana	**xollitu**	**rattax**	**roose**	**kemb**	**delluseek**
banana	peel	slippy	to water	peanut	come back with
aajo	**fajug**	**regg**	**mbaax**	**solaay**	**mànke**
need	resolve	sate	kindness	clothing	lack
bamba	**barke**	**maam**	**ibra**	**seex**	**rasululaay**
bamba	grace	grand-pa	ibra	sheikh	prophet

Table 2. Examples of Wolof words (in bold) with their five nearest neighbours according to skip-gram embedding. Note that the English translations (non bold) are only given for indicative purposes. Our models only exploit Wolof words occurrences.

Word	n_1	n_2	n_3	n_4	n_5
afrig	**oseyaani**	**asi**	**saalumu**	**sowwu**	**tugal**
africa	oceania	asia	south	west	france
bànk	**dugal**	**kont**	**jàngi**	**monjaal**	**xareñal**
bank	put in	account	go to school	world-wide	to teach
banaana	**soraas**	**màngo**	**guava**	**xollitu**	**koko**
banana	orange	mango	guava	peel of	coconut fruit
aajo	**fajug**	**aajowoo**	**faj**	**faji**	**drepanositoos**
need	resolution	want	to resolve	resolve	sickle cell disease
bàmba	**matub**	**taalubey**	**lumumbaa**	**seex**	**bijaahi**
bamba	completeness	student	lumumba	Sheikh	from his grace

5.2 Quantitative Analysis

One of the classical procedures to quantitatively evaluate word embeddings is to solve word analogy tasks. It consists in considering a set of words and a set of semantic relations, and selecting a set of word pairs where the words in each pair are related by one of the considered relations. Then, a set of questions are stated as follows: considering two pairs $(word_k, word_l)$ and $(word_n, word_m)$, the question "word k is to word l as word n is to word __?", where according to the considered relation, the expected response is word m. Because the embedding space is linear, assuming that relation between words are represented as

Table 3. Examples of Wolof words (in bold) with their five nearest neighbours according to GloVe embedding. Again, English translations (non bold) are only given for indicative purposes. Our models only exploit Wolof words occurrences.

Word	n_1	n_2	n_3	n_4	n_5
afrig	**oseyaani**	**asi**	**gànnaaru**	**sowwu**	**tefesi**
africa	oceania	asia	north	south	shore of
bànk	**fmi**	**kont**	**nafa**	**monjaal**	**xaalisu**
bank	imf	account	wallet	world-wide	money of
banaana	**soraas**	**màngo**	**guyaab**	**xob**	**kànja**
banana	orange	mango	guava	leaf	gumbo
aajo	**fajug**	**tekki**	**faju**	**lew**	**réeral**
need	resolve	mean	resolved	legal	lose
bàmba	**xaadimu**	**rasuul**	**coloniales**	**seex**	**murid**
bamba	Khadim	prophet	colonial	sheikh	mourid

Table 4. French Wikipedia GloVe words neighbours. The first column gives the target words and the three other columns give the first, second, and third nearest neighbours. These sample results show that neighbours are semantically related to the target word.

Target word	n_1	n_2	n_3
atom	atomes	isotope	cathode
mathématique	mathematiques	axiomatique	probabilites
art	contemporain	deco	abstrait
peinture	figurative	picturaux	picturales
agriculture	arboriculture	cerealieres	cerealiere
boudhisme	hindouisme	brahmanisme	jainisme
uranus	jupiter	saturne	pluton
planete	extraterrestre	lointaine	orbitant
mer	caspienne	baltique	ocean
fleuve	baikal	fleuves	embouchure

translations, we expect in the embedding space linear relation word vectors of the form

$$E_l - E_k = E_m - E_n \tag{5}$$

where we remind that E_k, E_l, E_m, E_n are the embedding vectors of words k, l, m, n. Questions about relations between words can be stated as finding the word $\hat{u}$ verifying:

$$\hat{u} = \arg\min_u ||E_l - E_k - E_u + E_n||_2 \tag{6}$$

where $||.||_2$ is the D-dimensional euclidean distance. Then, if the embedding vector of word m is among the ten word vector closest to $E_l - E_k + E_n$ we consider the system has provided the correct answer, otherwise the response is considered false.

To quantitatively assess the effectiveness of our models, we measured to which extent, for a given selected quadruple words k, l, m, n, one of the selected semantic relations is verified. We mainly focus our analysis on relations such as, male-female, derivative, synonym, country-capital. Table 5 gives the selected relations and word pairs, and the performances achieved by each model of the benchmark.

Table 5. Word analogy tasks and scores of the CBOW, skip-gram, and GloVe models. The first column specifies relations to be discovered from Wolof word pairs given in the second column. The third column gives the English translation of the word pairs. The fourth, fifth, and sixth columns give the models' results on the task (1 for correct answers and 0 otherwise).

Relations	Word pairs (Wolof)	English translation	CBOW	SG	GloVe
Country-capital	(senegaal, dakaar)	(senegal, dakar)	1	0	1
Country-capital	(faraas, pari)	(france, paris)	0	0	0
Male-female	(janq, waxambaane)	(girl, boy)	0	0	1
Male-female	(jigéen, góor)	(female, male)	0	0	0
Male-female	(yaay, baay)	(mother, dad)	1	1	0
Male-female	(jëkkër, jabar)	(husband, wife)	0	0	0
Synonym	(rafet, taaru)	(pretty, beautiful)	1	1	1
Synonym	(teey, yem)	(prudent, cautious)	1	1	1
Synonym	(tàmbale, sumb)	(to start, to begin)	1	1	1
Synonym	(metit, naqar)	(pain, grief)	0	0	1
Synonym	(suux, diig)	(sink,)	1	1	1
Synonym	(taarix, cosaan)	(history, story)	1	1	1
Derivation	(xam, xami)	(know, to know)	0	1	1
Derivation	(ajoor, kajoor)	(cayor resident, cayoor)	0	1	1
Derivation	(jàng, jàngale)	(to learn, to teach)	0	0	1
Model performances			47%	53%	73%

The scores in Table 5 show that among the tested models, GloVe achieves best performances, followed by the skip-gram model. This corresponds to the conclusions of the qualitative analysis conducted in Sect. 5.1.

6 Conclusion

In this paper, we presented preliminary investigations we conducted to build vector representations of Wolof words from a corpus gathered from the internet. We considered three word embedding models: the CBOW, the skip-gram, and the GloVe. Our experiments demonstrate that the models are able to build effective representations that encode semantic relations between words in the embedding spaces. These relations can be recovered using euclidean distance between word embedding vectors.

In the future, we plan to extend the corpus in two directions. First, we are looking forward to significantly enlarge the corpus from 47000 to 200000 phrases. Also, we will make the corpus bilingual by pairing the Wolof phrases with their corresponding French translations. This dataset will be used to build neural machine translation models.

References

1. AAI State of Education in Africa Report 2015. http://www.aaionline.org/wp-content/uploads/2015/09/AAI-SOE-report-2015-final.pdf
2. Bengio, Y., Ducharme, R., Vincent, P., Janvin, C.: A neural probabilistic language model. J. Mach. Learn. Res. **3**, 1137–1155 (2003)
3. Collobert, R., Weston, J.: A unified architecture for natural language processing: deep neural networks with multitask learning. In: Proceedings of the ICML (2008)
4. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. In: ICLR (2013)
5. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J.: Distributed representations of words and phrases and their compositionality. In: NIPS (2013b)
6. Pennington, J., Socher, R., Manning, C.: Glove: global vectors for word representation. In: EMNLP (2014)
7. Arora, S., Li, Y., Liang, Y., Ma, T., Risteski, A.: A latent variable model approach to PMI-based word embeddings. Trans. ACL **4**, 385–399 (2016)
8. Dione, C.B.: LFG parse disambiguation for Wolof. J. Lang. Model. **2**(1), 105–165 (2014)
9. Dione, C.B.: Valency change and complex predicates in Wolof: an LFG account. In: LFG Conference (2013)
10. Dione C.B.: An LFG approach to Wolof cleft constructions. In: LFG Conference (2012)
11. Khoule, M., Thiam, M.N., Nguer, E.M.: Towards the establishment of a LMF-based Wolof language lexicon. Traitement Automatique des Langues Africaines (TALAf) (2014)
12. Pauw, G.D., Wagacha, P.W., de Schryver, G.-M.: Towards English - Swahili machine translation. In: Research Workshop of the Israel Science Foundation (2011)
13. Ombui, E.O., Wagacha, P.W., Ng'ang'a, W.: InterlinguaPlus machine translation approach for under-resourced languages: Ekegusii & Swahili. In: Workshop on the Use of Computational Methods in the Study of Endangered Languages (2014)
14. Gebreegziabher, M., Besacier, L.: English-Amharic statistical machine translation. In: Workshop on Spoken Language Technologies for Under-Resourced Languages (2012)
15. Sichel, H.S.: On a distribution law for word frequencies. J. Am. Stat. Assoc. **70**, 542–547 (1975)
16. Pathe, D.: Grammaire de wolof moderne, Edition Presence Africaine (1971)
17. Cisse, M.T., Diagne, A.M., Campenhoudt, M.V., Muraille, P.: Mise au point d'une base de données lexicale multifonctionnelle : le dictionnaire unilingue wolof et bilingue wolof-français. Journées LC (2007)
18. Young, T., Hazarika, D., Poria, S., Cambria, E.: Recent trends in deep learning based natural language processing. arXiv:1708.02709 (2018)

19. Wild, F., Stahl, C.: Investigating unstructured texts with latent semantic analysis. In: Decker, R., Lenz, H.-J. (eds.) Advances in Data Analysis. SCDAKO, pp. 383–390. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-70981-7_43
20. Ba, M.: So long a letter. Nouvelles Editions Africaines. https://en.wikipedia.org/wiki/So_Long_a_Letter (1979)