Named Entity Recognition in Low-resource Languages using Cross-lingual distributional word representation.

M. F. MBOUOPDA, P. MELATAGIA YONTA

Department of Computer Sciences
University of Yaounde I, Cameroon
IRD, UMMISCO, F-93143, Bondy, France
mf.mbouopda@gmail.com, paulinyonta@gmail.com

ABSTRACT. Named Entity Recognition (NER) is a fundamental task in many NLP applications that seek to identify and classify expressions such as people, location, and organization names. Many NER systems have been developed, but the annotated data needed for good performances are not available for low-resource languages, such as Cameroonian languages. In this paper we exploit the low frequency of named entities in text to define a new suitable cross-lingual distributional representation for named entity recognition. We build the first Ewondo (a Bantu low-resource language of Cameroon) named entities recognizer by projecting named entity tags from English using our word representation. In terms of Recall, Precision and F-score, the obtained results show the effectiveness of the proposed distributional representation of words

RESUME. La reconnaissance des entités nommées (REN) est une tâche fondamentale du TALN dont le but est d'identifier les expressions telles que les noms de personnes, de lieux et d'organisations dans un texte. Il existe de nos jours plusieurs systèmes de REN, cependant les données nécessaires pour les utiliser dans le traitement des langues peu dotées telles que les langues camerounaises ne sont pas disponibles. Nous exploitons le fait que les entités nommées apparaissent rarement dans les textes pour définir une nouvelle représentation distributionnelle interlingue des mots, qui soit adaptée à la REN. En utilisant notre représentation, nous projectons les entités nommées de l'anglais vers l'ewondo (une langue bantou du Cameroun); nous obtenons donc le tout premier modèle de reconnaissance des entités nommées en langue ewondo. Les résultats en terme de précision, rappel et f-mesure montrent l'efficacité de notre représentation

KEYWORDS : Natural Language Processing, Named Entity Recognition, Neural Network, Annotation Projection, Low resource language, Cross-lingual

MOTS-CLES : Traitement Automatique du Langage Naturel, Reconnaissance des Entités Nommées, Réseau de neurones, Projection d'annotations, Langue peu dotée, Inter-linguistique

Special Issue CRI 2019
Maurice Tchuenté, Paulin Melatagia Yonta, Eric Badouel, René Ndoudam, Eds.

ARIMA Journal, Vol. 33, (2020), pp. 1-11

1. Introduction

Expressions such as people names, organization names and location names are so important in almost all Natural Language Processing (NLP) applications so that there is a dedicated task that focuses on them. This task is called Named Entities Recognition (NER) and its goal is to firstly identify those expressions and secondly classify them.

NLP aims to equip computers with the capability to communicate with humans in natural way, that means using human language. As NLP is important for resource-rich languages such as English, French, or German, it is also important for low-resource languages. A language is said to be low-resource when there is not (eventually not enough) annotated corpora, name dictionaries, good morphological analyzers, POS taggers, treebank, ... in that language. Ewondo is one of the most studied (by linguists) language in Cameroon and, as some Cameroonian languages, it is already taught in high schools and specialized training centers; however, it is a resource poor language.

Nowadays, machine learning (ML) approaches are popularly used in NER because they are easily trainable and are adaptable to different languages while rule-based approaches lack the ability of copying with the problem of portability and robustness. However the most performant ML approaches use supervised learning algorithm [3]. Hence they cannot be used to process low-resource languages because annotated data needed for learning are not available. More recently, new NLP systems based on cross-language projection of linguistic annotations from a resource-rich language to a low-resource language were developed and they achieve good performances [17]. The key idea of projection of linguistic annotations can be summarized as follows: through word alignment in parallel text corpora, the annotations are transferred from the source (resource-rich) language to the target (under-resourced) language, and the resulting annotations can be used for a supervised training in the target language. However, the poor performance of word alignment algorithms constitutes a principal limitation of approaches based on projection of linguistic annotations. As many Cameroonian languages, Ewondo is a morphologically rich language compared to English and French since it has agglutinative or inflectional morphology. In such languages, the production of hundreds of words from a given root is possible, which results in the data sparsity problem. The first consequence of this property is that a word in Ewondo can be translated to a sequence of words in English and hence English-Ewondo word alignment algorithms would perform poorly. For instance the Ewondo phrase "$Akore\ ya\ won\ awu$" is translated by the Glosbe platform [1] as "$Freedom\ from\ a\ morbid\ fear\ of\ death$".

In this paper, we present a new cross-lingual distributional representation that give very good performances for the NER task. We use this representation to efficiently project NER tags from English to Ewondo using a model inspired from [19]. The model would perform as well in any other language since the projection is done without any word alignment, without any knowledge about the target language and there is no need to choose a similarity measure.

The rest of this paper is as follows: Section 2 presents related work; Section 3 presents the Zennaki et al. [19] POS tagging model, adapted for NER. The proposed distributional representation is described in Section 4. The details of experiments and results are reported in Section 5 and Section 6 concludes this paper.

---

1. https://glosbe.com/en/ewo/death

2. Related work

Regarding the target language, named entity recognition methods can be classify in two groups: the first group contains methods for resource-rich languages and the second one is for methods that are applicable on low-resource languages.

2.1. Named entity recognition for resource-rich language

State of the art of named entity system for resource-rich language rely heavily on hand-crafted features which are domain and language-specific [7]. A Hidden Markov Model (HMM) based English named entities recognizer in which the basic input feature is character instead of word as it is most of the time have been constructed by [4]. This charater-level HMM got 83% of F-score on the CoNLL-2003 English test set which is better than the corresponding word-level HMM by a difference of about 10%. A Support Vector Machine (SVM) has been used to construct a named entity recognizer with hundreds of thousands of features [11] that give an F-score of 84%. The goal was to be able to use many features without overfitting and without high computational cost. As for other SVM models of NER found in the state of the art, it uses different contextual information (surrounding words) along with the variety of orthographic word-level features. SVM based NER models are more efficient than HMM to deal with the non-independent and overlapping features of the language. Most recently, researchers proposed methods in which the effort needed for feature engineering is considerably reduce. For instance a neural architecture that automatically detect word and character-level features using hybrid bidirectional LSTM and CNN architecture [1], eliminating the most feature engineering.

Because these methods are language-specific and need large amount of annotated data to perform, they can not be used for low-resource languages.

2.2. Named entity recognition for low-resource languages

For low-resource languages, many works are based on annotation projection. Introduced in 2001, annotation projection is the process of transferring annotations from one language to another language [17]. For instance, in order to build the first named entity annotated corpus for Malay, named entity tags have been projected from English (resource-rich) to Malay (low-resource) language with an F-score of 90% on their test set [18]. Based on a Malay-English align corpus, their method can be resume is three steps: they have firstly annotated the English part of the corpus using UIUC named entity tagger [14]. Secondly they have used a Malay-English dictionary to align English and Malay words according to a bi-gram Dice coefficient. Finally, they projected named entity tags using the word alignment obtained previously; however, they applied some language-specific rules (mostly regular expressions) for more confidence. One of the most important difference between this method and the process we are proposing in this paper is that no assumption on the target language is required and hence no language-specific rule is used.

A multilingual NER has been built by projecting entity tags from English to French, Spanish, German, Czech and Russian with 94% of recall [3]. After the annotation of the English side of the corpus with an in-house NER system [15], they used a phrase-based statistical machine translation [6] to align named entities found in the English side with their corresponding portion of text on the target side. They used three different methods for projection, starting from the most rigorous to the most tolerant. They started by the string matching and if it fails, they continue with consonant signature matching. For cases

where consonant signature return no result, they apply a Levenshtein edit distance on consonant signature and projection is applied when this distance is greater or equal to 0.7. The approach we are using in this paper is different as it uses a neural network that learn itself the best similarity measure, and no word alignment between language is required. Furthermore, it is not easy to get a phrase-based machine translation for resource-low languages.

Distributional regularities in monolingual data, projected information across closely related languages and human linguist judgments have been exploited to quickly build a NER system for Sorani Kurdish and Tajik resource-low languages of Western Iranian [8]. They wanted to build a usable NER system in very few time, although not perfect. Instead of relying on elaborated and robust linguistic resources built ahead of time, they opportunistically incorporate linguistic theory, informal and non-expert intuitions about the language and the target application, and resources adapted from closely related languages while avoiding extensive manual annotation. Non-experts are linguists that do not speak the target languages but are familiar with the structure of human languages in general. They have some knowledge about the family of target languages, and they can rapidly have intuitions about the languages by reading reference grammar and exploring data. The approach used in this paper makes no assumption about the target language or its family.

Using a set of language-independent features grouped in lexical, non-lexical, gazetteer and wikifier features, a cross-lingual NER trained only on English have been built [16]. The model can be used to perform NER in any language on Wikipedia. The strength of the model is that it only requires multilingual Wikipedia, no sentence-aligned or word-aligned parallel text is needed, however the model perform a word-alignment in background in order to align words in the target language to English word.

Mayhew et al. [12] used Cheap translation to project named entities from one language to another. Because lexicons are much cheaper and easier obtainable than parallel corpus [10], they decided to rely their translation only on a lexicon instead of a parallel corpus as in [6]. [2] trained a conditional random field model for NER in low-resource language by augmenting the small amount of data in the low-resource language with data coming from a set of rich languages related to the target low-resource language. As stated by the authors, this approach is different from annotation projection since it requires the low-resource language data to be annotated.

State of the art named entity tags projection methods have two main sources of error in common: word alignment and similarity measure. Solving those issues would lead to more accurate projection. Furthermore, making no assumption on the target language is better for generalization. In the next section we present a model that overcomes these limitations for the Part Of Speech Tagging (POS Tagging) task and that we are going to modify in Section 4 for NER. The modification is based on a new cross-lingual distributional representation. There is no report of named entities recognition work prior to our experiment for Ewondo and more generally for any Cameroonian language. Moreover, the proposed approach is, in our knowledge the first attempt to tackle NER for low-resource languages by projecting annotations without word alignment nor similarity measure, and in which no language-specific feature is used.

3. The Zennaki et al. model

Part Of Speech Tagging is a task of NLP that aims to assign to each word in a text its part of speech. Zennaki et al. [19] used a Recurrent Neural Network (RNN) to projects 12 part of speech tags from English to French, German, Spanish and Greek. The model is based on a simple principle stipulating that if we can find a space where a word in the source language and its translation in the target language have the same representation then an artificial neural network can learn from words in the source language and make predictions on the words in the target language. Since a word and its translation will always appear in the same bi-phrase in a parallel corpus, they represent each word w in the bi-text (parallel corpus) by a binary vector V_w of size N (the number of bi-phrases in the bi-text) called *distributional signature* where:

$$V_w[i] = \begin{cases} 1 & \text{if } w \text{ appears in the } i^{th} \text{ bi-phrase,} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The neural network architecture is described as follow:

– the input layer has N neurons,

– there is a first hidden layer with 640 neurons followed by the second hidden layer which contains 160 neurons,

– the output layer has 12 neurons and softmax is applied,

– the activation function of every neuron in the network is sigmoid

– and there is a feedback loop from the first hidden layer to the input layer.

The model has been tested using a multilingual corpus extracted from Europarl [5] and achieved more than 75% of accuracy for the POS tagging. The model suffers of two main problems: firstly, words that are not in the vocabulary are not represented and are tagged just according to their context; secondly, the memory and time needed to use the model increase too much with the size of the bi-text used. However the model has three great advantages:

1) no need to choose a similarity measure, the best one is implicitly learned by the RNN,

2) the same model can be used for many languages without any change. Indeed the only thing to do is to translate the resource-rich language text used for training in the new language,

3) the model is trained only on the source language (resource-rich language).

We have exploited these properties to build a named entity tagger for low-resource languages. This tagger is described in the next section.

4. Named entities tagger for low-resource languages

Transferring knowledge from one language to another requires the ability to transfer features from the source language to the target language; however features such as surrounding words, prefixes or suffixes are not easily transferable between languages because of their syntactical differences. Hence, the more the source and the target languages are linguistically similar, the better is the projection.

The distributional representation described in [19] is very impressive because it is not based on any language, we say it is language-free. Our model also uses distributional signature to represent words, but it is not binary. Indeed, named entity expressions are less frequent in text than expressions that are not named entities; considering this observation, we believe that using word frequency will lead to a better distributional representation for named entities. The *term frequency* or the frequency of a word in a document is the weight of that word in the document, computed according to its number of occurrences in the document [9]. This is a powerful notion, widely used in information retrieval. The frequency of a word in a document is the number of its occurrences divided by the total number of word in the document. It is a way to measure the importance of that word in the document. To reuse this notion, we have to decide what we call a document in our context. We have identify three possibilities:

– *Corpus-based term frequency* where the corpus (the bi-text) is considered as being a whole document. In this case, the term frequency is language independent and we will have a single document.

– *Side-based term frequency* that is only based on the words from the same language. Here, each side of the bi-text is a document, hence we have two documents: The rich side and the low-resource side.

– *Phrase-based term frequency* where each phrase in considered as being a document. Here, a word has as much frequencies as there are phrases in the corpus.

Given the definition of the document, the term frequency of a word w is given as follows:

$$tf_w = \frac{nbOcc(w)}{nbWC} \tag{2}$$

where $nbOcc(w)$ is the number of occurrences of w in the document containing it and $nbWC$ is the total number of words in that side.

In information retrieval, it is sometimes a good idea to normalize the term frequency. We used the following normalizations [9]:

1) **natural**: $norm(tf_w) = tf_w$

2) **binary**: $norm(tf_w) = 1$ if $tf_w > 0$, and 0 otherwise

3) **logarithmic**: $norm(tf_w) = 1 + \log(tf_w)$

4) **maximal**: $norm(tf_w) = 0.5 + 0.5 \times \frac{tf_w}{\max(\{tf_m | \forall m \in document\})}$

5) **log average**: $norm(tf_w) = \frac{1 + \log(tf_w)}{1 + \log(avg(\{tf_m | \forall m \in document\}))}$

Given the previous definition of the frequency of a word, we define a general cross-lingual distributional representation of a word w as being a vector V_w, of size N and with each component defined as follows:

$$V_w[i] = \begin{cases} norm(\frac{1}{tf_w}) & \text{if } w \text{ appear in the } i^{th} \text{ bi-phrase,} \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Hence, the distributional representation used by [19] is the side-base term frequency coupled with the binary normalization. This representation does not take advantage of the observation that named entities are less frequent in text than non name entities. Using natural, logarithmic, maximal or log average normalization, it would be easier to distinguish between name entities and non named entities since all these alternatives clearly take the word frequency (or more precisely the inverse term frequency) into account. Using one of these four representation, the distributional representation (given by Equation 3) of a

named entity is more likely to be further from the null vector; while the distributional representation of a non named entity is more likely to be closer to the null vector. Therefore, a binary classifier can be easy trained to differentiate named entities from non-named entities.

Our approach is based on a feedforward neural network with two hidden layers, more precisely the architecture is as follow:

– the number of neurons in the input layer is set to the size of distributional signature (the number of bi-phrases in the bi-text). For experiments described in Section 5 it is 210 neurons,

– the first hidden layer has 640 neurons each using sigmoid activation function,

– the second hidden layer has 160 neurons and use sigmoid activation function,

– the output layer has four neurons for our four named entity classes PERSON, OR-GANIZATION, LOCATION and MISCELLANEOUS plus a neuron for the non named entity class OTHER.

– the output is normalized using softmax function.

We have an empirical guess of the internal behavior of our model: some of its first neurons behave like a binary classifier to check if the input is a named entity or not. If the answer is "NO" the model outputs OTHER. If the answer is "YES", the remaining neurons are used to identify the specific named entity class of the input. In the next section, the experimental evaluation of our model is presented.

5. Experimental Results

To test our model, we need a parallel corpus containing the target low-resource language aligned with a rich language. The alignment must be phrase-based and not word-based. The corpus must have named entity annotations at least on the rich side. As far as we know, such a dataset do not exists in the literature; hence we manually created one using English as the rich side and Ewondo, a Bantu language of Cameroon as the low-resource side. Data are extracted from an online bible[2]. For each verse we have an English version and an Ewondo version. Each side of the bi-text has been manually annotated using Brat[3] and the help of an Ewondo native speaker. The named entities tagset used is described in Table 1 and the statistic of the bi-text is presented in Table 2, Table 3 and Table 4. As shown in Table 2 the number of words in Ewondo is not the same as in English; there is no bijection between English and Ewondo, hence word alignment might not be suitable here. Table 3 and Table 4 confirm the observation that named entities are less frequent in text than other expressions; more precisely the proportion of named entities is only 14% in the whole text and about 10% when considering only the vocabulary.

We used only the vocabulary to train and test our model. The training set is the rich side (RS) of the corpus, and the test set is the low-resource side (LRS). We trained the model over 10 epochs using 10-fold cross validation. We stop the training if the validation loss is the same between two consecutive epochs. In order to assess the robustness of our model to random initialization of the network weights, the model is trained 10 times with different random initialization of the weights. We reported the mean and standard

---

2. `https://www.bible.com`
3. `http://brat.nlplab.org`

NE tags	Meaning	Examples
PER	a person name	Yesus → Jesus
ORG	an organization name	bësadukea → Sadducees
LOC	a location name	Yerusalem → Jerusalem
MISC	others named entities	Mfufub Nsisim → Holy spirit
O	non named entities	kòbò → to speak

Table 1 – Named entities tagset

	English	Ewondo
# of sentences	210	210
# of words	4753	4185
# of distinct words	912	1029

Table 2 – Corpus statistics

	PER	ORG	LOC	MISC	O
English	5.87%	0.02%	0.95%	2.51%	90.62%
Ewondo	7.08%	0.05%	1.09%	2.68%	89.07%

Table 3 – Named entities distribution in the whole text

	PER	ORG	LOC	MISC	O
English	8.77%	0.10%	1.95%	1.84%	87.32%
Ewondo	8.18%	0.19%	1.82%	1.15%	88.64%

Table 4 – Named entities distribution in the vocabulary

deviation of precision, recall and F_1-score. The source code and the dataset we used are available here [4]

We have built our baseline by performing the projection using the distributional signature proposed by [19]. The network architecture is the feed forward neural network described in Section 4. The result are reported in Table 5.

	Precision	Recall	F_1-Score
Train set (English)	80.20% ± 3.49	56.53% ± 2.78	69.09% ± 3.3
Val. set (English)	78.79% ± 4.02	59.15% ± 3.23	72.03% ± 1.48
Projection (Ewondo)	69.99% ± 2.67	47.83% ± 2.16	60.16% ± 2.59

Table 5 – Zennaki et al. Model performance (average on 10 runs)

For each term frequency definition (phrase, side or corpus-based) and for each normalization (natural, logarithmic, maximal and log average) presented in Section 4, we have run our model and reported the precision, recall and F1-score in Table 6. This table shows in bold the highest precision, recall and F1-score for each definition of the term frequency; the lowest performances are underlined.

Regardless of the definition of the term frequency, the best results are obtained using the natural normalization. In fact, with the phrase-based term frequency, the projection is

_____

4. https://github.com/frankl1/Word2vec-For-NER-In-Low-Resource-Languages

performed with a precision of about 75%, a recall of about 67% and a F1 score of about 70%; comparing to the baseline, this is an improvement of about 6% for the precision, 20% for the recall and 10% for the F_1-score. The worst results are given by the log average normalization, and is even worse than the baseline.

Using the side-based term frequency leads to a slightly better projection comparing to the phrase-based term frequency. By the way, the projection improves up to about 78%, the recall up to about 67% and the F_1-score up to about 71%. It is also important to notice that the worst results here (obtained using the maximal normalization) is very closed the the baseline. This is because the side-based term frequency is more likely to give the same representation to a word and its translation and hence produce a better cross-lingual representation than the phrase-based term frequency

The corpus-based term frequency also produced better results than the phrase-based. These results are very closed those produced by the side-base term frequency. Indeed, the projection precision is 78%, the recall is 64% and the F_1-score is about 70%. In fact, using the corpus-based term frequency is also more likely to give very similar distributional representations to a word and its translation.

In any case, experimental results validate our hypothesis: taking into account the low frequency of named entities is important for the projection of named entity tags. Furthermore, performances on the test set (projection) show that unlike the model from [19], the proposed model performs pretty well as an English named entity annotator. In fact, the validation set is classified with a precision of about 85%, a recall of about 80% and an F_1-score of about 80%. Instead of the feed forward neural network, we have also run experimentations with the recurrent neural network from [19] and the performance was not significantly different. Therefore, we think the knowledge about the past (coming from the recurrence) is too much different from English to Ewondo.

6. Conclusion

In this paper we have proposed a way to quickly construct a named entity annotator for low-resource languages using cross-lingual distributional representation. We have defined a generalized cross-lingual representation based on the observation that named entities are less frequent in text than non named entities. We have coupled the proposed representation with a 2 hidden layers network to project named entities from English to Ewondo (a Bantu language of Cameroon). We have done several experiments with different instances of the proposed cross-lingual representation. In each case the model performs well on the used dataset. More precisely, we have projected named entity tags from English to Ewondo with a precision of 78.15%, a recall of 67.29% and an F1-score of 71.23%. We think that the proposed distributional signature can be used in any classification task where the classes of interest are more frequent (or less frequent) than the class of other expressions. We will continue to verify this hypothesis on other NLP tasks. Moreover, we intent to assess our model on a larger corpus, eventually from different domains. Elsewhere, we will try to reduce the size of our representation using cross-lingual word embedding techniques.

	Normalization	Split	Precision	Recall	F_1-Score
Phrase-based	natural	Train set (RS)	87.13% ± 0.98	83.59% ± 1.41	85.11% ± 0.41
		Val. set (RS)	84.79% ± 1.49	80.09% ± 1.25	81.46% ± 0.69
		Projection (LRS)	**75.72% ± 1.65**	**67.18% ± 2.34**	**70.72% ± 1.38**
	logrithmic	Train set (RS)	85.89% ± 1.43	72.07% ± 3.70	77.22% ± 2.56
		Val. set (RS)	78.04% ± 3.63	67.57% ± 4.66	75.66% ± 3.15
		Projection (LRS)	74.60% ± 1.46	57.82% ± 3.45	63.63% ± 2.66
	maximal	Train set (RS)	79.24% ± 4.24	57.80% ± 1.38	68.12% ± 3.20
		Val. set (RS)	75.92% ± 4.24	58.69% ± 3.14	71.81% ± 2.99
		Projection (LRS)	68.73% ± 3.09	49.14% ± 1.30	59.44% ± 3.00
	log avg	Train set (RS)	79.24% ± 4.24	69.15% ± 2.55	73.19% ± 2.06
		Val. set (RS)	80.76% ± 5.70	65.68% ± 5.66	73.36% ± 4.55
		Projection (LRS)	61.97% ± 2.27	28.87% ± 2.18	37.75% ± 2.07
Side-based	natural	Train set (RS)	87.35% ± 0.99	83.81% ± 2.80	84.50% ± 2.40
		Val. set (RS)	85.63% ± 2.6	81.44% ± 1.45	82.29% ± 1.18
		Projection (LRS)	**78.15% ± 1.61**	**67.29% ± 1.95**	**71.23% ± 1.89**
	logrithmic	Train set (RS)	85.58% ± 1.07	77.76% ± 2.33	80.89% ± 1.31
		Val. set (RS)	83.86% ± 2.70	75.17% ± 2.44	78.18% ± 1.86
		Projection (LRS)	73.96% ± 1.29	65.12% ± 1.88	68.60% ± 0.96
	maximal	Train set (RS)	74.44% ± 1.65	55.89% ± 4.23	69.13% ± 2.96
		Val. set (RS)	77.00% ± 3.42	53.04% ± 4.16	65.78% ± 3.19
		Projection (LRS)	66.20% ± 1.47	47.58% ± 3.91	59.80% ± 2.97
	log avg	Train set (RS)	84.26% ± 4.00	75.62% ± 3.01	80.13% ± 2.43
		Val. set (RS)	79.69% ± 2.47	72.03% ± 4.98	77.09% ± 4.03
		Projection (LRS)	71.83% ± 3.40	63.16% ± 2.78	67.37% ± 2.30
Corpus-based	natural	Train set (RS)	88.26% ± 1.08	83.34% ± 2.50	85.21% ± 1.85
		Val. set (RS)	85.42% ± 1.60	78.77% ± 2.37	80.70% ± 1.28
		Projection (LRS)	**78.68% ± 1.25**	**64.95% ± 2.02**	**70.52% ± 1.53**
	logrithmic	Train set (RS)	80.78% ± 4.30	73.37% ± 3.01	80.13% ± 2.23
		Val. set (RS)	76.98% ± 2.06	72.45% ± 2.50	79.26% ± 3.15
		Projection (LRS)	68.86% ± 3.42	62.47% ± 2.69	80.13% ± 2.23
	maximal	Train set (RS)	72.68% ± 5.02	51.08% ± 2.22	68.14% ± 3.36
		Val. set (RS)	68.65% ± 2.55	52.51% ± 2.52	72.29% ± 2.50
		Projection (LRS)	61.27% ± 2.56	43.59% ± 1.83	61.93% ± 3.09
	log avg	Train set (RS)	86.50% ± 2.07	79.27% ± 2.43	82.09% ± 1.3
		Val. set (RS)	82.44% ± 3.35	76.13% ± 3.49	77.53% ± 2.30
		Projection (LRS)	73.87% ± 2.05	67.50% ± 2.00	69.87% ± 0.89

Table 6 – Our Model performance using phrase-based term frequency (average on 10 runs)

7. References

[1] J. Chiu and R. Nichols 2015. "Named Entity Recognition with Bidirectional LSTM-CNNs", *TACL* vol. 4: 357-370

[2] R. Cotterell and K. Duh 2017. "Low-Resource Named Entity Recognition with Cross-Lingual, Character-Level Neural Conditional Random Fields", *InProceedings of the Eighth International Joint Conference on Natural Language Processing* vol. 2:91-96.

[3] M. Ehrman, M. Turchi, and R. Steinberger 2011. "Building a Multilingual Named Entity Annotated Corpus Using Annotation Projection", *Ranlp*: 118-124.

[4] D. Klein, J. Smarr, H. Nguyen and C. Manning 2003. "Named Entity Recognition with Character-Level Models", *Conference on Natural language learning at HLT-NAACL* vol. 7: 180-183.

[5] P. KOEHN 2005. "A Parallel Corpus for Statistical Machine Translation", *The MT Summit* vol. 5: 79-86

[6] P. KOEHN, F. OCH, AND D. MARCU 2003. "Statistical phrase-based translation", *Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - NAACL* vol. 1: 48-54.

[7] G. LAMPLE, M. BALLESTEROS, S. SUBRAMANIAN, K. KAWAKAMI AND, C. DYER 2016. "Neural Architectures for Named Entity Recognition", *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* : 260-270

[8] P. LITTELL, K. GOYAL, D. MORTENSEN, A. LITTLE AND C. DYER 2016. "Named Entity Recognition for Linguistic Rapid Response in Low-Resource Languages : Sorani Kurdish and Tajik", *COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers* vol. 26: 998-1006

[9] C. D. MANNING, P. RAGHAVAN AND H. SCHÜTZE 2008. "Scoring, term weighting and the vector space model", in *Introduction to Information Retrieval*, Cambridge University Press.

[10] MAUSAM, S. SODERLAND, O. ETZIONI, D. WELD, K. REITER, M. SKINNER, M. SAMMER AND J. BILMES 2010. "Panlingual lexical translation via probabilistic inference", *Artificial Intelligence* vol. 174: 619-637

[11] J. MAYFIELD, P. MCNAMEE AND C. PIATKO 2003. "Named Entity Recognition Using Hundreds of Thousands of Features", *Conference on Natural language learning at HLT-NAACL* vol. 7: 184-187.

[12] S. MAYHEW, C. TSAI AND R. ROTH 2017. "Cheap Translation for Cross-Lingual Named Entity Recognition", *Conference on Empirical Methods in Natural Language Processing*: 2536-2545.

[13] D. NADEAU 2007. "A survey of named entity recognition and classification", *Linguisticae Investigationes* vol. 30: 3-26.

[14] L. RATINOV AND D. ROTH 2009. "Design Challenges and Misconceptions in Named Entity Recognition", *CoNLL: Conference on Computational Natural Language Learning* vol. 30: 144-155.

[15] R. STEINGERGER AND B. POULIQUEN 2007. "Cross-lingual Named Entity Recognition", *Lingvisticae Investigationes* vol. 30: 135-160

[16] C.T. TSAI, S. MAYHEW AND D. ROTH 2016. "Cross-lingual named entity recognition via wikification", *CoNLL: Conference on Computational Natural Language Learning* vol. 20: 219-228.

[17] D. YAROWSKY, G. NGAI AND R. WICENTOWSKI 2001. "Inducing Multilingual Text Analysis Tools Via Robust Projection Across Aligned Corporas", *International conference on Human language technology research* vol. 1: 1-8.

[18] N. ZAMIN, A. OXLEY, AND Z. ABU BAKAR 2013. "Projecting Named Entity Tags From Resource-Rich Language to a Resource Poor Language", *Journal of ICT* vol. 12: 121-146.

[19] O. ZENNAKI, N. SEMMAR AND L. BESACIER 2015. "Utilisation des Réseaux de Neurones Récurrents pour la Projection Interlingue d'Etiquettes Morpho-syntaxiques à Partir d'un Corpus Parallèle", *Traitement Automatique des Langues Naturelles, Caen* vol. 22: 529-36.