A
SURVEY REPORT
ON

SEQUENCE LABELLING FOR LOW-RESOURCE LANGUAGES

DEPARTMENT OF COMPUTER SCIENCE
IIT BHU, VARANASI
2019

Submitted by:
ADITYA ANAND
Roll: 17045009

Under the Guidance of
Dr. A.K. SINGH
ASSOCIATE PROFESSOR
DEPARTMENT OF COMPUTER SCIENCE
IIT BHU, VARANASI 221005

1. ABSTRACT

Developing natural language processing tools for low-resource-languages often require creating resources from scratch. While a variety of methods exist for training different learning models from complete or incomplete data, namely supervised and unsupervised learning. There are open question regarding what types of training data should be used, how much data is necessary and what kind of learning should be implemented for better performance.

This report focuses on low-resource-languages which has lesser amount of annotated or labelled corpus, with little amount of research and language processing implementations like sequence labelling. Common NLP models require large amount of training data and/or sophisticated language-specific engineering. However, such amount of data is unavailable for low resource languages and, in many cases, we cannot find a linguistically trained speaker to build a language model for many languages, even if they are spoken by millions of people, like Maithili, Punjabi, Khasi and many other.

So building NLP applications for such languages can at the same time reinforce the ties between the world and ensure its diversity. Thus this report aims to brings out the essential NLP- Sequence-Labelling techniques and compare the methodologies, toolkit, dataset, accuracies with pros and cons for these kinds of languages

2. Introduction

2.1 NLP

The goal of Natural Language Processing (**NLP**) is to build computational models of natural language for its analysis and generation. Natural language processing (NLP) is a field of computer science and linguistics concerned with the interactions between computers and human (natural) languages; it began as a branch of artificial intelligence. In theory, natural language processing is a desirable method of human-computer interaction. The tools of work in NLP are grammar formalisms, algorithms and data structures, vocabulary knowledge, reasoning mechanisms etc. Many of these have been taken inherit results from Computer Science, Artificial Intelligence, Linguistics, Logic, and Philosophy.

2.2 Pipeline of NLP

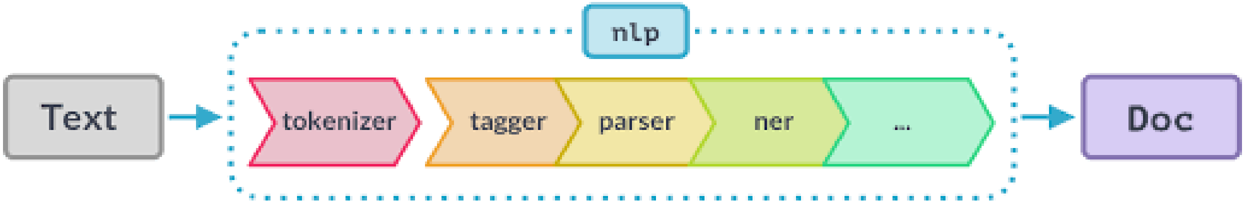

Fig. 2.1

Fig. 2.1 shows the basic pipeline of NLP. Whereas on context of this report we will only focus on Sequence Labelling Part, that include Parts Of Speech Tagger, Parser and Named Entity Recognizer.

A. **Parts Of Speech Tagger (POST) :** It, determine the part of speech for each word. Many words, especially common ones, can serve as multiple parts of speech. For example, a "book" can be a noun ("the book on the table") or verb ("to book a flight"); and "out" can be any of at least five different parts of speech. Some languages have more such ambiguity than others. Languages with little morphology, such as English are particularly prone to such ambiguity. Chinese is prone to such ambiguity because it is a tonal language during verbalization. Such inflection does not convey intended meaning so along with word sense disambiguation POST determine tags for each word.

B. **Parser:** In NLP, we mainly deals with three types of parsing:

   i. Shallow Parsing (or Chunking): It adds a bit more structure to a POS tagged sentence. The most common operation is grouping words into Noun Phrases (NP). You can also group stuff into VP (Verb Phrases) and PP (Prepositional Phrases).

ii. Constituency Parsing (or Deep Parsing): Adds even more structure to the POS tagged sentence. Such a parse is actually a tree, with words as leaves. The other nodes are part of sentence tags: NP, VP, PP etc
iii. Dependency Parsing: Probably the most popular type of parsing. It implies finding the dependencies between the words and also their type.

C. **Named entity recognition (NER)** : Also known as entity extraction , is a popular technique used in information extraction or Sequence labelling to identify and segment the named entities and categorize them under various predefined classes. Some common entity tags include PERSON, ORGANIZATION and LOCATION.

2.2 Deep Learning aspects of NLP

The majority of methods used for NLP problems employed shallow machine learning models and time-consuming hand-crafted features. This lead to problems such as curse of dimensionality. However, after the popularity and success of word embeddings, neural-based models have achieved superior results on various language-related work as compared to traditional machine learning models like logistic regression or SVM. These basic models are listed below.

1. Word2vec: It consist of CBOW and skip-gram models. CBOW is a neural approach to construct word embeddings by computing the conditional probability of a target word given the context words in a given sentnece. On the other hand, Skip-gram is a neural approach to construct word embeddings, by predicting the surrounding context words given a central target word. For both models, the word embedding dimension is determined in an unsupervised manner.

2. Character Embeddings: For tasks such as POS tagging and NER it is useful to look at morphological information in words, such as characters or combinations thereof by analyzing text at the character level, these type of embeddings help to deal with the unknown word issue. This is also helpful for morphologically rich languages such as Portuguese, Spanish, and Chinese.

3. Convolutional Neural Network (CNN): A CNN is basically a neural-based approach. sentences are first tokenized into words and then to word embedding some dimension. Then, convolutional filters are applied on this embedding layer. This is then followed by a max-pooling operation to obtain a fixed length output and reduce the dimensionality of the output. And finaly producing the final sentence representation.

4. Recurrent Neural Network (RNN): RNNs are specialized neural-based approaches has capacity to memorize the results of previous computations and use that information in the current computation. It recursively applies a computation to every instance of an input sequence conditioned on the previous computed result. RNNs have been used to study various NLP tasks such as machine translation, image captioning, and language modeling, among others.

5. RNN Variants: An LSTM consist of three gates (input, forget, and output gates), and calculate the hidden state through a combination of the three. GRUs are similar to LSTMs but consist of only two gates and are more efficient because they are less complex. They are usually picked based on the computing power available and are suitable for machine translation, text summarization, question answering or image-based language generation.

3. REPORT

The survey was done on different kind of models based on different algorithm and tool. The classifications are done on the basis of Language, tools/technique, type of sequence labelling work with pros and cons. Given below is the table of survey:

Sl. No.	Paper Title	Category	Language	Method/ Algorithm	Tools/ Techniques/ Approach	Pros/Merits/ Advantages	Cons/Demerits/ Disadvantages
1.	Cross-Lingual Transfer Learning for POS Tagging without Cross-Lingual Resources	POS TAGGING	Swedish, Danish, Dutch, German, Slovenian, Polish, Slovak, Bulgarian, Romanian, Portuguese, Italian, Spanish, Persian & Hungarian	Bidirectional Long Short-Term Memory (BLSTM) based model for both knowledge transfer and language specific representation.	Cross-lingual approach	Working on a cross lingual learning usually requires linguistic knowledge and resources about the relation between the source language and the target language but in this it has been done without ancillary resources such as parallel corpora. Also, from 14 languages the proposed model improves the POS tagging performance of the target languages without exploiting any linguistic	There are some opposite cases, which use all the tag-labelled training sentences in the target languages, and they showed mixed results. For example, the accuracy of German with the target only model is 93.31% while that of the proposed model is 93.04%. This is expected since transfer learning is effective when the target train set is small.

						knowledge between the source language and the target language.	
2.	Bootstrapping method for developing part-of-speech tagged corpus in low resource languages tagset- a focus on an African Igbo	POS TAGGING	High resource language- English Low resource language - Igbo	Bootstrapping method	Cross-lingual approach	This model aims at Transformation Based Learning (TBL) a certified machine learning approach good for small amount of data. The TBL arrangement enables the automatic transformation of English tags to their corresponding Igbo equivalents.	There are some limitation to this method such as poor accuracy of the word-alignment between the language, inconsistency in the word matching pattern due to translation shortfails.
3.	A Multi-lingual Multi-task Architecture for Low-resource Sequence Labelling.	Name Entity Recognition (NER) & POS TAGGING	FOR NAME TAGGING: Chechen, Dutch and Spanish- target language English, Russian- Source language FOR POS TAGGING: English, Dutch, Spanish, and Russian	Based on the LSTM-CNNs model	Multilingual and Multitasking Approach	As Name Tagging a target task, the model achieved 4.3% - 50.5% absolute F-score gains compared to the mono-lingual single-task baseline model.	This model can only be particularly effective for low-resources settings, when there are less than 200 training sentences for the target task.
4.	A Grounded Unsupervised Universal Part-of-Speech Tagger for Low-Resource Languages	POS TAGGING & Name Entity Recognition (NER)	POS TAGGING: English (en), German (de), French (fr), Italian (it), Spanish (es), Japanese (ja), Czech (cs), Russian (ru), Arabic (ar), Farsi (fa) and Swahili (sw),	Conditional probability model named as *Cipher Grounder.*	Multilingual and Multitasking Approach	The model describe an approach for low-resource unsupervised POS Tagging that yields fully grounded output and requires no labelled training data. Incorporating this POS tagger into a Name tagger also leads to state of art tagging performance in Sinhalese and Kinyarwanda for which nearly no labelled POS data were available.	Compared to the performance (P, R and F1) of Vanilla supervised neural model over NOUN tag only, the Cipher-Avg model robustly achieves mid-range performance (mid to low accuracy).

			NAME TAGGING: Kinyarwanda and Sinhalese				
5.	End-to-end Sequence Labelling via Bi-directional LSTM-CNNs-CRF	POS TAGGING & Name Entity Recognition (NER)	English	Combination of bidirectional LSTM, CNN and CRF	End to End approach	It is a truly end-to end model relying on no task-specific resources, feature engineering or data pre-processing. The model achieved state-of-the-art performance on two linguistic sequence labelling tasks, comparing with previously state-of-the-art systems.	The model is not that much accurate for the OOV words i.e. Out Of Vocabulary words. It shows accuracy of only 82.49% on POS Tagger and 82.69% on NER which is comparatively lesser than in-vocabulary words (IV), out-of-training-vocabulary words (OOTV) and out-of-embedding-vocabulary words (OOEV)
6.	Cross-Lingual Morphological Tagging for Low-Resource Languages	POS TAGGING	Source Language: English Target Language: Bulgarian, Czech, Danish, Dutch, Finnish, Italian, Polish, Portuguese, Slovene, Spanish and Swedish.	Model performs on par with a baseline weakly-supervised HMM	Multilingual MorphologicalApproach	Multilingual experiment on the model shows that the method performs best when projecting between related language pairs. Despite the inherently lossy projection, the model also improved the performance of the parser by +0.6 on average.	The model worked on different source target language pair but more work is required on choosing which source language to use for a low resource target language. May be use of multiple source language or additional dictionaries such as wiktionary can give better performance.
7.	Cross Lingual Transfer Learning for POS tagging	POS TAGGING	Source Language: English Target Language: Hindi	Two tags are obtained using LSTM on Hindi and CRF on English and later concatenated to predict the Label.	Cross-lingual Approach	The model has a lot of potential benefit for low resource languages like Hindi, for which annotated data is not readily available. Using tag in auxiliary feature improved the performance significantly and one can use this technique as weak supervision where large amounts of	Some of the demerit of this model is that the number of words in Hindi sentence might not be equal to the number of words in the translated version. Even if the above conditions holds, there is no guarantee that the (ith) word in the Hindi sentence would correspond

						annotated data is not available.	to the (ith) in the translated English sentence. So one to one mapping is required between two languages which is time consuming. Also the sentence level accuracies were on relatively lower side, on the training set it was 57.71% while on the test set it was 38.80%.
8.	Part-of-speech Taggers for Low-resource Languages using CCA Features	POS TAGGING	Source Language: English Target Language: BG, CS, DA, DE, ES, EU, HU, IT, NL, PT, TR	A probability-based confidence model to identify words with highly likely tag projections and use these words to train a multi-class SVM using the CCA (Canonical Correlation Analysis) features	Multilingual Approach	Experimentally, the model yields accuracy of about 85% for languages with nearly no resources available, beating a state-of the-Art of partially observed CRF formulation. In future this technique may enable to tag POS for hundreds of Low resource language altogether	The corpus used here is Bible scripts which are not exactly translated by sentences but by verses. therefore the assumption has been made that each verse in a chapter has the same meaning if the number of verses is exactly same in a same chapter. It also assumed that the whole chapters have the same meaning if the number of chapters in a book are exactly the same. Based upon these the best translation for the language is done.
9.	Real-World Semi-Supervised Learning of POS-Taggers for Low-Resource Languages	POS TAGGING	Source Language: English Target Language: Kinyarwanda (KIN) and Malagasy (MLG)	Semi Supervised Algorithm with Human generated annotation.	Additional Data (Morphological Transducer)	Very few language have enough data for standard supervised models to work. The collection of resource is time consuming and expensive for an under studied language. Including semi supervised learning approach using morphological	Supervised learning methods can provide high accuracy for part-of-speech (POS) tagging, but they perform poorly when little supervision is available. Care must be taken when drawing conclusions from small-scale annotation studies

						analyser provides better result than previous it obtained 87.15 accuracy on English and 81.9% for KIN.	such as those presented in this paper.
10.	DATA-DRIVEN PHRASING FOR SPEECH SYNTHESIS IN LOW-RESOURCE LANGUAGES	Phrase break, POS Tagger & Phrasing Model	English, Portuguese and Marathi	Phrase Break by HMM tool using Bootstrap alignment & POS Tagging by unsupervised approach using Ney-Essen clustering algorithm	Multitasking data driven approach	This method allows building models without depending on the availability of part of speech taggers, or corpus with hand annotated breaks. The model works better than the baseline model and other several model for given languages also.	However the model works better than other model for given language but overall performance was not up to the mark.
11.	Neural Architectures for Named Entity Recognition	Named Entity Recognition (NER)	English, Spanish, German & Dutch	The hybrid architecture of LSTM & CRF.	Multilingual	The models rely on two sources of information about words: character-based word representations learned from the supervised corpus and unsupervised word representations learned from unannotated corpora. the model also obtained state-of-the-art performance in NER in four languages without resorting to any language-specific knowledge or resources such as gazetteers	On these three languages, the LSTM-CRF model significantly outperforms all previous methods, but the only exception is Dutch, where the model of Gillick et al. (2015) can perform better by leveraging the information from other NER datasets.
12.	Unsupervised POS Tagging of Low Resource Language for Machine Translation	POS TAGGING	Low resource: Myanmar High Resource: English, Japanese	Novel bilingual infinite HMM approach	Bilingual (En-My) & (Ja-My) approach	The main contribution of this paper is a bilingual infinite HMM POS tagging technique to aid in the machine translation of low resource languages. From the overall results, the model can give rise to improvements in translation quality	The BLEU on My-Ja translation were lower than the baseline, but that the RIBES on the same experiment were similar to the baseline. Moreover, the BLEU and RIBES were totally different for my-en and en-my. The possible

						for EN-My, My-En language pairs.	explanation can be the sentence structure as Ja & My are SOV where as En is SVO. The SOV-SVO language pairs in our experiments (Myanmar and Japanese are SOV, whereas English is SVO)
13.	TagMiner: A Semisupervised Associative POS Tagger Effective for Resource Poor Languages	POS TAGGING	Resource-rich English Resource-moderate Hindi and Resource-poor Telugu, Tamil and Bengali	Semi-supervised associative classification method	Data Mining and Context rules based approach	The use of semi-supervised learning provides the advantage of not requiring a large high quality tagged corpus. These properties make it especially suitable for resource poor languages. This experiments on various resource-rich, resource-moderate and resource-poor languages show good performance without using any language specific linguistic information. Results also show that for smaller training data sizes this tagger performs better than state-of-the-art CRF tagger using same features as our tagger.	For test words not in the cluster the model use criterion 1 and 2. Based on that the cluster is assigned and if not able to find any suitable cluster it returns the "NOTAG". Criteria 1 : highest priority is given to the matching word of the test set from the cluster Criteria 2 : 2nd highest priority is given to the presence of matching preceding words. Criteria 3 : last priority is given to the frequency of the test word in the clusters.
14.	Conditional Random Field based part of speech tagger for Chhattisgarhi language.	POS TAGGING	CHHATTISGARHI	Conditional Random Field (CRF) based POS Tagging	Direct Approach	Chhattisgarhi is a low resource language for which POS tagger, morphological analyser and parser has not been developed. This model plays a vital role in the development of POS Tagger that is an important tool which is used to develop machine	In some cases exact categorization noun tag becomes difficult and differentiation between main verb and auxiliary verb is also an important issue.

						translation (MT) system.	
15.	Tagger Voting for Urdu	POS TAGGING	Urdu	Combination of 3 different tools for tagging: Humayoun's morphological Analyser, Urdu Shallow(SH) Parser & SVM Tool tagger	Voting Technique - Indirect Approach	The tagger was unified over different tagset data and tools available for Urdu POS tagging. And on an independent test set, the tagger clearly outperforms the other tools. Additionally, a voting scheme is also implemented through which the combined tagger reaches the accuracy of 87.98%.	The accuracy of the HUM analyser and SH parser appears to be surprisingly low even in the coarse tagset. Also the voting is done for the suitable tag and the priority is given to the SVM tagger. So the accuracy mainly depend upon the SVM tagger.
16.	Cross-Lingual Named Entity Recognition via Wikification	Name Entity Recognition(NER)	Low resource: Turkish, Tagalog, Yoruba, Bengali, and Tamil High resource: English, Spanish, German, and Dutch	Cross Lingual Approach	Cross lingual & Additional data: Wikipedia	The proposed NER model can be applied to all languages in Wikipedia. the model works over a wide range of languages in both monolingual and the cross lingual settings and show significant improvement over strong baselines.	Since this model works on all languages in Wikipedia so there is requirement of wikipedia dump for all languages. Also an analysis shows that the quality of the wikifier features depends on the Wikipedia size of the test language
17.	PD3: Better Low-Resource Cross-Lingual Transfer By Combining Direct Transfer and Annotation Projection	Sentence-level Argumentation Mining & POS TAGGING	ArgMin Language: English POS TAGGING : Source: English Target: German & French.	For ArgMin: CNN with 1-max-pooling For POS TAGGING: Bidirectional LSTM with CRF output layer.	Cross lingual & Multi tasking Approach	The model gains particularly in the small dataset scenario with 50-500 parallel sentences. This is arguably the most realistic scenario for a good portion of the world's languages, for which several dozens of parallel sentences are readily available e.g. from Bible translations consistently outperforming the two baselines it is built upon in the setting of little available parallel sentences	For POS, the projection system that uses only 50 parallel sentences suffers because the parallel corpus is tiny, getting high-quality alignments from fast-align on it is difficult because the aligner lacks statistical evidence. On the short pairs, 26% of the alignment decisions of fast-align were wrong. On the long pairs, 46% were wrong. In contrast, with 5000 parallel sentences error rates were considerably

							lower: 11% and 16%, respectively.
18.	Challenges and Issues in Developing an Annotated Corpus and HMM POS Tagger for Khasi	POS TAGGING	Khasi language	Hidden Markov Model (HMM)	Additional Data: Natural Language Toolkit.	The performance of the HMM tagger conditioned with the features has shown that it also provides good performance as reported in the literature relating to HMM POS taggers. This work for under resourced language such as Khasi bring a new initiative by annotating the corpus and developing the tagger which is limited by available resources.	The BIS tagset states that only a finite number of nouns of location, space and time that can also function as postpositions are tagged as N_NST. And in Khasi it also happens as a certain group of words can function as Noun or a preposition or as an adverb.
19.	NE Tagging for Urdu based on Bootstrap POS Learning	Name Entity Recogniti on(NER) & POS TAGGING	Urdu	Conditional Random Field (CRF) learning approach	Bootstrap Technique	The basic idea behind this work is to increase the size of the available training data by using bootstrapping so as to maximize learning for NE tagging. This work was undertaken as a precursor to achieve the objective as discussed to understand human sentiments and social behaviour through analysis of verbal communication. The proposed four stage model shows an F-Score of 68.9% for NE tagging which is much higher than that obtained by the simple two stage model. This model also overcomes the limitation imposed by the availability of limited ground truth data required	In NE tagging many of the words contain ambiguities in tagging PERSON and LOCATION. Also many of the avenues remain unexplored to further improve the performance like using bootstrapping approach too for NE data as well.

						for training a learning model.	
20.	End-to-End Korean Part-of-Speech Tagging Using Copying Mechanism	POS TAGGING	Korean	character-level seq2seq model with Input-Feeding and a Copying mechanism	End- to- End Approach	This model shown that a character level seq2seq model with Input-Feeding and a copying mechanism can achieve competitive performance with state-of-the-art approaches, including carefully tuned SVM models. This model also does not require either manual feature selection or knowledge such as a lexical-original form pattern dictionary.	In some cases the model outperform for example of POS tagging results with/without the Copying mechanism. The proper noun , which is a very rare word in the training data (occurring only 3 times) is wrongly generated as (where occurs 47 times in the training data) by the baseline model, but it is correctly tagged with the Copying mechanism (Ex. 1). Also at some, the proper noun is tagged correctly by the Copying mechanism, but the segmentation is wrong.
21.	An End-to-End Discriminative Approach to Machine Translation	Machine translation & POS TAGGING	English-French	Phrase-based models which require a translation table and language model for decoding and feature computation. The language model was a Kneser-Ney interpolated trigram model generated using the SRILM toolkit	End- to- End Approach	The system can take advantage of learned features in all stages of decoding. At the same time, discriminative methods have provided substantial improvements over generative models on a wide range of NLP tasks. They allow one to easily encode domain knowledge in the form of features. Moreover, parameters are tuned to directly minimize error rather than to maximize joint likelihood.	In this paper, we focus on two aspects of the problem of discriminative translation: the inherent difficulty of learning from reference translations, and the challenge of engineering effective features for this task.

22.	TRANSFER LEARNING FOR SEQUENCE TAGGING WITH HIERARCHICAL RECURRENT NETWORKS	POS TAGGING, text chunking, and Name Entity Recognition(NER).	English and Spanish	Three Neural architectures for the settings of cross-domain, cross-application, and cross-lingual transfer based on CNN/RNN and CRF model sharing scheme with each other.	Transfer based learning	This transfer learning approach achieves significant improvement on various datasets under low-resource conditions, as well as new state-of-the-art results on some of the benchmarks.	This transfer learning approach achieves new state-of-the-art results on all the considered benchmark datasets except PTB POS tagging, which indicates that transfer learning can still improve the performance even on datasets with relatively abundant labels.
23.	Using Neural Transfer Learning for Morpho-syntactic Tagging of South-Slavic Languages Tweets	Morpho-sytactic Tagging & POS TAGGING	Slovene, Croatian & Serbian	Neural Network (Bidirectional Gated Recurrent Units) model trained in an end-to-end manner combining both character and word level representations.	Transfer Based Learning	This TL method significantly improves results on all languages. Results further shows that the improvements made by cross-domain TL for Slovene and Croatian (+2.8% , +2.7%) are more important than improvements made by cross-domain TL for Serbian (+0.9%).	As shown in previous studies on TL (Mou et al., 2016), the parameters perfectly trained on a source data-set may be too specific to it, hence, this model may underfit on the target data-set.
24.	Exploiting Linguistic Resources for Neural Machine Translation Using Multi-task Learning	POS TAGGING & Name Entity Recogniti on(NER)	German to English	Attentional encoder-decoder model. The encoder uses word embedding of size 256 and a bidirectional LSTM with 256 hidden layers for each direction. For the attention, we use a multi-layer perceptron with 512 hidden units	Multi Tasking approach	In this work, we show that multi-task learning is a successful and a easy approach to introduce an additional knowledge into an end-to-end neural attentional model. By jointly training several NLP tasks in one system, this model is able to improve the performance of the individual task. It also suggests that multi-task learning is a promising	The error rate of POS TAGGER was high and the MT data is significantly larger than the POS data. Also a common problem of many neural MT systems is that they do not translate parts of the source sentence, or that parts of the source sentence are translated twice. The baseline system suffers from this.

				and tanh activation function. The decoder uses conditional GRU units with 512 hidden units		approach to exploit any linguistic annotated data, which is especially important if we have a lower source condition.	
25.	Improving part-of-speech tagging via multi-task learning and character-level word representations	POS TAGGING	English-Russian	As a baseline, we utilized a BiLSTM tagger, which is able to achieve state-of-the-art results on the sequence labelling tasks. We developed a new method for character level word representation using feed forward neural network	Multi tasking approach	The multi task learning with auxiliary losses is a popular method to improve generalization as achieved by this model. The word2vec and similar frameworks also give an opportunity to pre-train word vectors on the unlabelled data. Such pre-trained vectors usually improve the performance of most NLP models.	The model achieved almost 40% ERR on Syntagrus dataset by applying this process. Also the final model is worse than the best PTB models. On the other hand, it does not use word embeddings. That means that this model is much smaller. To the best knowledge, the achieved result is the best for model. To conclude, some few tricks are applied to achieve results that are on par with state-of-the-art results on considered datasets.
26.	Named Entity Recognition with Bidirectional LSTM-CNNs	Name Entity Recognition (NER)	English	A hybrid architecture of bidirectional LSTM and CNN that automatically detect words and character level features.	Word embedding	The model has surpassed the previous highest reported F1 scores for both CoNLL-2003 and OntoNotes. It doesn't use any external knowledge other than word embeddings. Also FFNNs are inadequate for larger domain like OntoNotes so LSTM model are essential for NER.	While, BLSTMM models completed training without difficulties on the CoNLL-2003 dataset BLSTMCNN model fails to converge around 5-10% of the time and similarly 1.5% of trials failed on OntoNote. Using clipping gradients and AdaDelta eliminated such failures but made training moe expensive with no gain in model performances.

4. CONCLUSION

The above survey brings the light on works done on low resource languages. This survey examined different deep learning models on different low resource languages and also brings out several metric parameter which measured the model's accuracy. Each model comes out to be unique in their ways of annotation, learning techniques, tools and corpora used.

Since in normal NLP implementation work most of the success is achieved for rich source languages like English that have text corpora of hundreds of millions of words. But we have only about 20 languages from approximately 7,000 languages in the world. There are several languages that does not have any written script which leads in faiing diverse communication. The majority of human languages are in dire need of tools and resources to overcome this resource barrier and in this NLP can deliver more widespread benefits. Therefore, building NLP applications for such languages can benefit such languages in various ways like:

1. Preservation: There are several tribal languages that may contain important information and an obvious task for NLP is to process and document these languages that do not have a writing system before they are gone forever.
2. Educational applications: Decoding ancient manuscripts and researching on historical civilization or knowing the tribal life NLP comes out to be a boon. Sometimes even languages that were extinct come back to thrive. For example the revival of Hebrew or Gaelic, and with the help of NLP techniques.
3. Knowledge expansion: Much of the world's knowledge is not in text, and these basically are low resource languages, corpora contain what people said, but not what they meant. Further research and development in NLP might give insights to connect the word and the meaning of low resource languages— not by pure statistics, but by comparing more diverse languages.