

# Part-of-Speech Tagging using Contextual Word Embeddings

Juhani Dickinson  
jdickin9@uwo.ca

Paul Moore  
email address or ORCID

**Abstract—ABSTRACT**

**Index Terms—pos tagging, part of speech, nlp, word embedding, contextual word embedding**

## I. INTRODUCTION

### A. Parts of Speech

In a sentence, each word has a syntactic role to play. In the sentence “The brown dog runs quickly towards food.”, “dog” denotes a thing, “brown” modifies or describes “dog”, “runs” denotes an action, and so on. These syntactic roles are commonly referred to as “parts of speech”, and are mapped to words using part-of-speech tags. However, this is not a one-to-one mapping. Many words have multiple meanings, and in many cases, these different meanings play different syntactic roles, mapping them to different parts of speech. For example, the word “orange”, spelled and pronounced identically, can be either a noun or an adjective, as seen in sentences (1) and (2).

“I often eat an orange after working out.” (Noun) (1)

“The orange car has arrived.” (Adjective) (2)

Figuring out what part of speech a word has is not an unsolvable task, however. Where a word is in the sentence and what words surround it play a major role in disambiguating between syntactic roles. In (1), the word after “orange” is “after”, which is an adverb. “Orange” is either an adjective or a noun, and since the following word is an adverb, it cannot be an adjective. In (2), the words surrounding “orange” are “the” and “car”, which are a determiner and a noun respectively. In this position, “orange” describes “car”, making it an adjective.

### B. Part-of-Speech (PoS) Tagging

The task of PoS tagging (often simply called tagging) is as follows: given a sentence, label each word in the sentence with the PoS tag that describes its syntactic role, seen below for sentence (3).

(3) Can you spot the large spot on your nose ?  
AUX D VB D ADJ N PP D N .

What makes the task of tagging challenging is the contextual requirement: without it, both instances of “spot” in sentence (3) would be tagged as either VB or N. Because of this, non-naive tagger implementations leverage some level of context to perform their task. Simple implementations like  $n$ -gram taggers assign tags to words based on both the word itself and the tags of the preceding  $n$  words.

### C. Word Embeddings

A common issue with simple tagging implementations is resistance to parallelisation. Any operation below the level of sentence must be linear, as the tag of the current word depends on the tag(s) of the previous words. To get around the linear-time requirement, some taggers use word embeddings, which represent words as vectors. This embedding takes into account the surrounding words, so that the meaning and context of each word is stored in the embedding. Using embeddings, tagging can be parallelised at the level of individual words.

1) *Static Word Embeddings*: pass

2) *Contextual Word Embeddings*: pass

## II. RELATED WORK

pass

## III. MOTIVATION

Part-of-speech tags are useful and sometimes key pieces of information in many natural language processing (NLP) tasks such as text-to-speech (TTS), word-sense disambiguation, marking word order, and named entity recognition (ref<sub>1</sub>). In TTS applications, the pronunciation of a word can change depending on what its syntactic role is or sometimes even what tense it is. The word “resume” is pronounced differently based on whether it is meant as “continue doing” or as “a document detailing professional experience”, which can be easily disambiguated based on whether it acts as a verb or a noun. This disambiguation-by-tag is also used in word-sense disambiguation, like in sentence (3). In a search-engine setting, understanding whether “spot” refers to an activity or a visual marker allows for more accurate search results.

While high-resource languages like English no longer use PoS tagging in some popular NLP applications like AI assistants by making use of the massive collections of curated data now available, most languages do not have the luxury of large, high-quality datasets. In these low-resource languages, PoS tags are still critical for NLP tasks.

For a task like machine translation, (ref<sub>2</sub>) found that using linguistic features derived from PoS tags improved the translation performance not only between the low-resource language pair of Thai and Myanmar but also between Thai and English and Myanmar and English. j-MORE-<sub>i</sub>

j-INTRO SENTENCE-<sub>i</sub>. (ref<sub>3</sub>) <sub>i</sub>. As well as being useful, the efficacy of cross-linguistic transfer is predictable. (ref<sub>4</sub>) found that language families and morphological structures

have a major impact on the performance of cross-linguistic performance. ¶–TALK ABOUT RESULTS–¶.

Finding the best PoS tagger for English, a high-resource language, has clear benefits for related low-resource languages. One example is Scots, a language in the Anglic family currently considered “vulnerable” by the UNESCO Atlas of the World’s Languages in Danger (¶ref<sub>ℓ</sub>). Research suggests that NLP can support and help revitalise endangered languages (¶ref<sub>ℓ</sub>) through machine translation, TTS, and integration into learning materials. ¶–MORE–¶

#### IV. METHODS

pass

#### V. EXPERIMENTAL RESULTS

pass

#### VI. CONCLUSIONS

pass

#### REFERENCES

[1] Pass