Annotation Projection With Pattern Based Neural Networks For Antonym and Synonym Detection in the Yorùbá Language

by

Kamoya Ikhofua

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2021

Author .....
Department of Electrical Engineering and Computer Science
Certified by .....
Regina Barzilay, PhD
Delta Electronics Professor MIT CSAIL
Thesis Supervisor

Certified by .....
Jiaming Luo
PhD Candidate MIT CSAIL
Thesis Supervisor

Accepted by .....
Katrina LaCurts
Chair, Master of Engineering Thesis Committee

Annotation Projection With Pattern Based Neural Networks For Antonym and Synonym Detection in the Yorùbá Language

by

Kamoya Ikhofua

Submitted to the Department of Electrical Engineering and Computer Science
on , in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

In this thesis, I explore the task of automatically distinguishing between synonyms and antonyms in the Yorùbá language. Previous work by Nguyen et al. [16] relies on linguistic resources such as WordNet to provide supervision for this task in English, however at this time there is no WordNet for Yorùbá. I investigate whether we can bridge this resource gap by utilizing machine translation systems, parallel corpora and natural language parsers, following the annotation projection paradigm [10, 13, 18]. In Chapter 1, I provide an overview of the entire work and my technical contributions. Next, I state the philosophical motivations of this work and its larger goals. In Chapter 2, I provide a background on annotation projection and on previous work on the task of distinguishing between antonyms and synonyms. In Chapter 3, I present the preliminary results of this work and explain its limitations as well as how they can be overcome in future work. I conclude with a summary about this work and its contributions.

Thesis Supervisor: Regina Barzilay, PhD
Title: Delta Electronics Professor MIT CSAIL

Thesis Supervisor: Jiaming Luo
Title: PhD Candidate MIT CSAIL

Acknowledgments

I would like to thank my father for pushing me to pursue a Master's degree and encouraging me through it, despite my fits and complaints. This thesis is mostly for him and his desires to see me achieve a higher level of educational attainment than he did. I would also like to thank my mother for her encouragement and emotional support through the stressful and turbulent times.

Next, I must thank my academic supervisor, Dr. Julie Greenberg, for her guidance and advice during my undergraduate and graduate years at MIT. It was with her that I first discussed my ideas for building an online Yorùbá dictionary and making some aspect of that project serve as my Master's thesis project. She was very supportive and encouraged me to seek supervision for the project from a member of the MIT faculty. I would also like to thank Dr. Regina Barzilay for recognizing the importance of my ideas and for agreeing to supervise me, Dr. Katrina LaCurts for her understanding and encouragement at a critical moment, and the soon-to-be Dr. Jiaming Luo, who was most gracious and kind in sharing his time, knowledge, and energy to help me to rein in my different ideas into this particular thesis project and supporting me through it.

Furthermore, I must thank my classmates Ian and Mahi for their assistance with certain parts of this project, including an early iteration for which we co-wrote a paper. It was a pleasure to work with them in Dr. Andreas's excellent Neurosymbolic Models for NLP seminar. And without question, I must thank Jing for persuading me to continue with the degree when I was intensely stressed and was considering quitting. He told me that people would be less interested in working with me if I did not have a Master's. Whether or not that is true, the comment served its intended effect and I continued on, grudgingly. I also give thanks to the other unnamed people who helped me in one way or the other to get to this point. I am very grateful.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

My central argument in this introductory chapter is that language is one of the fundamental pillars of any society and that a decline in language is an indicator of the decline of the society. I refer to the decline of the Yorùbá language [17, 4] and the observation that it lacks a comprehensive, digital monolingual dictionary, to the best of my knowledge. I reason that a digital, online dictionary is of utmost importance if the language is to avoid extinction and regain relevance in the 21st century and beyond. Not only do dictionaries provide reliable definitions, but they also show the correct usage of words in example sentences, provide word and phrase etymologies, and list the synonyms, antonyms, etc. for any given word. This thesis focuses on the specific task of synonym and antonym recognition since it is an area in which automation might expedite the development of an online dictioary for Yorùbá. The ideal outcome would be that as a lexicographer works on a particular Yorùbá word, an assistant engine should automatically provide a list of candidate synonyms and antonyms to ensure that the human catalogs as many synonyms and antonyms that really exist in the language for the given word. The pursuit of this ideal led me to survey the literature for previous work on synonyms and antonyms [16, 15] and use them as inspiration for this work.

However, previous work on distinguishing between synonyms and antonyms in

English relies on linguistic resources which do not currently exist for Yorùbá, such as WordNet. As a result, I propose to bridge this resource gap by using machine translation systems, parallel corpora, and dependency parsers in the annotation projection paradigm to train two models that accomplish this task. The first model is a purely pattern-based neural network which achieves 0.642 precision, 0.860 recall, 0.735 F1, and 98/129 accuracy. The second model is a combined pattern-and-distribution-based model which achieves 1.000 precision, 0.981 recall, 0.990 F1, and 128/129 accuracy. These results are from preliminary experiments which require further refinements to the data-processing pipeline, but they provide good indications that this research design is effective at this task.

In the remainder of this chapter, I provide some personal experiences, observations, and opinions which led me to pursue this work, as well as support for the idea that language plays an important role in societal development or decline. I also explain how Yorùbá differs from English, particularly in its use of diacritcs to distinguish between otherwise heteronymous words. In chapter 2, I explain the research design, annotation projection, data collection and preparation, and other methods used. In chapter 3, I present the results, provide analysis and discuss limitations of this work so far, in addition to suggesting ways to build on this in future work.

1.2 Language As A Signal for Societal Development or Decline

As is expected, my entire life's direction is the result of my experiences. I was born in Lagos, Nigeria and from an early age, I understood that things were wrong with the country. At home, electricity supply was unreliable. On the go, the roads were bumpy and often riddled with potholes. Mild-to-heavy rains were liable to result in flooding, sometimes mud-colored flooding. Adults frequently lamented about a distant past when the currency was more valuable and things were better. As I grew and started developing an understanding of my environment, I decided to see the problems in my

society as a challenge to me to learn how to reduce the unpleasantness of existence in the country. Around the age of nine or ten, I came to understand that engineers and scientists were the people who specialised in building things and solving the aforementioned problems, I decided for myself that I should direct myself towards the sciences and engineering.

But along the way, I began to question the origins of the very society I found myself in. I read about its history and those of other societies with greater achievements. It eventually became clear to me that I was born into an artificial nation in which many disparate societies had been bound together for the administrative conveniences of an external colonial power and that momentum had kept the contraption trudging along, at the cost of the self actualisation of the trapped societies. Frederick Forsyth wrote that "through all the years of the pre-colonial period Nigeria never was united, and during the sixty years of colonialism and the sixty-three months of the First Republic only a thin veneer hid the basic disunity... not only was Nigeria neither happy nor harmonious, but it had for the five previous years stumbled from crisis to crisis and had three times already come to the verge of disintegration. In each case, although the immediate spark had been political, the fundamental cause had been the tribal hostility embedded in this enormous and artificial nation. For Nigeria had never been more than an amalgam of peoples welded together in the interests and for the benefit of a European power." [6]

This information helped me to understand why the country is riddled with problems, disorganized, and dangerous to live in. With over 250 distinct ethnic groups (each indigenous to their respective region of the country) and languages, could one really expect anything different? After all, statistical analyses by Tatu Vanhanen shows that "the degree of ethnic conflict is strongly related to the degree of ethnic divisions" and since, as Vanhanen notes, humans are evolved to be predisposed to ethnic nepotism, i.e. the tendency of members of an ethnic group to favor their group members over nonmembers because they are more closely related to their group members than to outsiders [21]. Furthermore, according to Philip Atkinson, cultures clash rather than mix, since "the difference in the understanding of cultures makes different

cultures enemies, who will struggle to impose their wills upon each other." [3]

My understanding of the situation led me to believe that it is futile to continue to think in terms of Nigeria. It is not an organic country that formed on the basis of shared ancestry, language, and culture, or even from wars between the natives that fused them together. It dawned on me that not only did the country not always exist, but that some of my ancestors could never have imagined that it would exist and that the fate and identity of their descendants would be what it is today. In the introduction to Samuel Johnson's tome about the history of the Yorùbá people (my mother's ethnic group), the very first sentence describes a "Yorùbá country. "The Yorùbá country lies to the immediate West of the River Niger (below the confluence) and South of the Quorra (i.e, the Western branch of the same River above the confluence), having Dahomey on the West, and the Bight of Benin to the South." [11] In fact, for the remainder of the book the author only ever refers to the Yorùbá country as a nation separate from others in the region.

It then seemed to me that, whatever contributions I wanted to make towards making my country better, less hazardous, less stressful, that I needed to update the notion of "my country" from Nigeria to the original indigenous societies that originally existed. In thinking about what I could do in my current capacity to make contributions to my ancestral societies, I decided to address the issue of the decline of the languages, beginning with Yorùbá, the language of my mother's ethnic group, and then with Edo (Bini), the root language of my father's ethnic group. In particular, I lamented about the fact that I cannot speak either of these languages with the same command and flair with which I speak the English language. Why should this be if not for the hegemonic status of English resulting from its use as a common language between people of different native tongues? As a consequence, natives' competence in their own languages is eroding with each successive generation, especially as the so called government has no incentive to invest in the development of language tools for over 200 languages. [4, 17]

My belief is that language plays an important role in societal development and that if the decline of my ancestral societies is to be reversed, the languages must

first be revived so that natives' can have access to meaning and expression in these languages and thus regain their sense of identity. In the first place, a society may be described as being shaped by different traditions of which language is an essential part. And as Atkinson writes, language is the shared understanding of the citizens of a society. "The society will embark upon an ever-improving cycle of gaining wisdom and strength by refining and extending its traditions. This improvement will be revealed in its improving customs, manners, laws and institutions but especially in its use of language." [3] He also notes that social decay itself is caused by a widespread decay in the clarity of thought which obtains a widespread decay in the clarity of expression." [3]

Samuel Johnson wrote in 1897 that "educated natives of Yorùbá are well acquainted with the history of England and with that of Rome and Greece, but of the history of their own country they know nothing whatever! This reproach it is one of the author's objects to remove." [11] I would say that it is increasingly the case today that educated natives of Yorùbá are better at expressing themselves in English than in their own language and that for all intents and purposes, English is actually their first and primary language. I think this is a sign of the decline of the Yorùbá society and this reproach it is my mission to remove by developing a robust online dictionary, amongst other resources, for the Yorùbá (and subsequently the Edo/Bini) language.

1.3 Yorùbá

Yorùbá is the third most spoken African language, after Swahili and Amharic, with over 40 million speakers [1]. Its written form consists of 25 alphabets, with 17 consonants, 7 oral vowels, 5 nasal vowels, and 4 syllabic nasals. The alphabet differs from that of English in its use of diacritics to differentiate kinds of tones, which in turn denote different meanings. The low tone is represented with the grave symbol (\), the mid tone with no symbol, and the high tone with the acute symbol (/). The fourth diacritic is a dot below "e" and "o" and represents the open phonetic variants of those letters, while the dot below the s represents the long variant of "s" (i.e 'sh')

Oral vowels (7)	**Nasal vowels (5)**	**Syllabic nasals (4)**
a, e, ẹ, i, o, ọ, u	an, ẹn, in, ọn, un	ḿ, m̀, ń, ǹ

Consonants (17)

b, d, f, g, gb, j[dz], k, l, m, n, p[kp], r, s, ṣ, t, w, y

Tones (3)

Low: à, Mid: a, High: á

Figure 1-1: An overview of the different categories of characters within written Yorùbá, including the diacritical tone markers.

[9]. These characters are displayed in more detail in Figure B-1. Overall, Yorùbá is typologically different from English at both the character and syntactic levels.

Although Yorùbá has more language resources (through various mass media forms including books) than most of its neighboring languages, it is still classified as a LRL because there exists no readily available corpora for computational analysis [9] and no comprehensive online dictionary based on my investigation so far.

However, Ishola and Zeman created a Yorùbá treebank (YTB) by applying the Universal Dependencies (UD) framework to the Yorùbá Bible which provides an opportunity for dependency analysis of Yorùbá. UD provides an accessible, open source universal inventory of categories and tagsets with the capability for extensions and universal guidelines for consistent annotations across languages. The UD framework coherently synthesizes Stanford Dependencies, Google part-of-speech tags, and the Interset, which contains interlingua for morphosyntactic features. Ishola and Zeman cite various Yorùbá-specific linguistic considerations, for example: Yorùbá has a strict subject-object-verb order; some hyphenated words cannot be correctly annotated individually; there exists prepositions, but not postpositions [9]. The YTB outlines a UD framework for Yorùbá and also provides a small dataset of around 1000 Yorùbá sentences. It is nevertheless the current best resource available for parsing digital Yorùbá text and so it serves as the foundation for this work on using the annotation projection paradigm to automatically distinguish between synonyms and antonyms in the language.

Chapter 2

Research Design

2.1 Overview

As Yorùbá does not have a WordNet or other comprehensive lexical resource for easy access to list of synonyms or antonyms, I created a novel dataset of Yorùbá antonyms and synonyms in the annotation projection paradigm using the Bible as a parallel corpus. I use WordNet to identify verses in the English Bible which contain synonym pairs and antonym pairs respectively, extract the pairs, then feed them into the Google translate API to obtain their translations into Yorùbá. Finally I check if the translated pairs exist in the parallel verses in the Yorùbá Bible. In this way, I obtain a list of triples, each of which comprises a synonym or antonym pair along with a Yorùbá sentence which contains the pair of words. To process the resulting dataset, I train a CoNLL-U dependency parser using the Universal Dependencies Yorùbá treebank by Ishola and Zeman [9] and derive patterns from the simple paths between synonymous or antonymous words in a syntactic parse tree. I encode the resulting patterns into vector representations using the Facebook AI's FastText word embeddings for Yorùbá. With these, I trained two models, one which uses the pattern vectors patterns and a second model which concatenates the vectors of the pair of words with the pattern vectors and show that this method may applied to other resource-poor languages for this task. I provide more details about the research design in the sections below.

2.2 Annotation Projection

Annotation projection is a technique which makes use of parallel corpora to generate linguistically annotated corpora for resource-poor languages with the help of a resource-rich language which has tools to automatically generate the annotations. Parallel sentences from the target resource-poor language are aligned with sentences from the source language and the annotations in the source are projected into the target sentence. Guasekara et al. [10] used this technique to create the first-ever semantic role labeller for Sinhala, a language spoke mainly in Sri Lanka. Smith and Eisner [18] use annotation projection for cross-lingual parser projection to learn a dependency parser for a target language, whereby they learn a dependency parser for a target language by using parallel corpora, an English parser, and automatic word alignments. McDonald et al. [13] use a constraint driven learning algorithm to project delexicalized dependency parsers from multiple source languages with labeled training data to target languages without labeled training data. Lohk et al. [12] use annotation projection as part of an automatic method for composing synsets with multiple synonyms by using Google Translate and Semantic Mirrors' method.

2.3 Data

To develop a larger dataset of Yorùbá sentences, and since a variety of well-known synonyms and antonyms exist in the Bible, I use the Yorùbá translation of the King James Bible as our source of Yorùbá sentences. I collect verses from the English Bible found to contain synonym or antonym pairs, translate the English synonyms and antonyms to Yorùbá, and retain the Yorùbá verses which contained the translated word pairs.

Specifically, I remove co-occuring synonyms and antonyms with the same lemma using the WordNetLemmatizer in NLTK. I then use the Google Translate API to identify the most probable translation of the English synonyms and antonyms to Yorùbá, and verify that each of those translated Yorùbá words is proper and exists in

the corresponding Yorùbá verse. If a match exists, I have identified the co-occurring Yorùbá synonyms and antonyms in their sentences. If a match does not exist, I manually inspect the verse to identify the correct co-occurring Yorùbá synonyms and antonyms. In addition, WordNet imperfectly identifies synonym and antonym pairs in English; for example, WordNet improperly identifies "have" and "give" as synonyms and extracts the phrase "have given" as a synonym pair. For these limited numbers of verses with improper pairs, I manually remove them from the dataset. Although tedious, manual inspection to verify automated results is critical for accuracy. I also plan to release this Yorùbá synonym-antonym dataset since none currently exist. I then use the sentences with synonyms and antonyms in the Yorùbá Bible to create syntactic parse trees for our vector embeddings and then use these vector embeddings to train the pattern-based model. Examples of how the dataset is automated and verified for both synonyms and antonyms are included in B and B-3, respectively. Notably, I will use the Yorùbá Dependency Treebank and Yorùbá word embeddings in FastText [2].

2.4 Methods

Following the approach outlined in [16] and 2.3, I created the Yorùbá dataset outlined in Table A.1. The synonym and antonym patterns are combined into one larger dataset and then randomly shuffled. Out of the entire sample set, 70 percent is randomly sampled into a training set, 15 percent is sampled as a validation set, and 15 percent is used as a test set.

Table 2.1: Yorùbá Synonym/Antonym Dataset

Class	# of samples
Synonyms	1073
Antonyms	423
Total	1496

The table above catalogs the novel dataset of Yorùbá antonyms and synonyms I created using the annotation projection paradigm with the Bible as a parallel corpus.

Synonyms				
Ex.	Mode	Word 1	Word 2	Verse
1	English	creature	beast	Gen$1:24 And God said, Let the earth bring forth the living creature after his kind, cattle, and creeping thing, and beast of the earth after his kind: and it was so.
	Yorùbá (auto)	ẹran	ẹran	Gen$1:24 “Ọlọ́run sì wí pé, “Kí ilẹ̀ kí ó mú ohun alààyè jáde ní onírúurú wọn: ẹran ọ̀sìn, àwọn ohun afàyàfà àti àwọn ẹran inú igbó, ọ̀kọ̀ọ̀kan ní irú tirẹ̀.” Ó sì rí bẹ́ẹ̀.“
	Yorùbá (verified)	ẹ̀dá	ẹranko	Gen$1:24 “Ọlọ́run sì wí pé, “Kí ilẹ̀ kí ó mú ohun alààyè jáde ní onírúurú wọn: ẹ̀dá ọ̀sìn, àwọn ohun afàyàfà àti àwọn ẹranko inú igbó, ọ̀kọ̀ọ̀kan ní irú tirẹ̀.” Ó sì rí bẹ́ẹ̀.“
2	English	land	nation	Exod$9:24 So there was hail, and fire mingled with the hail, very grievous, such as there was none like it in all the land of Egypt since it became a nation.
	Yorùbá (auto)	ilẹ	None	Exod$9:24 Yìnyín rọ̀, mọ̀nàmọ́ná sì bẹ̀rẹ̀ sí bù sí orí ilẹ̀ èyí ni ó tí ì buru jù ti ó ṣẹlẹ̀ láti ìgbà ti Éjíbítì ti di orílẹ̀-èdè.“
	Yorùbá (verified)	ilẹ	orílẹ̀-èdè	Exod$9:24 Yìnyín rọ̀, mọ̀nàmọ́ná sì bẹ̀rẹ̀ sí bù sí orí ilẹ̀ èyí ni ó tí ì buru jù ti ó ṣẹlẹ̀ láti ìgbà ti Éjíbítì ti di orílẹ̀-èdè.“
3	English	let	have	Gen$1:26 And God said, Let us make man in our image, after our likeness: and let them have dominion over the fish of the sea, and over the fowl of the air, and over the cattle, and over all the earth, and over every creeping thing that creepeth upon the earth.
	Yorùbá (auto)	jẹ ki	ni	Gen$1:26 "Léyìn náà ni Ọlọ́run wí pé, “Ẹ jẹ́ kí a dá ènìyàn ní àwòrán ara wa, gégẹ́ bí àwa ti rí, kí wọn kí ó jọba lórí ẹja òkun, ẹyẹ ojú ọ̀run, ohun ọ̀sìn, gbogbo ilẹ̀ àti lórí ohun gbogbo tí ń rìn lórí ilẹ̀.”
	Yorùbá (verified)	jẹ ki	ni	Gen$1:26 "Léyìn náà ni Ọlọ́run wí pé, “Ẹ jẹ́ kí a dá ènìyàn ní àwòrán ara wa, gégẹ́ bí àwa ti rí, kí wọn kí ó jọba lórí ẹja òkun, ẹyẹ ojú ọ̀run, ohun ọ̀sìn, gbogbo ilẹ̀ àti lórí ohun gbogbo tí ń rìn lórí ilẹ̀.”

Figure 2-1: Synonym pair examples of how the dataset is constructed from a KJV version of the English bible. In each example, I use NLTK to identify synonyms and antonyms and corresponding verse in English KJV Bible, then use the unique identifier of the verse to find the corresponding verse in Yorùbá KJV Bible. Next, I use Google API to translate the English words to Yorùbá, verify if in Yorùbá verse. Example 1 demonstrates a verse in which the automation is incorrect, and I manually correct it. Example 2 demonstrates a verse in which the automation provide "None" (i.e. no translation existed from Google Translate API), and I manually fill it in. Example 3 demonstrates a verse in which the automation is correct, so the verified version is a repeat.

Antonyms				
Ex.	Mode	Word 1	Word 2	Verse
1	English	day	night	Gen$1:16 And God made two great lights; the greater light to rule the day, and the lesser light to rule the night: he made the stars also.
	Yorùbá (auto)	ọjọ	alẹ	Gen$1:16 Ọlọ́run dá ìmọ́lẹ̀ ńlá-ńlá méjì, ìmọ́lẹ̀ tí ó tóbi láti ṣe àkóso ọ̀sán àti ìmọ́lẹ̀ tí ó kéré láti ṣe àkóso òru. Ó sì dá àwọn ìràwọ̀ pẹ̀lú.
	Yorùbá (verified)	ọ̀sán	òru	Gen$1:16 Ọlọ́run dá ìmọ́lẹ̀ ńlá-ńlá méjì, ìmọ́lẹ̀ tí ó tóbi láti ṣe àkóso ọ̀sán àti ìmọ́lẹ̀ tí ó kéré láti ṣe àkóso òru. Ó sì dá àwọn ìràwọ̀ pẹ̀lú.
2	English	bad	good	Gen$24:50 Then Laban and Bethuel answered and said, The thing proceedeth from the Lord : we cannot speak unto thee bad or good.
	Yorùbá (auto)	None	búburú	Gen$24:50 “Lábánì àti Bétúélì sì dáhùn pé, “Lọ́dọ̀ OLÚWA ni èyí ti wá, nítorí náà àwa kò le sọ rere tàbí búburú fún ọ.”
	Yorùbá (verified)	rere	búburú	Gen$24:50 “Lábánì àti Bétúélì sì dáhùn pé, “Lọ́dọ̀ OLÚWA ni èyí ti wá, nítorí náà àwa kò le sọ rere tàbí búburú fún ọ.”
3	English	father	mother	Gen$20:12 And yet indeed she is my sister; she is the daughter of my father, but not the daughter of my mother.
	Yorùbá (auto)	baba	ìyá	Gen$20:12 "Yàtọ̀ fún ìyẹn, òtítọ́ ni pé arábìnrin mi ni. Ọmọ baba kan ni wá, bí ó tilẹ̀ jẹ́ pé, a kì í ṣe ọmọ ìyá kan."
	Yorùbá (verified)	baba	ìyá	Gen$20:12 "Yàtọ̀ fún ìyẹn, òtítọ́ ni pé arábìnrin mi ni. Ọmọ baba kan ni wá, bí ó tilẹ̀ jẹ́ pé, a kì í ṣe ọmọ ìyá kan."

Figure 2-2: Antonym pair examples of how the dataset is constructed from a KJV version of the English bible. Example 1 demonstrates a verse in which the automation is incorrect, and I manually correct it. Example 2 demonstrates a verse in which the automation provide "None" (i.e. no translation existed from Google Translate API), and I manually fill it in. Example 3 demonstrates a verse in which the automation is correct, so the verified version is a repeat.

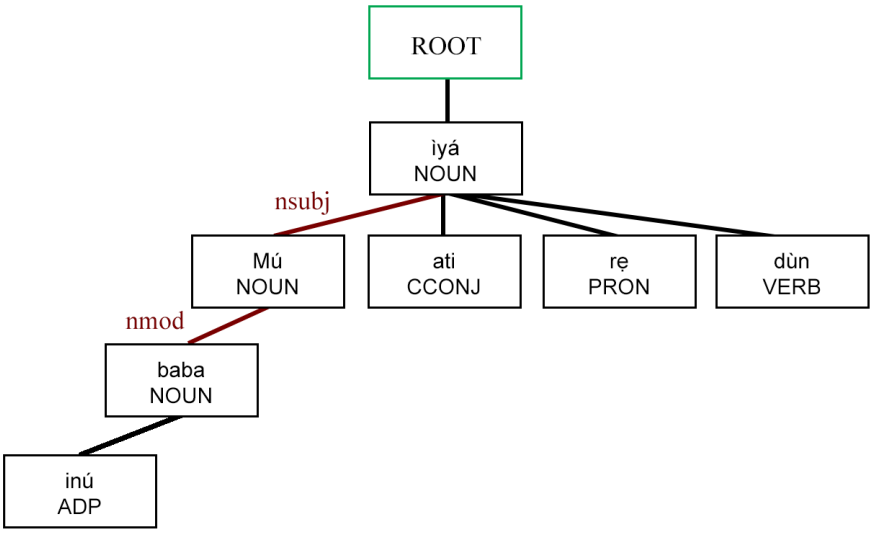

Figure 2-3: An example of a Yorùbá dependency parse from UDPipe, with the path from baba (English: father) to ìyá (English: mother) highlighted. The pattern-based neural network classifies word relations via these paths. The phrase is taken from Proverbs in the Yorùbá Bible.

2.4.1 UDPipe

I trained our dependency parser using the Universal Dependencies treebank, as mentioned previously [9]. This includes a set of hand-annotated sentences from the Yorùbá Bible and Wikipedia. The parser outputs information in CoNLL-U format, from which I collect patterns [19, 20].

2.4.2 Patterns

Nguyen et al. derive patterns from the simple paths between synonymous or antonymous words in a syntactic parse tree [16]. The patterns consist of the nodes along the simple path connecting the concerned words, with each node represented by a vector of four features: lemma, part of speech (POS), dependency label, and distance label. I adopt the same approach, using our UDPipe parser and corpus of Yorùbá text. Figure B-4 shows an example of a dependency parse for a Yorùbá phrase, with POS tags and dependency labels.

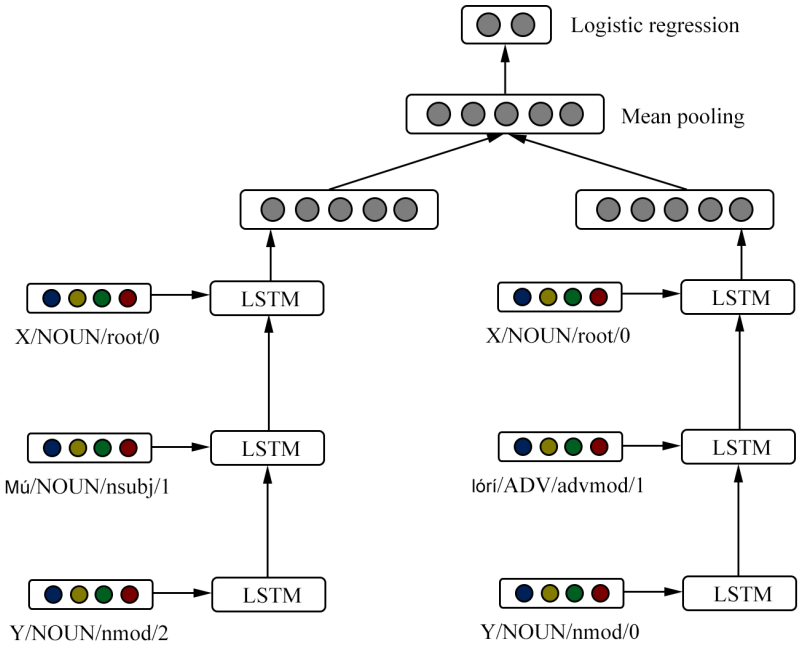

Figure 2-4: An overview of the LSTM-based model, one of the two models that I train in this work. Vectors, comprised of the concatenated lemma/POS/dependency tag/distance label, are used as inputs to LSTMs, which then feed into a logistic regression layer. The examples here are from the same word pair in Figure B-4.

2.4.3 Models

Nguyen et al. employ two different approaches to distinguish between antonyms and synonyms. The first approach uses LSTM units to encode the patterns as vector representations, which are then fed to a logistic regression layer for classification as synonymous or antonymous. This model is shown in Figure B-5. The second approach concatenates the vectors derived from the co-occurrence distribution of the words with the vectors derived from the syntactic path patterns before feeding them to the classifier. Both of these approaches outperformed the previous baseline. I use the same two approaches on the Yorùbá dataset.

Both models were randomly initialized and trained for 50 epochs. Weights were optimized via Adadelta with a learning rate of 0.0001 and a dropout rate of 0.5. All of these parameters are unchanged as much as possible from the original Nguyen et al. paper.

Chapter 3

Results

Table 3.1: Test Set Performance

Model Type	Precision	Recall	F1 Score	Accuracy
Pattern-based Model	0.642	0.860	0.735	98/129
Combined Model	**1.000**	**0.981**	**0.990**	**128/129**

The results of preliminary experiments showing that the combined model outperforms the pattern-based model on all metrics are shown in Table A.2. Examples of correctly and incorrectly classified word pairs are shown in Figure B-6.

3.1 Analysis

In general, the models were successfully able to distinguish synonyms from antonyms in Yorùbá text, although some bugs in the data processing pipeline resulting in a truncated dataset still need to be fixed.

Notwithstanding, these results show that the combined model outperforms the pattern-based model in precision, recall, F1, and accuracy scores. Notably, my combined model shows higher overall scores than those reported in the Nguyen et al. [16] paper and my pattern-based model has slightly worse precision but comparable recall and F1 scores. However, further work needs to be done to further validate these results. The fact that I use Facebook's FastText embeddings in Yorùbá whereas Nguyen et al. use dLCE embeddings as well as Google's GloVe English word embeddings in

Word 1	Word 2	Verse	Ground truth	Predicted label
ibi	bátà	Acts$7:33 "Olúwa sì wí fún un pé, 'Tú bátà rẹ kúrò ni ẹsẹ rẹ, nitorí ibi tí ìwọ gbé dúró sí yìí ilẹ mímọ́ ni."	0	0
èbè	àdúrà	2Chr$6:19 "Síbẹ ṣe àfiyèsí àdúrà ìránṣẹ́ rẹ àti èbè rẹ fún àánú, OLÚWA Ọlọ́run mi, gbọ́ ẹkún àti àdúrà tí ìránṣẹ́ rẹ ń gbà níwájú rẹ."	0	0
mú	ilẹ	Ezek$20:10 "Nítorí náà, mo mú wọn jáde ní ilẹ Éjíbítì mo sì mú wọn wá sínú ihà."	0	0
òkùnkùn	ojiji	Job$34:22 "Kò sí ibi òkùnkùn, tàbí ojiji ikú, níbi tí àwọn oníṣẹ́ ẹ̀ṣẹ̀ yóò gbé sá pamọ́ sí."	0	1

Word 1	Word 2	Verse	Ground truth	Predicted label
ìyá	bàbá	Ps$27:10 "Bí ìyá àti bàbá bá kọ̀ mí sílẹ̀, OLÚWA yóò tẹ́wọ́ gbà mí."	1	1
gbóná	tutù	Rev$3:16 "Ǹjẹ nítorí tí ìwọ lọ wǫ́ǫ́rọ́, tí o kò si gbóná, bẹ́ẹ̀ ni tí o kò tutù, èmi yóò pọ̀ ọ́ jáde kúrò ni ẹnu mi."	1	1
obìnrin	ènìyàn	1Sam$25:3 "Orúkọ ọkùnrin náà sì ń jẹ́ Nábálì, orúkọ aya rẹ ń jẹ́ Ábígáílì; òun sì jẹ́ olóye obìnrin, àti arẹwà ènìyàn; ṣùgbọ́n òǹrorò àti oníwà búburú ni ọkùnrin; ẹni ìdílé Kálẹ́bù ni òun jẹ́."	1	1
alágbára	aláìlera	Rom$15:1 "Àwa tí a jẹ́ alágbára nínú ìgbàgbọ́ yẹ kí ó máa ru ẹrù àìlera àwọn aláìlera, kí a má si ṣe ohun tí ó wu ara wa."	1	0

Figure 3-1: Examples of correctly classified and incorrectly classified word pairs. The topmost four samples are true synonyms, while the bottom four are true antonyms. There are no obvious common features in all of the incorrectly classified samples. For a discussion of the *obìnrin/ènìyàn* example, see Section 3.1.1.

their models could be a contributing factor in these differences and work remains to be done to determine the extent of these contributions.

Overall, the results suggest that there are underlying similarities in the task of distinguishing between antonyms and synonyms in both Yorùbá and English. One hypothesis is that word occurrence statistics follow approximately the same distribution, resulting in similar proportions of antonyms and synonyms as "difficult" examples that occur less frequently in text. Another hypothesis is that the models explored in this paper can only represent enough information to correctly classify a certain proportion of samples.

3.1.1 Limitations

One limitation of this work is that I used the English and Yorùbá Bibles as the only source of sentences. In future iterations of this work, it would be beneficial to use the synonyms and antonyms dataset that I curated to extract sentences from other corpora, specifically natural Yorùbá corpora, as it will allow for comparisons between the performance of the model on a dataset generated by translating from a higher-resource language and the performance on an originally Yorùbá dataset, especially with regard to the impact of syntax. Furthermore, obtaining other corpora would give access to other domains and lessen whatever effects the peculiarities of the syntax of religious text have on the models' learned representations. To that end, I would like to expand to more general and contemporary lingual data sources by using web-based corpora Yorùbá Global Voices as a source of sentences in Yorùbá.

Another limitation of this work is that due to the imperfections of our dataset, some of the ground truth labels are actually not accurate. For example, in the third row of the second table in Figure 7, the words, "obìnrin" (woman) and "ènìyàn" (person) are labeled as antonyms when in fact they are synonyms. It is worth noting that the word "okùnrin" (man) does occur later in that verse, so the mislabeling in the dataset is likely caused by the translation API.

Furthermore, two otherwise synonymous or antonymous words may not actually be synonymous or antonymous in the context of a given sentence. For example, in the sentence *'Although my dark blue coat is very light, it has a special heat-keep system, so it always keeps me warm'*, I can see that 'dark' and 'light' are not an antonym pair in the context of this particular sentence. Nguyen et al. try to exclude such cases by discarding patterns which occur below some threshold frequency and we adopt the same contingency. Thus, the 'x is very y' pattern would be discarded as an antonym-pattern if it does not occur frequently enough for a specific pair of 'x' and 'y'. However, it is still possible that some false patterns such as the one mentioned were retained. I leave it to future work to improve the quality of the dataset.

3.2 Future Work

In future iterations of this work, I plan to design a new approach that uses transformer-based language models such as BERT instead of pattern-based neural networks. I suspect that a multilingual BERT (mBERT) model may outperform the pattern-based network. The task of distinguishing synonyms and antonyms in sentences requires both syntactic and semantic information, and this is one of the reasons why standalone distributional co-occurrence models such as Word2Vec or GloVe have poor performance on this task [14]. However, previous work has shown that BERT contextual word embeddings can embed the syntactic tree of a sentence [7]. Recent research has also demonstrated that WordNet semantic subtrees can be reconstructed by BERT embeddings [5] and that fine-tuned BERT models can distinguish different semantic senses from the same word [8]. Furthermore, other work has shown that the BERT model, which is trained with general self-supervised language masking, is highly transferable to various downstream tasks by fine-tuning. BERT, with its ability to embed the context and distinguish between word senses, is likely also more resistant to false patterns, as discussed in Section 3.1.1. As such, it may achieve better performance on this task than the pattern-based approach.

Thus, a next step should be to fine-tune BERT to distinguish between synonyms and antonyms, first in English to verify its performance and then in Yorùbá through transfer-learning techniques. One could then compare the pattern-based approach to the fine-tuned mBERT approach and report the results obtained both in English and Yorùbá, but I leave this for future work.

Chapter 4

Conclusion

This work was motivated by the thinking that language is a fundamental pillar of any society and that its decline is a signal for the decline of the society. Myself and others before me have noted the drop in the competence of Yorùbá natives in their indigenous language [17, 4] and stressed the need to reverse this trend with the creation of online dictionaries and other digital language tools. As part of the investigation into ways to assist Yorùbá lexicographers and linguists with automated systems for dictionary making, I identified the ability to automatically classify synonyms and antonyms in Yorùbá text as part of a larger pipeline for the construction of dictionaries and other lexical resources for the language.

To that end, I successfully trained two pattern-based neural network models in preliminary experiments to distinguish between synonyms and antonyms in the Yorùbá language. The pattern-based model achieves 0.642 precision, 0.860 recall, 0.735 F1, and 98/129 accuracy, while the combined model achieves 1.000 precision, 0.981 recall, 0.990 F1, and 128/129 accuracy.

To achieve this result, I created a novel dataset of Yorùbá antonyms and synonyms using the Bible as a parallel corpus and the annotation projection paradigm. This approach to dataset creation can be applied to any other language since it automatically generates enough samples to train a model that had comparable metrics with models trained in the source language. With this same combination of corpus and pattern-based model, future work could focus on replicating this experiment across

many languages. In fact, an interesting experiment could be to further automate the process of dataset creation and test the pattern-based model across a multitude of automatically generated datasets.

While the pattern-based approach has advantages over a simple co-occurrence model, it also some limitations of its own. First, the compositional nature of language means that there are a near-infinite number of ways to include a given antonym or synonym pair in a sentence. This means that there is a high probability that, given a pair of words with an unknown relationship in a sentence, the model will have never seen the pattern between the two words before. As a result, it is particularly important to have a large corpus available when training this model. This model may also be prone to over fitting when there are too few patterns in the corpus to train a robust model. This can result in very high or very low validation scores, depending on the train/test splits. Overall, these features make pattern-based models challenging to work with. I suspect that the BERT model fine-tuned on a synonym/antonym task described in Section 3.2 might be a more robust and more practical model.

Appendix A

Tables

Table A.1: Yorùbá Synonym/Antonym Dataset

Class	# of samples
Synonyms	1073
Antonyms	423
Total	1496

The table above catalogs the novel dataset of Yorùbá antonyms and synonyms I created using the annotation projection paradigm with the Bible as a parallel corpus.

Table A.2: Test Set Performance

Model Type	Precision	Recall	F1 Score	Accuracy
LSTM	0.642	0.860	0.735	98/129
Combined	**1.000**	**0.981**	**0.990**	**128/129**

The results in the table above that the combined model outperforms the pattern-based model on all metrics.

Appendix B

Figures

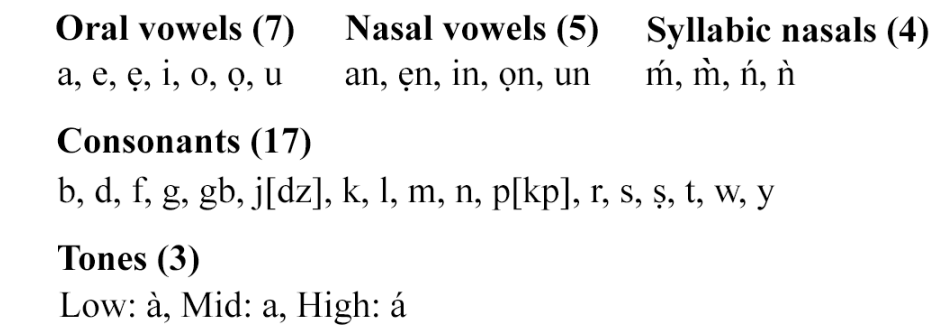

Figure B-1: An overview of the different categories of characters within written Yorùbá, including the diacritical tone markers.

Synonyms				
Ex.	Mode	Word 1	Word 2	Verse
1	English	creature	beast	Gen$1:24 And God said, Let the earth bring forth the living creature after his kind, cattle, and creeping thing, and beast of the earth after his kind: and it was so.
	Yorùbá (auto)	ẹran	ẹran	Gen$1:24 “Ọlọ́run sì wí pé, “Kí ilẹ̀ kí ó mú ohun alààyè jáde ní onírúurú wọn: ẹran ọ̀sìn, àwọn ohun afàyàfà àti àwọn ẹran inú igbó, ọ̀kọ̀ọ̀kan ní irú tirẹ̀.” Ó sì rí bẹ́ẹ̀.“
	Yorùbá (verified)	ẹ̀dá	ẹranko	Gen$1:24 “Ọlọ́run sì wí pé, “Kí ilẹ̀ kí ó mú ohun alààyè jáde ní onírúurú wọn: ẹ̀dá ọ̀sìn, àwọn ohun afàyàfà àti àwọn ẹranko inú igbó, ọ̀kọ̀ọ̀kan ní irú tirẹ̀.” Ó sì rí bẹ́ẹ̀.“
2	English	land	nation	Exod$9:24 So there was hail, and fire mingled with the hail, very grievous, such as there was none like it in all the land of Egypt since it became a nation.
	Yorùbá (auto)	ilẹ	None	Exod$9:24 Yìnyín rọ̀, mọ̀nàmọ́ná sì bẹ̀rẹ̀ sí bù sí orí ilẹ̀ èyí ni ó tí ì buru jù ti ó ṣẹlẹ̀ láti ìgbà ti Éjíbítì ti di orílẹ̀-èdè.“
	Yorùbá (verified)	ilẹ	orílẹ̀-èdè	Exod$9:24 Yìnyín rọ̀, mọ̀nàmọ́ná sì bẹ̀rẹ̀ sí bù sí orí ilẹ̀ èyí ni ó tí ì buru jù ti ó ṣẹlẹ̀ láti ìgbà ti Éjíbítì ti di orílẹ̀-èdè.“
3	English	let	have	Gen$1:26 And God said, Let us make man in our image, after our likeness: and let them have dominion over the fish of the sea, and over the fowl of the air, and over the cattle, and over all the earth, and over every creeping thing that creepeth upon the earth.
	Yorùbá (auto)	jẹ ki	ni	Gen$1:26 "Lẹ́yìn náà ni Ọlọ́run wí pé, “Ẹ jẹ́ kí a dá ènìyàn ní àwòrán ara wa, gẹ́gẹ́ bí àwa ti rí, kí wọn kí ó jọba lórí ẹja òkun, ẹyẹ ojú ọ̀run, ohun ọ̀sìn, gbogbo ilẹ̀ àti lórí ohun gbogbo tí ń rìn lórí ilẹ̀.”
	Yorùbá (verified)	jẹ ki	ni	Gen$1:26 "Lẹ́yìn náà ni Ọlọ́run wí pé, “Ẹ jẹ́ kí a dá ènìyàn ní àwòrán ara wa, gẹ́gẹ́ bí àwa ti rí, kí wọn kí ó jọba lórí ẹja òkun, ẹyẹ ojú ọ̀run, ohun ọ̀sìn, gbogbo ilẹ̀ àti lórí ohun gbogbo tí ń rìn lórí ilẹ̀.”

Figure B-2: Synonym pair examples of how the dataset is constructed from a KJV version of the English bible. In each example, we use NLTK to identify synonyms and antonyms and corresponding verse in English KJV Bible, then use the unique identifier of the verse to find the corresponding verse in Yorùbá KJV Bible. Next, we use Google API to translate the English words to Yorùbá, verify if in Yorùbá verse, and finally we manually inspect and translate words not in Yorùbá verse. Example 1 demonstrates a verse in which the automation is incorrect, and we manually correct it. Example 2 demonstrates a verse in which the automation provide "None" (i.e. no translation existed from Google Translate API), and we manually fill it in. Example 3 demonstrates a verse in which the automation is correct, so the verified version is a repeat.

Antonyms				
Ex.	Mode	Word 1	Word 2	Verse
1	English	day	night	Gen$1:16 And God made two great lights; the greater light to rule the day, and the lesser light to rule the night: he made the stars also.
	Yorùbá (auto)	ọjọ	alẹ	Gen$1:16 Ọlọ́run dá ìmọ́lẹ̀ ńlá-ńlá méjì, ìmọ́lẹ̀ tí ó tóbi láti ṣe àkóso ọ̀sán àti ìmọ́lẹ̀ tí ó kéré láti ṣe àkóso òru. Ó sì dá àwọn ìràwọ̀ pẹ̀lú.
	Yorùbá (verified)	ọ̀sán	òru	Gen$1:16 Ọlọ́run dá ìmọ́lẹ̀ ńlá-ńlá méjì, ìmọ́lẹ̀ tí ó tóbi láti ṣe àkóso ọ̀sán àti ìmọ́lẹ̀ tí ó kéré láti ṣe àkóso òru. Ó sì dá àwọn ìràwọ̀ pẹ̀lú.
2	English	bad	good	Gen$24:50 Then Laban and Bethuel answered and said, The thing proceedeth from the Lord : we cannot speak unto thee bad or good.
	Yorùbá (auto)	None	búburú	Gen$24:50 “Lábánì àti Bétúélì sì dáhùn pé, “Lọ́dọ̀ OLÚWA ni èyi ti wá, nítorí náà àwa kò le sọ rere tàbí búburú fún ọ.”
	Yorùbá (verified)	rere	búburú	Gen$24:50 “Lábánì àti Bétúélì sì dáhùn pé, “Lọ́dọ̀ OLÚWA ni èyi ti wá, nítorí náà àwa kò le sọ rere tàbí búburú fún ọ.”
3	English	father	mother	Gen$20:12 And yet indeed she is my sister; she is the daughter of my father, but not the daughter of my mother.
	Yorùbá (auto)	baba	ìyá	Gen$20:12 "Yàtọ̀ fún ìyẹn, òtítọ́ ni pé arábìnrin mi ni. Ọmọ baba kan ni wá, bí ó tilẹ̀ jẹ́ pé, a kì í ṣe ọmọ ìyá kan."
	Yorùbá (verified)	baba	ìyá	Gen$20:12 "Yàtọ̀ fún ìyẹn, òtítọ́ ni pé arábìnrin mi ni. Ọmọ baba kan ni wá, bí ó tilẹ̀ jẹ́ pé, a kì í ṣe ọmọ ìyá kan."

Figure B-3: Antonym pair examples of how the dataset is constructed from a KJV version of the English bible. Example 1 demonstrates a verse in which the automation is incorrect, and we manually correct it. Example 2 demonstrates a verse in which the automation provide "None" (i.e. no translation existed from Google Translate API), and we manually fill it in. Example 3 demonstrates a verse in which the automation is correct, so the verified version is a repeat.

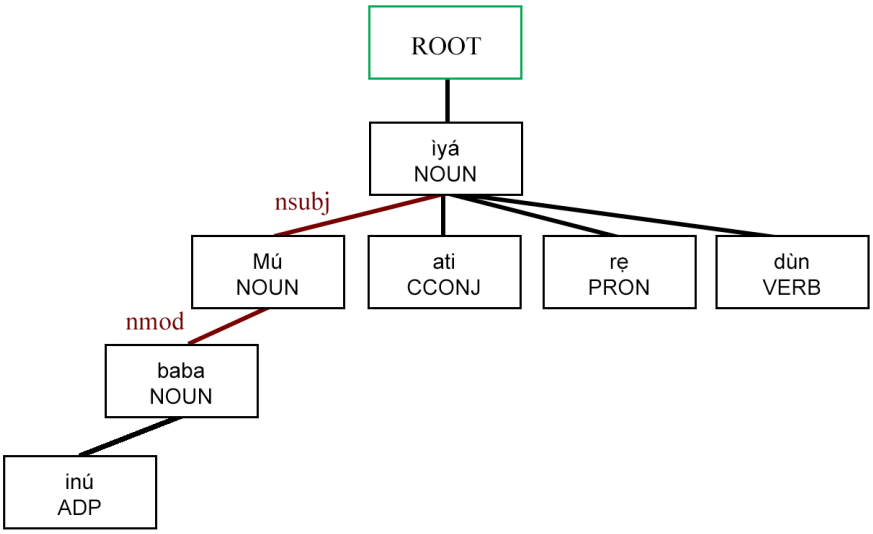

Figure B-4: An example of a Yorùbá dependency parse from UDPipe, with the path from baba (English: father) to ìyá (English: mother) highlighted. The pattern-based neural network classifies word relations via these paths. The phrase is taken from Proverbs in the Yorùbá Bible.

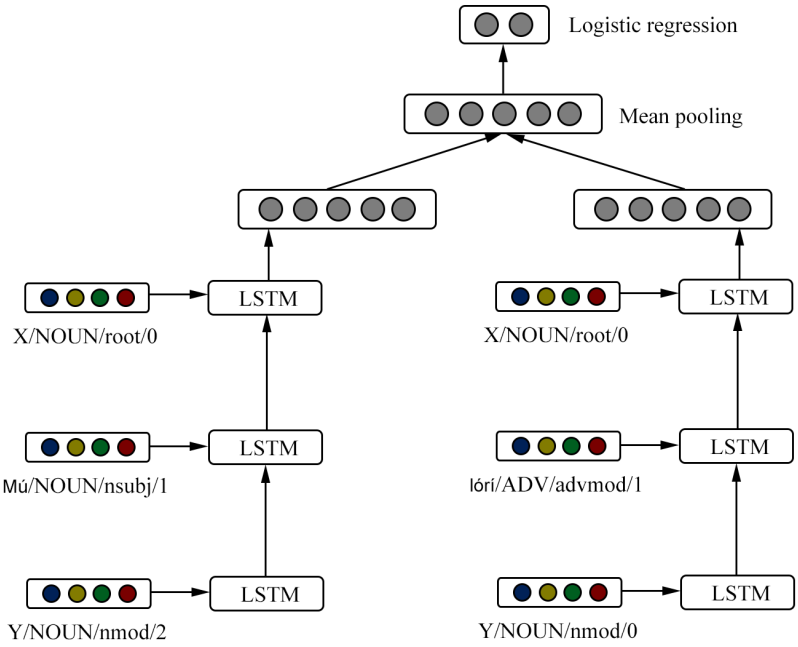

Figure B-5: An overview of the LSTM-based model, one of the two models that we train in this work. Vectors, comprised of the concatenated lemma/POS/dependency tag/distance label, are used as inputs to LSTMs, which then feed into a logistic regression layer. The examples here are from the same word pair in Figure B-4.

Word 1	Word 2	Verse	Ground truth	Predicted label
ibi	bátà	Acts$7:33 "Olúwa sì wí fún un pé, 'Tú bátà rẹ kúrò ni ẹsẹ̀ rẹ, nitorí ibi tí ìwọ gbé dúró sí yìí ilẹ̀ mímọ́ ni."	0	0
ẹ̀bẹ̀	àdúrà	2Chr$6:19 "Síbẹ̀ ṣe àfiyèsí àdúrà ìránṣẹ́ rẹ àti ẹ̀bẹ̀ rẹ̀ fún àánú, OLÚWA Ọlọ́run mi, gbọ́ ẹkún àti àdúrà tí ìránṣẹ́ rẹ ń gbà níwájú rẹ."	0	0
mú	ilẹ̀	Ezek$20:10 "Nítorí náà, mo mú wọn jáde ní ilẹ̀ Éjíbítì mo sì mú wọn wá sínú ihà."	0	0
òkùnkùn	ojìjì	Job$34:22 "Kò sí ibi òkùnkùn, tàbí ojìjì ikú, níbi tí àwọn oníṣẹ́ ẹ̀ṣẹ̀ yóò gbé sá pamọ́ sí."	0	1

Word 1	Word 2	Verse	Ground truth	Predicted label
ìyá	bàbá	Ps$27:10 "Bí ìyá àti bàbá bá kọ̀ mí sílẹ̀, OLÚWA yóò tẹ́wọ́ gbà mí."	1	1
gbóná	tutù	Rev$3:16 "Ǹjẹ́ nítorí tí ìwọ lọ wọ́ọ́rọ́, tí o kò si gbóná, bẹ́ẹ̀ ni tí o kò tutù, èmi yóò pọ̀ ọ́ jáde kúrò ni ẹnu mi."	1	1
obìnrin	ènìyàn	1Sam$25:3 "Orúkọ ọkùnrin náà sì ń jẹ́ Nábálì, orúkọ aya rẹ̀ ń jẹ́ Ábígáílì; òun sì jẹ́ olóye obìnrin, àti arẹ̀wà ènìyàn; ṣùgbọ́n òǹrorò àti oníwà búburú ni ọkùnrin; ẹni ìdílé Kálẹ́bù ni òun jẹ́."	1	1
alágbára	aláìlera	Rom$15:1 "Àwa tí a jẹ́ alágbára nínú ìgbàgbọ́ yẹ kí ó máa ru ẹrù àìlera àwọn aláìlera, kí a má si ṣe ohun tí ó wu ara wa."	1	0

Figure B-6: Examples of correctly classified and incorrectly classified word pairs. The topmost four samples are true synonyms, while the bottom four are true antonyms. To the extent of our analysis, there doesn't seem to be any obvious common features in all of the incorrectly classified samples. For a discussion of the *obìnrin/ènìyàn* example, see Section 3.1.1.

Bibliography

[1] Ethnologue 22nd edition. 2019.

[2] O. Adams, A. Makarucha, G. Neubig, S. Bird, and T. Cohn. Cross-lingual word embeddings for low-resource language modeling. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, 1, 2017.

[3] Philip Atkinson. *A Study Of Our Decline: Civilisation Explained*, pages 57, 163, 171. Lulu.com, 220th edition, 2021.

[4] Segun Awonusi. Linguistic hegemony and the plight of minority languages in nigeria. 07 2008.

[5] Catherine Chen, Kevin Lin, and Dan Klein. Inducing taxonomic knowledge from pretrained transformers. 2020.

[6] Frederick Forsyth. *The Biafra Story*, page 11. Pen & Sword Military, fifth edition, 2015.

[7] John Hewitt and Christopher D. Manning. A structural probe for finding syntax in word representations. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, 2019.

[8] Luyao Huang, Sun Chi, Xipeng Qiu, and Xuanjing Huang. Glossbert: Bert for word sense disambiguation with gloss knowledge, 08 2019.

[9] O. Ishola and D. Zeman. Yorùbá dependency treebank (ytb). 2020.

[10] Chamoda Jeewantha, Sandun Gunasekara, Dulanjaya Chathura, and Gihan Dias. Using annotation projection for semantic role labeling of low-resourced language: Sinhala. 12 2020.

[11] Samuel Johnson. *The History Of The Yorubas: From The Earliest Times To The Beginning Of The British Protectorate*, pages vii, xix. C.M.S (Nigeria) Bookshops Lagos, 1921.

[12] Ahti Lohk, Mati Tombak, and Kadri Vare. An experiment: Using google translate and semantic mirrors to create synsets with many lexical units.

[13] Ryan McDonald, Slav Petrov, and Keith Hall. Multi-source transfer of delexicalized dependency parsers. pages 62–72, 01 2011.

[14] T. Mikolov, I. Sutskever, K. Chen, GS Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. *Neural information processing systems*, 2013.

[15] Kim Anh Nguyen, Sabine Schulte Im Walde, and Thang Vu. Integrating distributional lexical contrast into word embeddings for antonym-synonym distinction. 08 2016.

[16] Kim Anh Nguyen, Sabine Schulte Im Walde, and Thang Vu. Distinguishing antonyms and synonyms in a pattern-based neural network. 04 2017.

[17] Eunice Omolara Olarewaju. Attitudes of nigerian yoruba - english bilinguals in assigning roles to english and mother tongue.

[18] David Smith and Jason Eisner. Parser adaptation and projection with quasi-synchronous grammar features. pages 822–831, 01 2009.

[19] M. Straka, J. Hajic, and J. Straková. Udpipe: trainable pipeline for processing conll-u files performing tokenization, morphological analysis, pos tagging and parsing. *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, 2016.

[20] M. Straka and J. Straková. Tokenizing, pos tagging, lemmatizing and parsing ud 2.0 with udpipe. *Proceedings of the CoNLL 2017*, 2017.

[21] Tatu Vanhanen. *Ethnic Conflicts Explained By Ethnic Nepotism*, volume 7, pages xiii, 186. Emerald Group Publishing Limited, 1999.