Multilingual Dependency Parsing for Low-Resource African Languages: Case Studies on Bambara, Wolof, and Yoruba

Cheikh Bamba Dione
University of Bergen / Sydnesplassen 7, 5007 Bergen
`dione.bamba@uib.no`

Abstract

This paper describes a methodology for syntactic knowledge transfer between high-resource languages to extremely low-resource languages. The methodology consists in leveraging multilingual BERT self-attention model pretrained on large datasets to develop a multilingual multi-task model that can predict Universal Dependencies annotations for three African low-resource languages. The UD annotations include universal part-of-speech, morphological features, lemmas, and dependency trees. In our experiments, we used multilingual word embeddings and a total of 11 Universal Dependencies treebanks drawn from three high-resource languages (English, French, Norwegian) and three low-resource languages (Bambara, Wolof and Yoruba). We developed various models to test specific language combinations involving contemporary contact languages or genetically related languages. The results of the experiments show that multilingual models that involve high-resource languages and low-resource languages with contemporary contact between each other can provide better results than combinations that only include unrelated languages. As far as genetic relationships are concerned, we could not draw any conclusion regarding the impact of language combinations involving the selected low-resource languages, namely Wolof and Yoruba.

1 Introduction

Treebanks constitute valuable resources for many Natural Language Processing (NLP) applications. They can be used as training and testing data for a wide range of NLP algorithms as well as to induce robust parsing models (Manning and Schütze, 1999). Unfortunately, developing treebanks in form of large annotated data used to be a very time- and resource-consuming task. As a consequence, annotated data (in particular the type required for parsing) is lacking for most languages, especially for low-resource languages.

To help speed up the treebank development process, various supervised learning techniques (Weiss et al., 2015; Straka and Straková, 2017; Straka, 2018) have been developed in recent past. The supervised monolingual approach based on syntactically annotated corpora has long been the most common approach to parsing. However, thanks to recent developments involving feature representation methods and neural network models, the idea of combining treebanks for multilingual UD parsing has become a more common approach. Multilingual modeling constitute a very attractive approach to circumvent the low-resource limitation, as it allows one to create models that can parse the language's text quite accurately in the absence of annotated data for the given language. This occurs through syntactic knowledge transfer across multiple languages. The multilingual approach has yielded encouraging results for both low-resource (Guo et al., 2015) as well as for high-resource (Ammar et al., 2016) languages.

The idea of combining treebanks for transfer learning was first introduced in Vilares et al. (2016), which train bilingual parsers on pairs of UD treebanks, showing similar improvements. Subsequently, in the CoNLL 2018 Shared Task, Smith et al. (2018) presented the Uppsala system, which follows the same idea. That system combines treebanks of one language or closely related languages together over 82 treebanks and parses all UD annotations in a multi-task pipeline architecture for a total of 34 models. This approach provides two main advantages. First, it reduces the number of models required to parse each language. Second, it can provide results that are no worse than training on each treebank individually, and in especially low-resource cases, significantly improved. In the same spirit, Kondratyuk and Straka (2019) con-

ducted a multilingual multi-task parsing study for 124 Universal Dependencies (Nivre et al., 2016) treebanks across 75 languages, and demonstrated that a multilingual model can yield better results than monolingual models for different languages.

In this paper, we use the approach described by Kondratyuk and Straka (2019) to produce a cross-lingual transfer model that can predict UD annotations for three extremely low-resource African languages by using knowledge from medium- to high-resource European languages. The UD annotations include universal part-of-speech (UPOS), morphological features (FEATS), lemmas (LEM), and dependency trees (DEPS).

The structure of the paper is as follows. Section 2 first provides a brief description of the low-resource languages used as case studies in this research work. Section 3 provides an overview of our approach, and section 4 details the neural network-based parsing model. Section 5 describes a series of experiments conducted on high-resource and low-resource languages to verify our assumptions. Section 6 presents an analysis of our results. Section 7 concludes the discussion.

2 Languages used as our case studies

The low-resource languages selected for this study are Bambara, Wolof and Yoruba. Bambara is spoken in Mali, Ivory Coast, Upper Guinea, in the western part of Burkina Faso and in eastern Senegal. Wolof is spoken in Senegal, in The Gambia and in Mauritania. Yoruba is spoken in West Africa, most prominently in Southwestern Nigeria.

These West African languages belong to two different subgroups of the larger Niger-Congo family of languages. Bambara is part of the Mande subgroup, while Wolof and Yoruba are Atlantic-Congo languages. While the ultimate genetic unity of Atlantic-Congo languages is widely accepted, the internal cladistic structure is not well established (Dixon et al., 1997), especially with respect to the connection of the Mande languages, which has never been demonstrated. For instance, the Mande languages lack the noun-class morphology that is the primary identifying feature of the Atlantic-Congo languages. Wolof and Yoruba are genetically related to each other, but not closely related to Bambara. Interestingly, while Wolof does not have much language contact with Yoruba (if any), it actually may share areal features with Bambara, since their common geographic location al-

lowed for a long history of contact between these two languages.

Bambara is highly isolating and has a very strict word order: Subject AUX / TAM (tense-aspect-mood markers) Object Verb (Creissels, 2007). It is a tone language, with two tones: high and low. Wolof is an agglutinative language with an SVO and head-modifier basic word order (Robert, 2018). Unlike many other languages of the Niger-Congo family, Wolof is not a tonal language. Yoruba is a highly isolating language and the sentence structure follows Subject Verb Object (Adelani et al., 2021). In addition, Yoruba is a tonal language with three tones: low, middle (optional) and high.

The three low-resource languages are fairly well documented. For Bambara, there exist hundreds of linguistic papers and few recent reference grammars published about that language (Dumestre, 2003; Vydrin, 2019). There are also some dictionaries available, including the Bamadaba online dictionary[1] and a 15k print dictionary (Dumestre, 2011). Likewise, Wolof has several descriptive grammars and few dictionaries, e.g. the French-Wolof print dictionaries (Diouf, 2003; Cisse, 1998) and an online Wolof dictionary.[2] Similarly, for Yoruba, there are many literary texts, newspapers, religious kinds of literature, and some blogs in the language. There are also academic papers, print dictionaries, e.g. the Yoruba-English dictionary by Odoje (2019), and online dictionary[3] published in the language.

Although these languages are well-documented, until very recently, they did not or still do not really have a Universal Dependency corpus. Bambara has a 12k tokens UD treebank (Aplonova and Tyers, 2017). Yoruba has a 8k tokens UD treebank.[4] For these two languages, only test set data are available (no training data), making them good candidates for zero-shot learning. Wolof has a 44k tokens UD treebank (Dione, 2019) that consists of a training, a development and a test set. As these numbers show, the sizes of these UD treebanks are extremely small. This alone does not make them low-resource languages, but they are poorly equipped with regard to NLP tools as well. For instance, while Yoruba and Bambara are documented with huge written corpora,[5] these are mostly not achieved for research

---

[1]http://cormand.huma-num.fr/bamadaba.html

[2]https://www.lexilogos.com/wolof_dictionnaire.htm

[3]https://www.lexilogos.com/english/yoruba_dictionary.htm

[4]https://universaldependencies.org/treebanks/yo_ytb/index.html

[5]For instance, Bambara has an 11 million corpus of writ-

and NLP purposes. For Wolof, resources and tools have only very recently begun to emerge, including a finite-state morphological analyzer (Dione, 2012), a small treebank (Dione, 2014) and computational grammar/parser (Dione, 2020) based on the Lexical-Functional Grammar (LFG) framework (Bresnan, 2001; Dalrymple, 2001). We chose to focus on three languages due to the availability of UD treebanks for these languages, even though for two of these languages only test data are available.

3 Approach

Our approach consists in developing several multilingual parsing models using different language combinations of medium- to high-resource languages (English, French, Norwegian) and low-resource languages (Bambara, Wolof, Yoruba) that have had some contemporary language contact. The languages used for training the models have been selected with the assumption that contemporary contact languages, at least in certain scenarios, share (structural) similarities with the low-resource languages in question. English has a long history of contact with Yoruba, leading to a variety of morphosyntactic changes and lexical borrowings in the latter language (Ogundepo, 2015). Our expectation is that the match rate between English and Yoruba should be somewhat high. Likewise, we expect to see similar patterns between French and the Bambara and Wolof languages with which it has had a long contact. For instance, through French influence there exists two varieties of Wolof: urban Wolof, used especially in cities, and Kajoor Wolof (also referred to as 'pure' Wolof), which is spoken mostly in rural areas (Ngom, 2003). In addition, we include Norwegian as a control language with no direct contact or genetic relation with any of the selected low-resource languages.

Recents studies, including (Lim et al., 2018) conducted similar experiments to explore the impact of using contemporary contact languages or genetically related languages (e.g. Finnish) in multilingual parsing scenarios involving low-resource languages (e.g. North Saami and Komi-Zyrian). Their findings showed that specific language combinations of contemporary contact languages or genetically related languages may enable improved dependency parsing.

ten texts: http://cormand.huma-num.fr/index.html

4 Method

Parsing approaches can be divided into two main types: transition-based (Nivre, 2004) vs. graph-based (McDonald et al., 2005) models. In transition-based dependency parsing, the parser starts in an initial configuration and, at each step, asks a guide to choose between one of several transitions (actions) into new configurations. The parser stops if it reaches a terminal configuration, returning the dependency tree associated with that configuration. In relatively recent past, transition-based dependency parsing using neural networks has enjoyed increasing success, starting with the fast and accurate parser presented by Chen and Manning (2014). Subsequently, many other neural network transition-based models have been developed using different techniques, including stack LSTM (Dyer et al., 2015), biaffine attention (Dozat and Manning, 2016), and recurrent neural networks (Kuncoro et al., 2017).

The basic idea of graph-based dependency parsing is to produce a dependency tree in form of a directed graph with some constraints by first generating all possible candidate dependency graphs for a given sentence. Subsequently, each tree is scored and the parser picks the one with the highest score. During training, the parser induces a model for scoring an entire dependency graph for a sentence. During parsing, it finds the highest scoring dependency graph, given the induced model. More recently, graph-based approaches have shown to outperform transition-based approaches when it comes to UD-type corpora, notably with the neural graph-based parser of Dozat et al. (2017), who won the CoNLL 2017 UD Shared Task by a wide margin.

In this study, we chose UDify (Kondratyuk and Straka, 2019), a neural model which uses the graph-based biaffine attention parser developed by Dozat and Manning (2016); Dozat et al. (2017). UDify is a single multitask model that produces UD annotations (UPOS, FEATS, LEM, DEPS) jointly. In a first step, UDify generates contextual embeddings for any input sentence by using the cased[6] pre-trained multilingual BERT network (Devlin et al., 2018), a self-attention (Vaswani et al., 2017) network of 12 layers, 12 attention heads per layer, and hidden dimensions of 768. The BERT model was trained by predicting randomly masked input

---

[6]https://github.com/google-research/bert/blob/master/multilingual.md

words on the entirety of the top 104 languages with the largest Wikipedias, including two African languages: Swahili and Yoruba. BERT segments texts into (unnormalized) sub-word units using the wordpiece tokenizer (Wu et al., 2016). In a second step, the UDify model integrates task-specific layer-wise attention similar to ELMo (Peters et al., 2018). Finally, each UD task is decoded simultaneously using softmax classifiers. During training, various regularization techniques are applied to the BERT network, including input masking, increased dropout, weight freezing, discriminative fine-tuning, and layer dropout.

5 Experiments

We conducted a series of experiments on Bambara, Wolof and Yoruba. For these languages, we tested different language combinations for the cross-lingual model.

The dataset used in our experiments are provided in Table 1. This consist of a total of 11 UD v2.3 treebanks drawn from three medium- to high-resource languages (English, French, Norwegian) and three low-resource languages (Bambara, Wolof, Yoruba). Table 1 shows the selected treebank(s) used for each language. For English and French, we used several treebanks. For Norwegian, we only selected the Bokmål treebank, leaving out the Nynorsk one in order to reduce computational expenses.[7]

The percentage distribution of the individual languages in our training corpus is shown in Figure 1. As can be seen, ca. 95,5% of the data used in our experiments are drawn from the high-resource languages' treebanks.

Table 2 displays information about the vocabulary of the combined treebanks, including the total number of tokens, BERT wordpieces, UPOS, XPOS, UD Features, lemmas and dependency relations (Deps). To tackle the issue related to a ballooning vocabulary, we use BERT's wordpiece tokenizer directly for all inputs.

For multilingual training with UDify, the 11 UD treebanks are concatenated into a single treebank, similar to McDonald et al. (2011); Kondratyuk and Straka (2019). This single treebank consists of a training, a development and a test set. For each epoch, sample input sentences were drawn randomly from the training data and fed to the neural

network in form of mixed batches, i.e. each batch may contain sentences from any language or treebank. The sentences are shuffled and bundled into batches of 8 sentences each. We employ a base learning rate of $1e^{-3}$ that is kept constant until we unfreeze BERT in the second epoch. We then linearly warm up the learning rate for the next 1,000 batches. Next, we apply inverse square root learning rate decay for the remaining epochs. Following Kondratyuk and Straka (2019), training was done for a total of 80 epochs (ca. 3 days) on a single GPU (RTX 2080). The hyperparameters used for our model are given in Table 3.

6 Results and analysis

For comparison, we show in Table 4 UDify scores obtained for Bambara and Yoruba as reported by Kondratyuk and Straka (2019). These scores are obtained by evaluating UDify on 124 treebanks with the official CoNLL 2018 Shared Task evaluation script.

The experiments reported by Kondratyuk and Straka (2019) did not include Wolof, since no UD treebank was available for that language at that time. For this purpose, we trained a customized monolingual UDify model on the Wolof training data and applied that model on the Wolof test set. The results of this monolingual training are shown in Table 5. These scores are used as a baseline to compare the impact of the monolingual and the multilingual models.

For multilingual dependency parsing, we run several experiments in which we keep the settings described in section 5, excluding only one language at a time. In an initial experiment, we used all the treebanks presented in section 5 for which training data are available. This consists of a total of 9 out of the 11 UD treebanks.[8] Then, the created multilingual model has been used to parse the test data of the selected low-resource languages. The results are given in Table 6 and indicate an improvement of ca. 5% and 4,38% in terms of UAS and LAS, respectively for Bambara. Likewise, a significant increase of 11,37% and 12,34% in UAS and LAS, respectively, has been observed for Yoruba. Also, for Wolof, we compared the scores displayed in Table 5 (i.e. monolingual model) with those presented in Table 6 (i.e. multilingual model). Such a comparison reveals that parsing quality increases by 2.23%

---

[7]In fact, we have tried to include the Nynorsk treebank as well, but this led to quite computationally expensive costs when training the multilingual model.

[8]Recall that the Bambara and Yoruba treebanks contain only test data.

| Language	Treebank Name	# Tokens	# Sentences
Bambara	Bambara CRB	13.82k	1.03k
English	UD_English_EWT	254.83k	16.62k
	UD_English-GUM	80.18k	4.40k
	UD_English-ParTUT	49.62k	2.09k
French	UD_French-FTB	556.06k	18.53k
	UD_French-ParTUT	27.67k	1.02k
	UD_French-Sequoia	68.64k	3.10k
	UD_French-Spoken	34.98k	2.79k
Norwegian	UD_Norwegian-Bokmaal	310.22k	20.04k
Wolof	Wolof WTB	42.83k	2.11k
Yoruba	Yoruba YTB	2.67k	0.10k

Table 1: UD treebanks used in the experiments
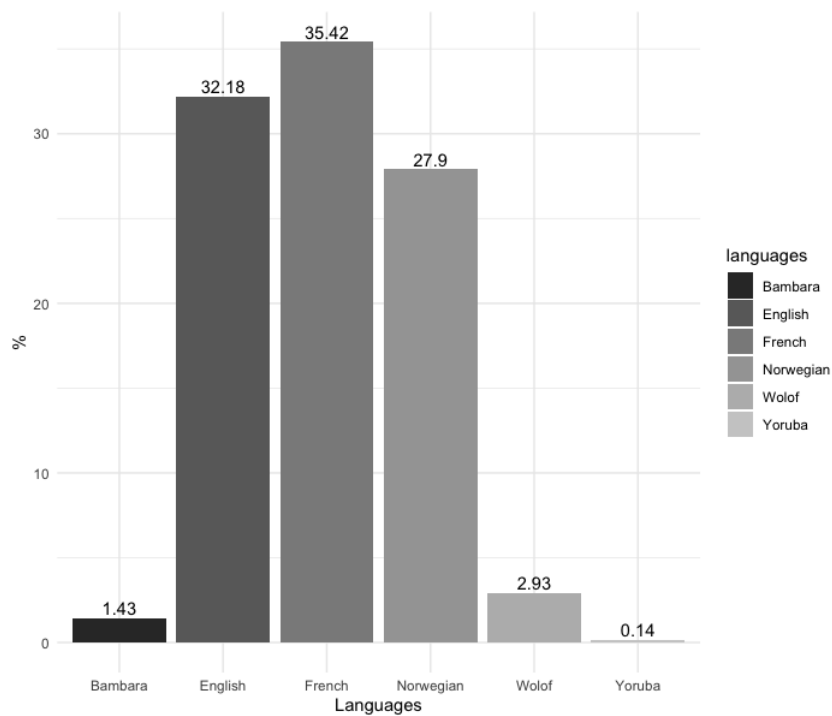
Figure 1

TOKEN	VOCAB SIZE
Word Form	178,214
BERT Wordpieces	119,547
UPOS	17
XPOS	206
UFeats	1300
Lemmas (tags)	3,834
Deps	86

Table 2: Vocabulary sizes of words and tags over the 11 UD v2.3 treebanks

hyperparameter	
Dependency tag dimension	256
Dependency arc dimension	768
Optimizer	Adam
β_1, β_2	0.9, 0.99
Weight decay	0.01
Label Smoothing	0.03
Dropout	0.5
BERT dropout	0.2
Mask probability	0.2
Layer dropout	0.1
Batch size	8
Epochs	80
Base learning rate	$1e-3$
BERT learning rate	$5e-5$
Learning rate warmup steps	1000
Gradient clipping	5.0

Table 3: Our model hyperparameters.

Treebank	UPOS	FEATS	LEM	UAS	LAS
Bambara	30.86	57.96	20.42	30.28	8.60
Yoruba	50.86	78.32	85.56	37.62	19.09

Table 4: The full test results on Bambara and Yoruba when training UDify on the **124** UD treebanks (Kondratyuk and Straka, 2019).

Language	UPOS	FEATS	LEM	UAS	LAS
Wolof	87.90	69.97	87.86	73.48	63.84

Table 5: The test results on Wolof when training UDify on monolingual Wolof data.

and 3.19% in terms of UAS and LAS, respectively. This provides a good indication that a transfer learning approach in a multilingual dependency parsing context seems to have positive impact (at least for Wolof), outperforming the monolingual model.

To assess the impact of some language combina-

Treebank	UPOS	FEATS	LEM	UAS	LAS
Bambara	36.29	45.18	23.89	34.28	12.98
Wolof	89.64	75.06	89.95	75.71	67.03
Yoruba	60.73	60.43	93.59	48.99	31.43

Table 6: The full test results on Bambara, Wolof and Yoruba when training UDify on the **9** UD treebanks (the Bambara and Yoruba treebanks are only used for testing).

Treebank	UPOS	FEATS	LEM	UAS	LAS
Bambara	35.05	44.32	23.84	33.83	12.94
Wolof	90.49	77.25	90.82	76.14	67.86
Yoruba	59.19	60.18	93.21	45.19	28.53

Table 7: The full test results on Bambara, Wolof and Yoruba when training UDify on the **8** UD treebanks, **excluding the English treebanks**.

tions, we run several additional experiments where we keep the same setting and language data as described above, excluding only one language at a time. Accordingly, we run a similar experiment as the previous one, excluding the English treebanks from the training. The results of this experiment are displayed in Table 7. For Bambara, excluding the English treebank did cause a very slight drop of parsing quality. In contrast, for Yoruba, we could observe a substantial decrease of 3.8% and 2.9% UAS and LAS, respectively. Interestingly, for Wolof, this actually led to a slight improvement in UAS (0.43%) and LAS (0.83%).

In the same way, we run a similar experiment where we exclude only the French treebanks to assess their impact on the overal results for the selected low-resource African languages. The results of this experiment are shown in Table 8. For Bambara, this caused a decrease of 3.78% and 2.85% UAS and LAS, respectively. Parsing quality also drops for Wolof in terms of UAS (-2.28%) and LAS (-1.74%). For Yoruba, no real impact on parsing quality could be observed.

Treebank	UPOS	FEATS	LEM	UAS	LAS
Bambara	33.89	42.86	23.66	30.50	10.13
Wolof	89.61	82.33	91.53	73.43	65.29
Yoruba	60.32	55.78	91.19	48.69	30.87

Table 8: The full test results on Bambara, Wolof and Yoruba when training UDify on the **7** UD treebanks, **excluding the French treebanks**.

As mentioned above, we used Norwegian as a control language to verify our assumption with re-

Treebank	UPOS	FEATS	LEM	UAS	LAS
Bambara	36.02	45.08	23.87	33.94	11.94
Wolof	90.76	83.55	90.91	75.35	67.25
Yoruba	59.94	58.68	88.30	48.09	30.87

Table 9: The full test results on Bambara, Wolof and Yoruba when training UDify on the **10** UD treebanks, **excluding the Norwegian treebank**.

Treebank	UPOS	FEATS	LEM	UAS	LAS
Bambara	29.44	69.08	13.36	30.91	11.21
Wolof	23.02	45.78	65.95	27.51	10.15
Yoruba	67.44	59.26	78.39	52.74	33.80

Table 10: The full test results on Bambara, Wolof and Yoruba when training UDify on the **10** UD treebanks, **excluding the Wolof training treebank**.

spect to the impact of using genetic or geographical relation. Norwegian is selected, as it is a language with no direct contact or genetic relation with any of the studied low-resource languages. To verify our assumption, we run an additional experiment where we removed the Norwegian data from the training, keeping the remaining 10 UD treebanks as before. The results of this experiment are provided in Table 9. This operation does not seem to have a substantial impact on parsing quality for any of the three low-resource languages. For instance, for both Bambara and Yoruba, a slight drop in UAS and LAS could be observe, but the decrease is less than 1% in all these cases. For Wolof, even a slight improvement could be observed of ca. 0.22% in LAS only (compared with -0.36% drop in UAS).

In a final experiment, we wanted to test the impact of not using the Wolof data during training. Thus, we trained a model using the 10 treebanks, excluding the Wolof UD training set and applied the model to the three test sets of the studied low-resource languages (this means that we evaluated Wolof for zero-shot learning). The results of this experiment are given in Table 10. Interestingly, for Bambara, this operation caused a decrease of parsing quality of 3.37% UAS and 1.77% LAS. For Wolof, as expected, we noted parsing accuracy dropped drastically by 48.2% UAS and 56.88% LAS. This large drop in parsing result can be explained by the fact that the Wolof test set is relatively large (e.g. compared to the test sets for Bambara and Yoruba). Surprisingly, for Yoruba, removing the Wolof data in the training had a positive impact. Parsing quality for Yoruba increased by ca. 3.75% UAS and 2.37% LAS. At first glance, this seems to contradict our expectation that genetically related languages may enable improved dependency parsing, at least for our case study. But a crucial question to consider is whether the genetic relationship between these two languages is just a classification issue and that, from the language char-

acteristics, these two languages are not so closely related as the classification would suggest. Based on our data and experiments, we could not draw any conclusion as to whether the obtained results emerge from an issue related to the data used or to the language classification or to something else. This might need further investigation.

7 Conclusion

In this paper, we have presented a multilingual approach to parsing that is effective for languages with few resources and no syntactically annotated corpora available for training. We have shown that specific language combinations involving contemporary contact languages can provide better results than combinations that only include unrelated languages. We should note, however, that for Wolof and Yoruba, which are supposed to be genetically related languages, we rather observed a decrease of parsing results from the Yoruba side when using the Wolof training data. It remains a question for further study whether this decrease observed here are actually attributable to a lack of real genetic relationship between the language or to the lack of (large) training data for Yoruba.

References

David Ifeoluwa Adelani, Jade Abbott, Graham Neubig, Daniel D'souza, Julia Kreutzer, Constantine Lignos, Chester Palen-Michel, Happy Buzaaba, Shruti Rijhwani, Sebastian Ruder, et al. 2021. Masakhaner: Named entity recognition for african languages. *arXiv preprint arXiv:2103.11811*.

Waleed Ammar, George Mulcaire, Miguel Ballesteros, Chris Dyer, and Noah A. Smith. 2016. Many languages, one parser. *Transactions of the Association for Computational Linguistics*, 4:431–444.

Ekaterina Aplonova and Francis M. Tyers. 2017. Towards a dependency-annotated treebank for Bambara. In *Proceedings of the 16th International Workshop on Treebanks and Linguistic Theories*, pages 138–145, Prague, Czech Republic.

Joan Bresnan. 2001. *Lexical-Functional Syntax*. Blackwell, Oxford.

Danqi Chen and Christopher Manning. 2014. A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 740–750, Doha, Qatar. Association for Computational Linguistics.

Mamadou Cisse. 1998. *Dictionnaire wolof-français*. Editions Karthala, 75013 Paris.

Denis Creissels. 2007. A sketch of bambara argument structure. In *Workshop on Grammar and Processing of Verbal Arguments. Leipzig*, pages 20–21.

Mary Dalrymple. 2001. *Lexical-Functional Grammar (Syntax and Semantics, Volume 34) (Syntax and Semantics)*. Academic Press.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Cheikh Bamba Dione. 2019. Developing universal dependencies for wolof. In *Proceedings of the Third Workshop on Universal Dependencies (UDW, SyntaxFest 2019)*, pages 12–23, Paris, France. Association for Computational Linguistics.

Cheikh M. Bamba Dione. 2012. A Morphological Analyzer For Wolof Using Finite-State Techniques. In *Proceedings of the Eigth International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey. ELRA.

Cheikh M Bamba Dione. 2014. LFG parse disambiguation for Wolof. *Journal of Language Modelling*, 2(1):105–165.

Cheikh M. Bamba Dione. 2020. Implementation and evaluation of an LFG-based parser for Wolof. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 5128–5136, Marseille, France. European Language Resources Association.

Jean-Léopold Diouf. 2003. *Dictionnaire wolof-français et français-wolof*. Editions Karthala, Paris.

Robert MW Dixon, Robert MW Dixon, and Dixon Robert Malcolm Ward. 1997. *The rise and fall of languages*. Cambridge University Press.

Timothy Dozat and Christopher D. Manning. 2016. Deep biaffine attention for neural dependency parsing. *CoRR*, abs/1611.01734.

Timothy Dozat, Peng Qi, and Christopher D. Manning. 2017. Stanford's graph-based neural dependency parser at the CoNLL 2017 shared task. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 20–30, Vancouver, Canada. Association for Computational Linguistics.

Gérard Dumestre. 2003. *Grammaire fondamentale du bambara*. KARTHALA Editions.

Gérard Dumestre. 2011. *Dictionnaire bambara-français, suivi d'un index abrégé français-bambara*. KARTHALA Editions.

Chris Dyer, Miguel Ballesteros, Wang Ling, Austin Matthews, and Noah A. Smith. 2015. Transition-based dependency parsing with stack long short-term memory. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 334–343, Beijing, China. Association for Computational Linguistics.

Jiang Guo, Wanxiang Che, David Yarowsky, Haifeng Wang, and Ting Liu. 2015. Cross-lingual dependency parsing based on distributed representations. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1234–1244, Beijing, China. Association for Computational Linguistics.

Dan Kondratyuk and Milan Straka. 2019. 75 Languages, 1 Model: Parsing Universal Dependencies Universally. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2779–2795, Hong Kong, China. Association for Computational Linguistics.

Adhiguna Kuncoro, Miguel Ballesteros, Lingpeng Kong, Chris Dyer, Graham Neubig, and Noah A. Smith. 2017. What do recurrent neural network grammars learn about syntax? In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 1249–1258, Valencia, Spain. Association for Computational Linguistics.

KyungTae Lim, Niko Partanen, and Thierry Poibeau. 2018. Multilingual dependency parsing for low-resource languages: Case studies on north saami and Komi-Zyrian. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Christopher D Manning and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA.

Ryan McDonald, Fernando Pereira, Kiril Ribarov, and Jan Hajič. 2005. Non-projective dependency parsing using spanning tree algorithms. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 523–530, Vancouver, British Columbia, Canada. Association for Computational Linguistics.

Ryan McDonald, Slav Petrov, and Keith Hall. 2011. Multi-source transfer of delexicalized dependency parsers. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 62–72, Edinburgh, Scotland, UK. Association for Computational Linguistics.

Fallou Ngom. 2003. The social status of arabic, french, and english in the senegalese speech community. *Language Variation and Change*, 15(3):351.

Joakim Nivre. 2004. Incrementality in deterministic dependency parsing. In *Proceedings of the Workshop on Incremental Parsing: Bringing Engineering and Cognition Together*, pages 50–57, Barcelona, Spain. Association for Computational Linguistics.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. 2016. Universal Dependencies v1: A Multilingual Treebank Collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France. European Language Resources Association (ELRA).

Clement Odoje. 2019. *Yoruba-English/ English-Yoruba Dictionary & Phrasebook* . Hippocrene Books.

Abimbola O Ogundepo. 2015. Contrastive study of english and yoruba morphological systems–implications for nigerian teachers and learners of english. *International Journal of English Language and Linguistics Research*, 3(4):1–8.

Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*.

Stéphane Robert. 2018. Wolof: a grammatical sketch. In Friederike Lüpke, editor, *The Oxford guide to the Atlantic languages of West Afric*. Oxford University Press.

Aaron Smith, Bernd Bohnet, Miryam de Lhoneux, Joakim Nivre, Yan Shao, and Sara Stymne. 2018. 82 treebanks, 34 models: Universal Dependency parsing with multi-treebank models. In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 113–123, Brussels, Belgium. Association for Computational Linguistics.

Milan Straka. 2018. UDPipe 2.0 prototype at CoNLL 2018 UD shared task. In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 197–207, Brussels, Belgium. Association for Computational Linguistics.

Milan Straka and Jana Straková. 2017. Tokenizing, pos tagging, lemmatizing and parsing ud 2.0 with udpipe. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 88–99, Vancouver, Canada. Association for Computational Linguistics.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need.

David Vilares, Carlos Gómez-Rodríguez, and Miguel A. Alonso. 2016. One model, two languages: training bilingual parsers with harmonized treebanks. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 425–431, Berlin, Germany. Association for Computational Linguistics.

Valentin Vydrin. 2019. *Cours de grammaire bambara*. Inalco presses.

David Weiss, Chris Alberti, Michael Collins, and Slav Petrov. 2015. Structured training for neural network transition-based parsing. *arXiv preprint arXiv:1506.06158*.

Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Lukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. *CoRR*, abs/1609.08144.