

New syntactic insights for automated Wolof Universal Dependency parsing

Bill Dyer

University of Florida / Dept. of Linguistics 4131
Turlington Hall P.O. Box 115454
Gainesville, FL 32611
wgdyer@ufl.edu

Abstract

Focus on language-specific properties with insights from formal minimalist syntax can improve universal dependency (UD) parsing. Such improvements are especially sensitive for low-resource African languages, like Wolof, which have fewer UD treebanks in number and amount of annotations, and fewer contributing annotators. For two different UD parser pipelines, one parser model was trained on the original Wolof treebank, and one was trained on an edited treebank. For each parser pipeline, the accuracy of the edited treebank was higher than the original for both the dependency relations and dependency labels. Accuracy for universal dependency relations improved as much as 2.90%, while accuracy for universal dependency labels increased as much as 3.38%. An annotation scheme that better fits a language’s distinct syntax results in better parsing accuracy.

1 Introduction

Wolof is a language of Senegal, where it is the lingua franca in a nation of more than twelve million people (McLaughlin, 2008). About six million speak Wolof as their first language (Eberhard et al., 2020). However, it is severely underrepresented in print, as well as in digital format. Because French is the official language of the Senegalese state, writing is more commonly practiced in French, while Wolof and other indigenous languages are used more in spoken communication.

Out of the almost 120 languages for which there are universal dependency treebanks available, only eight are indigenous African languages¹. African languages in particular are not well represented, given Africa’s large share of the world’s languages and the relatively large populations of even minority language groups. The presence of these annotated treebanks is promising for automated computational tasks, though.

¹<https://universaldependencies.org>

The aim of this study is to improve universal dependency (UD) parsing for Wolof. A UD treebank by Cheikh Bamba (Dione, 2019) is available for the Wolof language. The innovation proposed here is not only to train a parser, as (Dione, 2020) has already designed a Wolof language-specific parser. The purpose was to determine whether out-of-the-box parsers would show improvement on the Wolof treebanks after edits to the part of speech and universal dependency syntax annotations. Improved Wolof parser models may be used to inform other African language parsers, whose features can be analyzed on their own terms rather than through the lens of other major languages or existing annotation schemes.

2 Hypothesis

The assignment of syntactic dependencies in the Wolof UD treebank is based on syntactic and morphological analysis from lexical functional grammar (Dione, 2019). In some cases, natural language processing has ignored language-specific features, a sacrifice that is to some degree necessary to create a universal system like UD syntax. This is especially true of languages with less presence in scholarly literature, where tagging or parsing assignments attempt to fit languages into the mold of world languages like English (Tovey, 2019). The dependency structures of determiners, pronouns, and copulas in the Wolof UD Treebank comply with the traditional functions of those categories, but can be realigned to capture broader linguistic generalizations of their behavior while improving accuracy.

2.1 Relative clauses

Wolof determiners and relative clause pronouns are represented by identical morphemes. Determiners follow nouns, as in example 1 (Njie, 1982; Ka, 1994). They consist of a consonant that corresponds with the noun class and a vowel

that corresponds to deictic configuration (Njie, 1982; Robert, 2006). Wolof has a large number of classes (or genders) for nouns (McLaughlin, 1997), and 18 classes are represented in the Wolof UD Treebank. This class of words includes definite determiners and demonstratives that designate the distance of the object from the speaker.

- (1) a. *cin l-i*
 large.pot LClass-the
 ‘the large pot’
 b. *jamono j-ooju*
 era JClass-that.far
 ‘that time long ago’

In the Wolof UD Treebank, such determiners are tagged DET in both the universal UD tagset and the Wolof-specific tagset. Their UD dependency label is *det*, and they are dependents of the noun.

Wolof relative clauses appear with an overt noun head, or as ‘headless’ relative clauses. Examples of headed relative clauses from the training and development data of the Wolof UD Treebank are those in 2.

- (2) a. *làkk y-ii ñu*
 language YClass-these we
nàmp
 learn.as.mother.tongue
 ‘these languages here that we learn as a mother tongue’

In the Wolof UD Treebank, the relative pronouns in headed relative clauses are tagged as PRON in both the universal and Wolof-specific tag assignments. They are given the dependency relation label that corresponds to their role in the embedded clause, such as *nsubj* or *obj*, and are a dependent of the verb embedded in the relative clause.

The examples in 3 are headless relative clauses. The relative clause pronouns are made of the same class consonant and vowel combinations that signify distance from the speaker.

- (3) a. *k-i leen*
 ClassK-the them
taxawal-oon
 stand.up.against-PAST
 ‘the one who stood against them’

The relative pronoun in these headless relative clauses are also given the PRON tag in the Wolof UD Treebank, for both their universal part-of-speech (POS) tag, as well the Wolof tagging system established by Dione. Unlike the dependency relation for headed relative pronouns, the embedded relative clause verb is a dependent of the relative pronoun. Alternatively, adopting an SUD annotation scheme would also result in parallel structures where the closed class determiner is the head. SUD provides further evidence that the strict application of UD syntactic policy does not always result in the most accurate parsing (Gerdes et al., 2018).

There are other clauses that have the same structure as relative clauses, such as temporal and conditional clauses beginning with *bu* or *su* (Torrence, 2013). If relative clauses are uniformly assigned a similar dependency structure, regardless of whether they have a head or not, the parser should be able to better recognize them. The idea that the relative pronoun should consistently be represented as a functional head with the same position in the syntax is clear from minimalist syntax as outlined by (Chomsky, 1996).

Furthermore, all definite determiners, demonstratives, relative pronouns and quantifiers follow nouns, provide more information about the noun and agree in noun class. As such, I hypothesize that if they are all labeled with the same tag, the part-of-speech tagger in the parser pipeline will be more accurate.

An annotation scheme where the definite-determiner-like morpheme is a complementizer should result in a better trained parser than one where it is a determiner, following the syntactic analysis of (Torrence, 2013).

2.2 The existence of copulas

A similar trend occurs with those words that have been tagged COP in the Wolof language specific tags, all of which are AUX in the universal tag set. Syntactic analyses have identified several copulas in Wolof (Torrence, 2013), although each are associated with other function words. *La* indicates

complement (as opposed to subject or verb) focus in a clause. *Di* (and its allomorph *y*) indicates progressive aspect. The examples in 4 show them in copular sentences. The *-a-* morpheme in 4b indicates subject focus. Capital letters represent focus of any kind in the glosses.

- (4) a. *Kolle sama mag la*
 Kolle my older.sibling FOCUS
 ‘Kolle is my OLDER SISTER’
- b. *Abdu mo-o-y sama mag.*
 Abdu he-FOCUS-is my older.sibling
 ‘ABDU is my older brother’

The examples in 5, however, show the use of *la* and *di* as verbal auxiliaries that indicate focus and progressive aspect, respectively (Ka, 1994). In such cases, *la* is tagged as INFL in the Wolof tagset, and *di* is tagged as AUX. Both are tagged as AUX in the universal tagset.

- (5) a. *Kolle kànj la jënd.*
 Kolle okra FOC.he sell
 ‘Kolle has sold OKRA.’
- b. *Kolle kànj la-y jënd.*
 Kolle okra FOC-he-is sell
 ‘Kolle is selling OKRA.’
- c. *Kolle mo-o jënd kànj.*
 Kolle he-FOC sell okra
 ‘KOLLE has sold okra.’
- d. *Kolle mo-o-y jënd kànj.*
 Kolle she-FOC-is sell okra
 ‘KOLLE is selling okra.’

Wolof does allow null copulas, which must be the case in sentences like 5c. It is more likely that the function morphemes assigned as copulas are in fact function morphemes in all cases, and that there are no overt copular verbs, a phenomenon attested in many languages. While the interpretation of lexical functional grammar presented by (Dione, 2019) would attempt to match the function of a verb with the words present, the minimalist analysis elaborated by (Chomsky, 1996) does not require an overt lexical item to occupy a syntactic position. The verb position may be empty in certain cases, allowing the focus complementizer morphemes like *la*, following (Martinović,

2017), and imperfect morphemes like *di* to consistently maintain their roles rather than be circumstantially designated as verbs. Attempts to make a language fit the mold of other languages more commonly tested in natural language processing are what (Tovey, 2019) predicts will increase confusion in tasks like part-of-speech tagging.

The function words that determine focus and verbal aspect are consistent in their syntactic distribution, whether used in copular contexts or not.

3 Methods

The original data in the Wolof UD Treebank (Dione, 2019) consists of 42,832 tokens across 2,107 sentences. These sentences are from four different Wolof sources online: the Organisation Sénégalaise dAppui au Développement (Senegalese Aid and Development Organization) web site, Wolof Online, Wolof Wikipedia, and the news site Xibaaryi. They were divided by Dione into training, test, and development sets.

Table 1: Sources of Corpora for the Wolof UD Treebank (Dione, 2019)

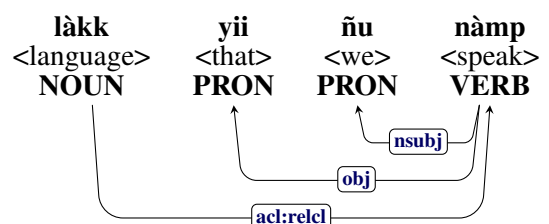
Source	# Doc.	# Tok.	# Sent.
OSAD	6	6,269	265
Wolof Online	18	12,988	673
Wolof Wikip.	12	9,232	500
Xibaaryi	17	15,095	669

Using Python, the test, development, and test files of the Wolof UD Treebank were edited to assign certain lemmas new tags in certain environments and assign new UD labels and hierarchies.

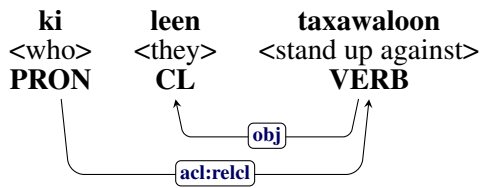
3.1 DEF model with relative pronoun dependency labeled ‘mark’

First, the universal dependency relations of headless relative clauses were edited. Examples of headed and headless relative pronouns can be seen in 6 and 7 respectively, illustrating the structural difference in the baseline.

(6) Headed Relative Clause

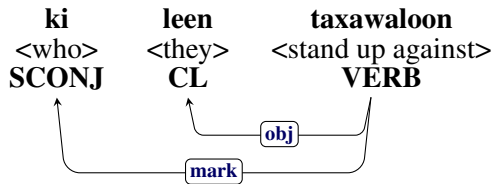


(7) Headless Relative Clause



Following the hypothesis that similar syntactic structures will have similar dependency structures, headless relative clauses like 7 were edited to take on the dependency hierarchy in 8. In this way, all relative clauses are given the same dependency structure, whether they have an overt noun head or not.

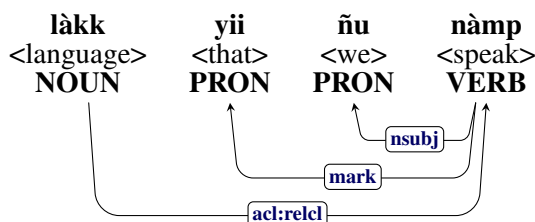
(8) Adjusted dependency relations for headless relative clauses to match headed relative clauses



The tags of each definite determiner, relative pronoun, post-nominal quantifier, or clausal complementizer that agrees in noun class are changed from DET, PRON, and COMP to a new class; DEF. The universal equivalent of the Wolof-specific COMP tag is SCONJ. This includes the complementizers *bu* and *su*, which (Torrence, 2013) analyzes as being the relative pronouns of headless conditional relative clauses. The edited treebanks will be the input for the first parser model.

In this model, all relative pronouns in headed and headless relative clauses are labeled as the universal dependency relation *mark*, which signifies complementizers in the UD annotation. Even the headed relative clauses had their part-of-speech tags and universal dependency relation labels changed.

(9) Adjusted dependency relations for headed relative clauses



3.2 DEF model with relative pronoun dependency labeled ‘det’

The analysis of the relative pronoun as complementizer is the one that Torrence favors, but another hypothesis that he tests is that they are determiners. This competing analysis is tested computationally by a second parsing model. The edited treebanks for this model use the label *det* for relative clause pronouns rather than *mark*.

3.3 Relabeled copula model

All copular tags are edited in treebanks designated to be the input to a third model. In this model, all COP tags for selected lemmas are changed to INFL and AUX. These are lemmas that are assigned INFL and AUX tags in non-copular contexts. The assignment of AUX or INFL is somewhat changed, however, based on the category of the lemma. The following lemmas that sometimes acted as copulars are given with their alternate POS in Table 2.

Table 2: Alternative tag assignment for select lemmas when not assigned COP

INFL		AUX	
Lem.	Funct.	Lem.	Funct.
la	Compl. Foc	ngi	Prog. Asp.
da	Verb. Foc	du	Neg.
daan	Pst. Hab Asp., foc. cl.	daan	Pst. Hab Asp., non-focus cl.
		di	Imp. Asp.

One issue with this classification is that it divides AUX and INFL into irregular categories. Some of the lemmas in each category designate focus, while others designate aspect. Instead, AUX and INFL are reassigned to these lemmas based on the classification given in Table 3. INFL will be assigned for focus particles and negative *du*, which appears in the same syntactic position as focus elements do. AUX will be assigned to auxiliaries denoting aspect, but is also used for particles that are not used as copulas and are not on this list. The universal tags for these lemmas goes unchanged, as the Wolof-specific tags for AUX and INFL are both labeled AUX in the universal tag system.

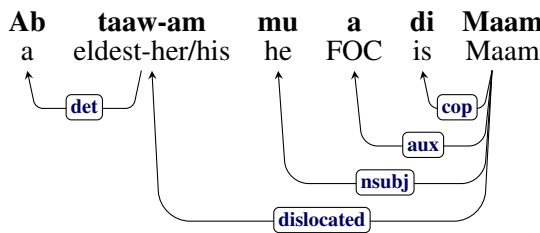
Table 3: Category reassigned to selected lemmas previously assigned COP, AUX, or INFL tag

INFL		AUX	
Lem.	Funct.	Lem.	Funct.
la	Comp. foc.	ngi	Prog. Asp.
da	Verb. foc.	daan	Pst. Hab. Asp.
du	Neg	di	Imp. Asp.

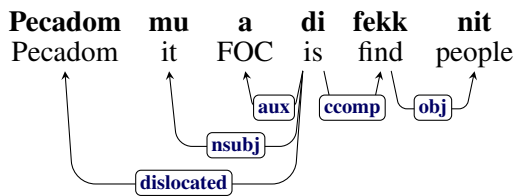
Words that were originally assigned as COMP are retagged, as well as those lemmas in the table that were assigned INFL or AUX. This leads into somewhat reduced granularity in part of speech tags, but there is a diverse distribution of each dependency structure outside of copulas.

The dependencies in subject focus constructions are also conflated into a similar structure. Nominals are treated as roots in copular clauses with subject focus and a nominal complement, as in 10. However, in copular clauses with subject focus and a clausal complement, as in 11, the imperfect auxiliary *di* is treated as the root.

(10) *Di* copula with nominal complement

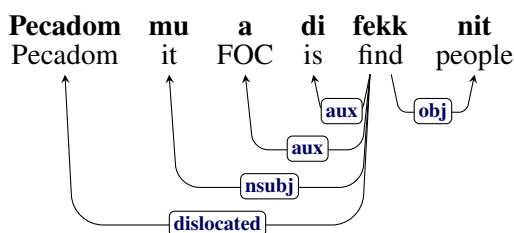


(11) *Di* copula with clausal complement



The dependencies for sentences like 11 will be changed into the structure in 12 in the treebanks for the third model.

(12) Reassigned dependency for clausal *di* copula to match nominal complement



The relative clause verb *fekk*, ‘find,’ is now the root. In the UD system, the verb of the clause acts as the root, meaning that copular verbs can be roots. In Diones lexical functional grammar analysis (Dione, 2019), the subject focus morphemes (*mooy = mu+a+di*) in 11 act as a copula. The copula implies a cleft, such as the English translation "Its Pecadom who finds people that are sick in their house." I reject the cleft analysis, and follow Martinovićs analysis for the similar *la* complement focus construction (Martinović, 2017). Like *la*, the *-a* in 11 and 12 are complementizers. *Di* is a morpheme that marks imperfect aspect. It is not a copula and does not result in a cleft. The main verb in 12 is *fekk*, ‘find’, rather than a copula, making it the root.

3.4 Parser pipelines

After the treebanks are edited, they are prepared for either the spaCy parser pipeline, or the Tree-Tagger+MaltParser parser pipeline.

3.5 The SpaCy pipeline

For spaCy, the treebank UD treebanks for all three models are converted to .json format. There are separate treebanks for the train, development, and test data. The parser is trained using each of the three sets of edited treebanks. A baseline parser model was also trained using the data from the unedited Wolof UD Treebank.

Four separate models have been created; one baseline model, one model from the DEF tag treebanks with relative pronouns as *mark*, a third from the DEF tag treebanks with relative pronouns as *det*, and a fourth from the treebanks that are edited to replace the COP tag. The trained models are evaluated using Python. The accuracy of the universal dependency label assigned was measured against the baseline, as well as the accuracy of the universal dependency hierarchies.

3.6 The MaltParser pipeline

The second parser consisted of two separate tools: TreeTagger (Schmid, 1994) and MaltParser (Nilsen and Nivre, 2008). The baseline and all three edited treebanks were used to train Tree-Tagger models. TreeTagger requires the tag SENT for punctuation marking the end of the sentence. The Wolof tags PERIOD, EXL-POINT, SEMICOLON, ELLIPSIS and INT-MARK were changed to SENT for use on TreeTag, as their corresponding lemmas ‘.’, ‘!’, ‘;’, ‘...’, and ‘?’

were used to separate sentences in the Wolof UD Treebank. After models were trained, a treebank file was produced that tagged the words from the test Wolof UD Treebank. The treebank file took combined tags from a universal tagger and Wolof-specific tagger that were trained for each parser pipeline model. A baseline tagger was also trained based on the original Wolof UD Treebank.

The treebank files only contained the word number, word form, lemma, universal POS tag, and Wolof POS tag for each word. This is all that could be produced by the tagger. After a treebank file was prepared for each model, it was used as input into MaltParser. A MaltParser model was trained on the baseline Wolof UD Treebank training data, as well as the edited treebank data for each model. The trained edited models were all tested against the baseline; for accuracy of the UD labels as well as the UD structural hierarchy.

3.7 Combination of DEF+'det' model and Relabeled Copulas model

After testing was completed, the COP model and the DEF tag model that showed the highest improvement in accuracy are combined. A combined model was made for both spaCy and the TreeTagger-MaltParser pipeline.

4 Results

Accuracy was improved for the models made for each parser pipeline; the spaCy and the TreeTagger+Malt Parser pipelines. Table 5 shows the results for both the labels assigned to the universal dependency relations, as well as the hierarchical structure of the universal dependencies.

Table 4: Accuracy for UD labels and relations with spaCy pipeline

#	Annot.	UD Label	Univ. Dep.
0	Baseline	76.4%	71.1%
1	DEF tag, RC pron as <i>det</i>	77.9%	71.7%
2	DEF tag, RC pron as <i>mark</i>	77.8%	71.4%
3	Copulas Relabeled	77.4%	71.2%
4	Combination of #1 and #3	78.0%	71.4%

Table 5: Accuracy for UD labels and relations with TreeTagger+MaltParser pipeline

#	Annot. Label	UD Dep.	Univ.
0	Baseline	72.7%	70.4%
1	DEF tag, RC pron as <i>det</i>	74.9%	72.9%
2	DEF tag, RC pron as <i>mark</i>	74.0%	73.2%
3	Copulas Relabeled	73.9%	70.7%
4	Combination of #1 and #3	76.1%	73.3%

Models 1 and 2 with the DEF tag showed drastic improvement in SpaCy UD Label, the Malt UD Label, and the Malt universal dependency accuracies when compared to the test data. The SpaCy UD label increased 1.5% for the model with the *det* label, the Malt UD labels increased 2.2%, and the Malt universal dependencies increased 2.5%. The model with the *mark* label improved SpaCy UD labels by 1.4%, Malt UD labels by 1.3%, and Malt universal dependencies by 2.8%. Targeting a separate set of syntactic dependencies, relabeling copulas also showed across the board increases in accuracy. Accuracy for the SpaCy UD labels improved 1%, .8% for the Malt UD labels, and .3% for the Malt universal dependencies. Improvement was less in the SpaCy universal dependencies, which showed a maximum of .6% improvement. When the changes made to the best DEF model treebanks were made to the Relabeled Copula treebanks, the SpaCy UD label accuracy increased 1.6%, SpaCy universal dependency accuracy increased by .3%, Malt UD label accuracy increased by 3.4%, and Malt universal dependencies increased by 3.3%.

Relative clauses pronouns were relabeled as determiners (*det*), which show an increase in recall and f1-score. As relative clause pronouns were made determiner dependents of the relative clause verb, and no longer confused with subjects (*nsubj*), objects (*obj*), obliques (*obl*), and indirect objects (*idobj*), their precision, recall and f1-scores increase in the improved parser. There are no copula dependency labels in the improved parser, and the relabeling of 182 copulas in the auxiliary (*aux*) category resulted in an increase in precision, recall, and f1-score for auxiliaries (*aux*).

5 Analysis

Overall, adopting a unified and streamlined syntactic approach to assigning UD relations improves accuracy in Wolof. Two different parsers both showed improvement in parsing when a DEF tag was added to definite noun modifiers, headless relative clauses had the same structure as headed ones, and copulas were relabeled to capture their universal function. This suggests that improved accuracy was not simply due to the parsers.

The results supported the hypothesis that a unified UD syntax for headed and headless relative clauses improves the accuracy in parsing universal dependencies and their labels. The hypothesis that treating definites as one part of speech category would improve parsing was supported by the results. The hypothesis the relative pronoun is a complementizer due to theoretical syntactic analysis was not supported by the data. In fact, the model that treats the relative pronoun as *det*, an extracted determiner, results in more accurate parsing.

The copular analysis carried over from English does not seem to ‘fit’ Wolof. The data from Wolof does not contradict an analysis where *di* and its allomorphs universally indicates imperfect aspect, rather than acting in some instances as a copula. The subject and object focus morphemes are the same whether the sentence is copular or not, suggesting that they are not copulas in copular sentences. The copula should be instead attributed to some null morpheme. Improved parsing accuracy resulting from the reassigning of copula tags and dependency relations in the Wolof UD Treebank supports this hypothesis.

As (Dione, 2019) mentions, the morpho-syntactic assignments of the Wolof UD treebanks, and the universal dependency program in general, are based on lexical functional grammar. In this view, whatever lexical item is the semantic head of the relative clause must have the rest of the relative clause as its dependents. The same is true for morphemes that were labeled as copulas; the subject, object and other arguments of the sentence would be dependents of this morpheme. In other cases, however, the same lexical item would be swapped and the dependency relationship completely inverted. These cases involve the same lexical items, but apply the semantic role of another missing element to them.

By adopting syntactic assumptions from the

minimalist syntax formalism (Chomsky, 1996), a unified structure can be preserved with the UD framework. The minimalist framework allows for the assumption that the missing element is simply not overtly pronounced. Although common in many languages, the need for a copular verb or overt relative clause head need not outweigh the evidence that verbs or relative clause heads may simply be null items. A legitimate UD structure can still be attained while maintaining a consistent roll for these words. Such consistency better reflects the findings of (Tovey, 2019) that language-specific particularities should not be diluted in annotation to accommodate more commonly analyzed languages.

As editing the treebanks improves parsing from the baseline, the morphological and syntactic annotations made here should improve future parsers. (Dione, 2020) trained a Wolof-specific lexical functional grammar parser with 67% recall, 93% precision and an f-score of 78%. The most accurate parser model in this study had 78% recall, 78% precision, and an f-score of 78%. The significance of this study is not the accuracy of the parser itself, but the improvement from the baseline. The baseline-trained spaCy parser had 76% recall, 76% precision, and an f-score of 76%, meaning that each measure improved by 2%. This improvement should carry over if implemented with future Wolof UD parsers.

6 Conclusion

This study proposed considerations for improving the parsing of Wolof, one of the few African languages represented in a UD treebank. In the cases of relative clauses, assigning tags and dependency relations of definites based on their distribution and features provides a better parse than trying to distinguish them as pronoun in certain cases and determiners and demonstratives in others, following patterns from Indo-European languages. Positing a unified dependency structure for relative clauses also improves parsing. The idea of a copula imported from copular sentences in other languages does not fit the syntax of Wolof. Rather, classifying part-of-speech tags and labels based on their function in the clause results in a more accurate parse.

Although the UD framework is lexically oriented, and is more readily translated from lexical functional grammar, insights from the minimalist

framework can inform morphological and syntactic annotation. These edited treebank annotations lead to improved parsing in the case of Wolof, and are likely to be useful for related African languages.

While the Wolof UD parser by (Dione, 2019) has similar accuracy, the fact that two out-of-the-box parsers showed improvements with the edited annotations is promising. The final accuracy achieved by this parser is similar to Dione’s, and suggest that future parsers can attain even greater accuracy if these treebank annotation edits were combined with Dione’s parser. Wolof is a low-resource language with only one treebank, also created by (Dione, 2019). 2% is a small but valuable improvement given accuracies of 75%–80% and a smaller treebank relative to languages like English, French, and Russian. The improvements made to parsing compared to the baseline provide guidance for future annotation of African language treebanks, which are not proportionally represented in the UD project.

References

- Noam Chomsky. 1996. *The Minimalist Program*. Current Studies in Linguistics. MIT Press.
- Cheikh Bamba Dione. 2019. Developing universal dependencies for wolof. In *Third Workshop on Universal Dependencies (UDW, SyntaxFest 2019)*, Paris, France.
- Cheikh Bamba Dione. 2020. Implementation and evaluation of an lfg-based parser for wolof. In *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, pages 5128–5136, Marseille, France. European Language Resources Association (ELRA).
- David M. Eberhard, Gary F. Simons, and Charles D. Fenning, editors. 2020. *Ethnologue: Languages of the World*, 23 edition. SIL International, Dallas, Texas USA.
- Kim Gerdes, Bruno Guillaume, Sylvain Kahane, and Guy Perrier. 2018. *SUD or surface-syntactic Universal Dependencies: An annotation scheme near-isomorphic to UD*. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pages 66–74, Brussels, Belgium. Association for Computational Linguistics.
- Omar Ka. 1994. *Wolof phonology and morphology*. University Press of America.
- Martina Martinović. 2017. Wolof wh-movement at the syntax-morphology interface. *Natural Language and Linguistic Theory*, 35:205–256.
- Fiona McLaughlin. 1997. Noun classification in wolof: When affixes are not renewed. *Studies in African Linguistics*, 26(1):1–28.
- Fiona McLaughlin. 2008. Senegal: the emergence of a national lingua franca. In Andrew Simpson, editor, *Language and National Identity in Africa*, pages 79–97. Oxford.
- Jens Nilsson and Joakim Nivre. 2008. Malteval: An evaluation and visualization tool for dependency parsing. In *Sixth international conference on Language Resources and Evaluation*, Marrakech, Morocco.
- Codu Mbassy Njie. 1982. *Description Syntaxique du Wolof de Gambie*. Les Nouvelles Éditions Africaines.
- Stéphane Robert. 2006. Deictic space in wolof: discourse, syntax and the importance of absence. In Maya Hickmann and Stéphane Robert, editors, *Space in languages: linguistic systems and cognitive categories*, volume 66 of *Typological Studies in Language*, pages 155–174. John Benjamins.
- Helmut Schmid. 1994. Probabilistic part-of-speech tagging using decision trees. In *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK.
- Harold Torrence. 2013. *The Clause Structure of Wolof*. John Benjamins, Philadelphia.
- Bethan Siân Tovey. 2019. You’re not the pos of me: part-of-speech tagging as a markup problem. In *Proceedings of Balisage: The Markup Conference 2019*, volume 23 of *Balisage Series on Markup Technologies*, Washington, DC USA. Basilage: The Markup Conference 2019.

A Data availability statement

The data that support the findings of this study are openly available in the directory wolofUDParsing at <https://github.com/BillDyer/wolofUDParsing>.