LAYING THE GROUNDWORK FOR NATURAL LANGUAGE PROCESSING (NLP) IN NGIEMBOON: A DESCRIPTIVE STUDY OF ITS PART-OF-SPEECH SYSTEM

Patrice Yemmene [1,4], Prosper Djiaffeua [2], Basile Difouo [3,4]

[1] Department of Computer Science, University of Wisconsin Milwaukee
[2] University of Yaoundé I, Cameroon
[3] University of Maroua, Cameroon
[4] Laboratoire,Technologies Educatives, Langues, Cultures et Humanités, ACETELACH, Yaoundé, Cameroun

ABSTRACT

In this paper, we discuss the necessity for enhancing NLP capabilities for African under-resourced languages, particularly those spoken in Cameroon. We use the Ngiemboon language as a focal point for developing innovative tagging solutions. We lay the groundwork for creating a part-of-speech (POS) tag set for the Ngiemboon language, focusing on a descriptive study of its parts of speech.

We establish NLP as an interdisciplinary field that automates language understanding and generation, highlighting various applications such as machine translation and chatbots. We emphasize the role of POS tagging as a fundamental step in NLP. We highlight the linguistic description of the language as a prerequisite for the development of POS. One aspect of linguistic description is the morphosyntactic analysis of the language, which is essential for understanding linguistic structures and enabling more complex language processing tasks. We emphasize the importance of a well-structured tag set, which should be informed by detailed linguistic analysis.

KEYWORDS

Part of speech, Natural Language Processing (NLP), Under-resourced language

1. INTRODUCTION

Natural Language Processing (NLP) encompasses the advanced application of several fields of study, including (but not limited to) computer science, artificial intelligence (AI), Mathematics, Statistics, Linguistics, and other related disciplines. NLP aims to automate processes that enable machines to read, decipher, understand, and make sense of natural human language.

NLP applications are powered by language models grouped based on NLP algorithms needed for their development. A few of the models include named entity recognition, word segmentation (tokenization), lemmatization, morphological segmentation, word-sense disambiguation, stemming, and part-of-speech tagging. The later (part-of-speech tagging) is of critical importance, especially for under-resourced languages. "*The significance of parts-of-speech … for language processing is the large amount of information they give about a word and its*

David C. Wyld et al. (Eds): CSIT, AMLA, IPDCA, NLPA, AIS, IPPR, SPTM – 2025
pp. 147-156, 2025. CS & IT - CSCP 2025 DOI: 10.5121/csit.2025.151411

neighbours... Knowing whether a word is a possessive pronoun, or a personal pronoun can tell us what words are likely to occur in its vicinity... This can be useful in a language model for speech recognition" (Jurafsky, D. & Martin, J., 2024). The applications of part-of-speech tagging are very useful for downstream NLP tasks such as automatic speech recognition (ASR), automatic syntactic parsing, automatic machine translation (AMT), and many other NLP tasks.

Generally, "*many challenges in NLP involve natural language understanding, that is, enabling computers to derive meaning from human or natural language input, and others involve natural language generation. The solution for language understanding is part-of-speech tagging.*" (Adhvaryu, N.; Balani, P. 2015). Part-of-speech tagging is a very early and important step in NLP research. In fact, "*a crucial step for a part-of-speech tagging system for a language is a well-designed, consistent, and complete tag set*" (Onyenwe, Ikechukwu E. et al., 2019).

2. OBJECTIVE OF THE STUDY

With the advancement of research in artificial intelligence and the constant influx of new deep learning algorithms, combined with the availability of high computing resources, there has been significant progress in the development of part-of-speech tagging in various languages. A few deep learning algorithms that exhibit a high degree of satisfaction when applied to part-of-speech tagging in NLP include Multilayer perceptron (MLP), recurrent neural networks (RNN) algorithms, Long Short-Term Memory (LSTM)with Multilayer perceptron (MLP), Bidirectional long short-term memory (BLSTM), Gate recurrent unit (GRU), Feed-forward neural network (FFNN), Deep neural network (DNN), Convolutional neural network (CNN) (Chiche, A., Yitagesu, B).

The absolute prerequisite for this implementation is a detailed linguistic study and analysis of the target language. In this paper, we propose a detailed study and description of the Ngiemboon part-of-speech system and its critical importance to the development of a state-of-the-art POS tagging system for this language. "*The significance of parts-of-speech for natural language processing is the large amount of information they give about a word and it's neighbours*". Knowing a word part-of-speech can also aid in its accurate pronunciation, when developing a speech synthesis system, or an automatic speech recognition. Field such as information retrieval (IR) would also benefit from a descriptive study of part of speech because the part-of-seapch with give a hint about the morphological affixes that it can take. NLP algorithms such as word parsing, word sense disambiguation, shallow information extraction will produce better results when a word part-of-speech is known.

3. THE CHOICE OF THE NGIEMBOON LANGUE

Large corporations such as Microsoft, IBM, Google, and the like are investing massively in languages with significant marketing and commercial potential. Minority languages, especially those spoken in sub-Saharan Africa, have very little attraction for these companies. The reason is obvious: these languages don't have a significant economic, scientific, or political strength, neither at a global level, nor at a regional level. At least, this is true today.

Yet, these languages are vectors of communication and development of communities where they are spoken. The reason is obvious: most of the speakers of these languages are monolingual and depend solely on their language to access resources and participate in decision-making. These languages constitute the medium used by members of these linguistic communities to connect to the imperative and joy of the information needed to develop themselves and their community. From personal observation, no community around the world has ever achieved sustainable

development using exclusively a foreign language. Therefore, investing in human language technologies in these languages will make a significant contribution to the sustainable development of communities where they're spoken. We want to contribute to the effort of bridging the gap between high-resourced languages and under-resourced languages by contributing to the democratization process of artificial intelligence**.**

4. LARGE LANGUAGE MODELS AND UNDER-RESOURCED LANGUAGES

Nowadays, large language models (LLMs) dominate the discussion when people talk about artificial intelligence. LLMs represent state-of-the-art, cutting-edge, advanced research in natural language processing. "*Building on traditional language models (LMs) like N-gram models, LLMs address limitations such as rare word handling, overfitting, and capturing complex linguistic patterns. Notable examples, such as GPT-3 and GPT-4, leverage the self-attention mechanism within transformer architectures to efficiently manage sequential data and understand long-range dependencies. Key advancements include in-context learning for generating coherent text from prompts and Reinforcement Learning from Human Feedback (RLHF) for refining models using human responses*" (Parthasarathy, Venkatesh B. et al. 2024).

However, LLMs require massive language corpora for efficient results. Nevertheless, existing pre-trained LLMs can be fine-tuned on lesser subject-specific datasets, specific fields of study, topics, or industries. "*Fine-tuning transfers the pre-trained model's learned patterns and features to new tasks, improving performance and reducing training data needs. It has become popular in NLP for tasks like text classification, sentiment analysis, and question-answering*" (Parthasarathy, Venkatesh B. et al. 2024).

With the popularity, accuracy level, and effectiveness of LLMs, the following question seems to carry a significant amount of weight: shouldn't we all turn to LLMs for answers to research questions in natural language processing? To answer the question, we say this: we don't have all the details to suggest a direction in one way or the other, especially about resource-rich languages such as English and others. However, it's important to note that LLMs are "*challenged by hallucination, reliability, huge computational complexity, and lack of incremental learning capabilities. To tackle these deficiencies, an alternative approach could be the decomposition of a generic LLM to several smaller domain-specific mid-size language models (MLM) that have a "multi-modal interface" to handle textual or visual input/output and a "knowledge core" implemented by knowledge graphs (KGs) for knowledge representation, data mining, and incremental learning. Such a modular design could improve the interpretability, reliability, computational complexity, and incremental capability of next-generation MLMs with justifiable and logical reasoning. In this direction, POS tagging is still a valuable step in building interpretable NLP models"* (Wei, Chengwei; Pang Runqi. 2024).

This apprehension is not only applicable to resourced languages but is critically extendable to under-resourced languages, where there are almost no existing language corpora. Therefore, we understand the need to extend the excitement for the development of part-of-speech taggers to under-resourced languages.

5. RAPID SURVEY OF THE NGIEMBOON LANGUAGE

The Ngiemboon language is an Eastern Grassfields Bantu language spoken in two divisions (Bamboutos and Menoua) of the West Region of Cameroon (Sonkoue, Eliane, 2020b). Initial scientific research work on the language dates back to the early 70s, when Stephen C Anderson, the then Summer Institute of Linguistics (SIL) researcher and PhD student at the Linguistics

Department of the University of Southern California in Los Angeles (UCLA), settled in the geo-linguistic area of the language to decipher stories that the language had to tell.

Over the years, he has remained a prolific researcher of the Ngiemboon language, contributing to the linguistic description that led to the development of a basic standardization of the language. Despite significant efforts made thus far, further scientific discovery is still needed to bring the language up to the status of a high-resourced language.

6. NGIEMBOON LANGUAGE: AN UNDER-RESOURCED LANGUAGE

Despite the significant amount of research carried out on the Ngiemboon language, and the enthusiasm observed by its linguistic community to promote it in the written form, the Ngiemboon language remains an under-resourced language (Yemmene, Patrice; Laurent Besacier. 2019). An under-resourced language is "*a language with some (if not all) of the following aspects: lack of a unique writing system or stable orthography, limited presence on the web, lack of linguistic expertise, lack of electronic resources for speech and language processing such as monolingual corpora, bilingual electronic dictionaries, transcribed speech data, pronunciation dictionaries, vocabulary lists, etc*." (Berment, 2004).

From personal observation, this statement is an accurate rendering of the linguistic description reality of the Ngiemboon language, the availability of written documentation in the language, and the presence of the language in the digital space.

This suggests that the development of a part-of-speech tagging for the under-resourced Ngiemboon language will present significant challenges, including but not limited to the availability of labelled data needed to properly train POS taggers, and many other challenges. Yes, these resources are critical for the development of NLP tasks in general, and POS taggers in general.

The morphonology of the language is very complex, and the current research level hasn't yet deciphered all the story that it has to tell. In fact, "*one of the most complex aspects of Eastern Grassfields languages is the number of tonal perturbations … Even more complex are the many tonal morphemes that affect verb roots in complicated verbal constructions. While Eastern Grassfields languages are noteworthy for their lack of productive verbal suffixes with segmental material, they more than make up for this lack by the number of tonal morphemes that surround the verb. The presence of these many tonal morphemes is only revealed by the vast variety of surface tones found on verb roots in their various verbal constructions*" (Anderson, 2001). This complexity will only add to the challenges of developing a tag set in this under-resourced language.

7. NGIEMBOON LINGUISTIC FEATURES WITH POTENTIAL NLP IMPLICATIONS

7.1. Phonological Characteristics

After several years of research, Stephen C. Anderson (2008) concluded that the Ngiemboon phonology is very complex, requiring in-depth and sustained analysis, and cannot be easily characterized. Nevertheless, he distinguishes globally two morpheme types in the language: roots and affixes.

Referring to the syllabic structure of the word root of the language, Anderson further states that "*Ngiemboon roots consist of the following: C1 (S1) V1 (C2) (V2), i.e., an obligatory root-initial consonant, an optional semivowel, an obligatory vowel, and optional final consonants and vowels*". Furthermore, "*Ngiemboon prefixes (on nouns and verbs) and some pronouns may, in addition to simple CV(C) syllables, have just a syllable nucleus, consisting of either a vowel or of a homorganic nasal consonant (symbolized as "N"). Ngiemboon suffixes are only two in number: /-te/ or a vowel*"

The language has 16 underlying consonants. The phonetic realization of many of those consonants varies depending on whether they appear before or after the word root. The realization of some of the consonants may also depend on the vowel that follows, as well as unpredictable tonal variation.

		Labials	**Coronals**	**Velars**
Stops:	Voiceless		t	k
	Voiced	b	d	g
Affricates:		pf	ts	
Fricatives:	Voiceless	f	s	
	Voiced	v	z	
Nasals:		m	n	ŋ
Semivowels:			j	w

Ngiemboon phonetic consonants proposed by Stephen C Anderson

The language has 7 underlying vowels. Some of the vowels can take a long form, while others can take a nasal form. Vowel variations may take place depending on morpho-phonological rules.

SHORT ORAL	[+ front] [- back] [- round]	[+ front] [- back] [+ round]	[- front] [- back] [- round]	[- front] [- back] [+ round]	[- front] [+ back] [- round]	[- front] [+ back] [+ round]
[+ high] [- low]	i	y			ɯ	u
[- high] [- low]	e		ə		ɤ	o
[- high] [+ low]	ɛ		a			ɔ

Ngiemboon phonetic vowels proposed by Stephen C Anderson

7.2. Tonal Characteristics

The Ngiemboon language is tonal. *"Ngiemboon and other Eastern Grassfields Bantu languages are known for the complexity of their tone systems. The most complex part of these systems is found in their tone perturbations, how tones of individual words change when they are put into sentences"* (Anderson, 2014). Some perturbations and complexity are observed by the occurrence of tonal morphemes, that represent, in some scenario, the grammatical category of a noun, or a verb tense, mode or aspect.

A tonal language is a language that has "*morphemes whose surface pitch (acoustically understood as the fundamental frequency with which corresponds the articulatory rate at which the vocal cords vibrate at any point in time) patterns contrast with each other in one or more*

comparable environments" (Snider, 2017). The Ngiemboon language has four main tone melodies on noun and verb stems: rising, downstepped high, low, and low-falling.

8. NGIEMBOON PART OF SPEECH

The notion of part-of-speech is deeply embedded in the notion of grammatical categories. The notion of grammatical categories in Bantu languages is different from the notion of grammatical categories in Indo-European languages, because of the unique peculiarities of Bantu languages (Mutaka, Nguessimo and Chumbow, Sammy. 2021). Based on the studies by Etienne Lonfo (2016), the Ngiemboon language has the following grammatical categories: nominal, verbal, adjective (qualifiers, Possessives, demonstratives, interrogatives, cardinal numbers, participle "N"), pronouns (personal, general subject pronoun, possessive pronouns, demonstrative, relatives and interrogatives, indefinite pronouns), prepositions, adverbs (quantity, place, time, manner, etc.), adverb of negation, ideophone adverbs, conjunctions, coordination, subordination (complementizers, other subordinate conjunctions) … The sub-sections below, describing the Ngiemboon grammatical categories, are inspired by the grammatical studies carried out by Lonfo.

8.1. Nominals

The notion of **nouns**, as a grammatical category, is extremely complex in African languages, especially Bantu languages (Mutaka, Ngessimo M. and Chumbow, Sammy. 2021). Noum in Bantu languages are not categorised based on masculine or feminine. Instead, they're grouped in noun classes, and each class is represented by an affix that is added to the word root based the prefix of the preceding word. At times, it's difficult to tell the different between a verb and a noun in the Ngiemboon language, when the word is in isolation.

For the context of this research, we'll be substituting ***nouns*** for **nominals**., because it more closely aligns with the reality of Bantu languages. Therefore, we'll be referring to the nominal grammatical category. BOT BA NJOCK (1970) describes the nominal as "*a constituent which, in a statement with a verbal predicate, is capable of ensuring non-predicative functions. This functional definition of the nominal allows us to establish two large groups of nominals which contrast from a syntagmatic perspective. Therefore, we can group as independent nominals and dependent nominals, because of the independence of some and the obligation for others to agree with the first*". Let's note that in Bantu languages spoken in Cameroon, there is no notion of genres (masculine or feminine).

In Ngiemboon, the nominal grammatical category has a very complex morphological structure. In a phrase or a sentence, the nominal has a great influence on other grammatical categories. This influence is translated by the nominal's latitude of combination with other parts of speech concerning nominal modalities. Based on these nominal modalities, Anderson, Stephen C. (1980a) grouped nominals in the Ngiemboon language into 10 classes, each of which has an associated nominal prefix, consonant concord marker, vowel concord marker, and tone concord marker. He proposed the chart below, as a high-level summary of the 10 Ngiemboon nominal classes.

Noun Class	Noun Prefix	Typical Example	Concord markers		
			Consonant	Vowel	Tone
1	Varies	Pfom "ant"	w	a	L
2	me-	me-pfom "ants"	p	e	H
3	N-	n-ka? "field"	w	e	H
4	meN-	men-ka? "fields"	m	e	H
5	le-	le-pfo "kolanut"	s	e	H
6	me-	me-pfo "kola nuts"	m	e	H
7	(a)-	a-poon "sack"	y	a	H
8	(e)	e-poon "sack"	s	a	H
9	N-	n-tu? "calabash"	y	e	L
10	N-	n-ko? "roosters"	y	e	H

There are morphological changes that take place in other grammatical categories such as adjectives and pronouns, to agree with the nominal class of the nominal that they appear in the same phrase or sentence. This morphological agreement modification also takes place in some cases with verbs in a sentence.

In addition to nominals, there are also verbo-nominals. These are nominals that are different from other nominals solely by their verbal origins. They are derived from verbs and have a verbal root. All of this makes the nominal morphology of the Ngiemboon language a very complex one.

8.2. Verbs

Verbs constitute another grammatical category in the Ngiemboon language. Anderson, Stephen C. (op cit) groups Ngiemboon verbs into two groups, based on their morphophonological changes at the supra-segmental level:

- High tone verbs: "*In the infinitive, high tone verb roots are preceded by high tone nasal prefixes and also carry a high tone. Examples: ńná to give; ḿbá' to weave; éshúate to crush and ńzá'te to carve up. In the second person imperative form, these roots also carry high tones: Examples: tóo come; náa give; pá'a weave; zá'te carve up and shúate crush."*
- Mid-tone verbs: "*In the infinitive, mid-tone verb roots are preceded by a high tone nasal prefix and carry mid-tones. Examples: éfa to blow; ḿbu' to perforate and ńcuŋte to spoil. In the second-person imperative form, these roots also carry mid-tones. Examples: faa blow; pu'u perforate and cuŋte spoi*l."

Verbs in Ngiemboon have a root, prefixes, and suffixes, and can have various tenses: the present, the past tense, and the future tense. There are also verb moods such as the infinitive, the indicative, the imperative, the subjunctive, and the conditional moods.

8.3. Adjectives

The Ngiemboon language has several types of adjectives:

- Qualifiers adjectives: They usually appear before the nominal in a noun phrase. In various cases, it will agree with the nominal class of the nominal.
- Possessive adjectives: Possessive adjectives usually occur after the nominal, and almost always agree with the nominal class of the nominal. Furthermore, there are singular possessive pronouns and plural possessive pronouns.

- Demonstrative adjectives: Demonstrative adjective morphology fluctuates based on the nominal class of the nominal in the noun phrase. It will also vary based on whether the object is close or far or somewhere in between.
- Interrogative adjectives: The derived form of the interrogative adjectives is subject to the nominal class of the nominal in the noun phrase where it appears.
- Cardinal Numbers: cardinal numbers always agree with the noun class of the nominal that they occur in a noun phrase with. Except for "ta", all cardinal numbers appear in a nominal syntagma after the nominal.
- Participle" N" adjectives: "*The participle is a verb root that becomes an adjective by adding a nasal prefix with low tone "/Ǹ-/*" (Anderson, Stephen C. (op cit).

8.3.1. Pronouns

The Ngiemboon language has various types of pronouns:

- Personal pronouns: There are singular and plural pronouns in the ngiemboon language. For each category, there is the 1st person, the second person, and the third person.
- General subject pronouns: These always occur before a verb and agree with the nominal class of the name that they replace.
- Possessive pronouns: The language has possessive pronouns, both plural and singular. Each category has 1st person, second person, and third person. Possessive pronouns agree as well with the nominal class of the nominal that they are referring to.
- Demonstrative pronouns: The Ngiemboon language equally has three demonstrative pronouns, that indicate whether the object is close, or far, or in between. Each one of the pronouns concords with the nominal classes indicated above.
- Relatives and interrogative pronouns: The Ngiemboon language has relatives and interrogative pronouns that concord with the nominal classes as well.
- Indefinite pronouns: There's only one indefinite pronoun in the language.

8.4. Prepositions

The Ngiemboon language has several propositions. The following have been identified: " *"á" at, in; "né" at, in, for; "lé" at, with, in, for; "ngwaa" before; "lâ" with; "pâ" with; "lê" with; "nê" with; "mbɔɔ" even, with, and; "tà" until; "ndùm" above; "tsèɛ" in; "tsiŋe" below; "tsetsèɛ" in the midst of, between; "nzèm" behind; "mvfò" in front of; "gwŏŋ" next to*" (Stephen C Anderson (op cit)

8.5. Adverbs

The Ngiemboon has several types of adverbs: adverbs of negation, ideophone adverbs, adverbs of quantity, adverbs of place, adverbs of time, adverbs of manner, adverbs of conjunction, and adverbs of coordination. Most of them occur after the verbs in a sentence.

9. CONCLUSION

In this paper, we proposed a descriptive study of the Ngiemboon part-of-speech. This research contributes to laying the groundwork further NLP research in the Ngiemboon language. One of the next tasks that we'll accomplish is the development of a part-of-speech tagging system for the Ngiemboon language. From there, other systems such as automated translation system or automatic system recognition, just to name a few could bd further developed.

REFERENCES

[1] A. Y. Kassahun and T. G. Fantaye, "*Design and Develop A Part of Speech Tagging for Ge'ez Language using Deep Learning Approach*," 2022 International Conference on Information and Communication Technology for Development for Africa (ICT4DA), Bahir Dar, Ethiopia, 2022, pp. 66-71, doi: 10.1109/ICT4DA56482.2022.9971153.

[2] Anderson, Stephen C. 1980a. *The Noun Classes of Ngyembɔɔn-Bamileke*, in Larry M. Hyman (ed.), Noun Classes in the Grassfields Bantu Borderland, Southern California Occasional Papers in Linguistics, Volume 8, Los Angeles: University of Southern California, pp. 37-56

[3] Anderson, Stephen C. 2001. *Phonological Characteristics of Eastern Grassfields Languages*. In Nguessimo M. Mutaka and Sammy B. Chumbow, ed. Research Mate in African Linguistics: Focus on Cameroon, 33-54.

[4] Anderson, Stephen C. 2008. *A Phonological Sketch of Ngiemboon-Bamileke*. In Nguessimo M. SIL, Cameroon.

[5] Anderson, Stephen C. and Lonfo, Etienne. (2016). Dictionaire ngiemboon – francais – anglais. SIL International ; Cameroon

[6] Bamba Dione, et al (2023) MasakhaPOS: *Part-of-Speech Tagging for Typologically Diverse African languages*. In: Rogers, Anna and Boyd-Graber, Jordan and Okazaki, Naoaki, (eds.) Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. (pp. 10883-10900). Association for Computational Linguistics: Toronto, Canada

[7] Berment, Vincent. 2004. Méthodes pour informatiser des langues et des groupes de langues peu dotées. Ph.D. Thesis, J. Fourier University – Grenoble I.

[8] Bot Ba Njok, H. M. (1970) Nexus et noninaux en Bassaa, these de doctorat d'Etat, Paris.

[9] Chiche, A., Yitagesu, B. Part of speech tagging: a systematic review of deep learning and machine learning approaches. J Big Data **9**, 10 (2022). https://doi.org/10.1186/s40537-022-00561-y

[10] Daniel Jurafsky and James H. Martin. 2024. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models, 3rd edition. Online manuscript released August 20, 2024. https://web.stanford.edu/~jurafsky/slp3.

[11] Horsmann, Tobias; Zesch, Torsten. 2015. *Tokenization and PoS Tagging of Social Media Text*. LTL-UDE @ EmpiriST 2015

[12] Kenneth Steimel. 2018. Part of Speech Tagging in Luyia: A Bantu Macrolanguage. In *Proceedings of the Fifth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2018)*, pages 46–54, Santa Fe, New Mexico, USA. Association for Computational Linguistics

[13] Lonfo, Etienne. 2016. *Grammatical sketch*. Unpublished manuscript

[14] Mutaka, Ngessimo M. and Chumbow, Sammy B. eds. 2021 Research Mate in African Linguistics: Focus on Cameroon. A Fieldworker's Tool for Deciphering the Stories Cameroonian Languages Have to Tell. Rudiger Koppe Verlag.

[15] N Adhvaryu**,** P Balani**;** Adhvaryu, N., Balani, P., (2015), "*Survey: Part-Of-Speech Tagging In NLP*",
International Journal of Research in Advent Technology, Special Issue; 1st International Conference on Advent Trends in Engineering, Science and Technology.

[16] Onyenwe, Ikechukwu E. et al. (2019) *Toward an effective Igbo part-of-speech tagger*. ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP), 18 (4): 42. ISSN 2375-4699.

[17] Parthasarathy, Venkatesh B. et al. 2024. The Ultimate Guide to Fine-Tuning LLMs from Basics to Breakthroughs: An Exhaustive Review of Technologies, Research, Best Practices, Applied Research Challenges and Opportunities. CeADAR: Ireland's Centre for AI, University College Dublin, Belfield, Dublin, Ireland.

[18] Sintayehu Hirpassa, G.S. Lehal. (2023), Improving part-of-speech tagging in Amharic language using deep neural network, Heliyon, Volume 9, Issue 7, 2023, e17175, ISSN 2405-8440,

[19] Sneider, Keith L. 2017. Tone Analysis for Field Linguists. SIL International.

[20] Sonkoue, Eliane Meli Epse Kamdem. 2020b. Tense-aspect categories and standard negation in five Bamileke languages of Cameroon: A descriptive and comparative study. Bayreuth: University of Bayreuth. (Doctoral dissertation).

[21] Wei, Chengwei; Pang Runqi; Kuo Jay 2024. GWPT: *A Green Word-Embedding-based POS Tagger*. arxiv.org/abs/2401.07475

[22] Yemmene, Patrice and Besacier, Laurent. 2019. Motivations, challenges, and perspectives for developing an Automatic Speech Recognition System for the under-resourced Ngiemboon Language. In Proceedings of the First International Workshop on NLP Solutions for Under-Resourced Languages (NSURL 2019), co-located with ICNLSP 2019 - Short Papers, pages 59–67, Trento, Italy. Association for Computational Linguistics.
[23] Author, F., Author, S.: Title of a proceedings paper. In: Editor, F., Editor, S. (eds.) CONFERENCE 2016, LNCS, vol. 9999, pp. 1–13. Springer, Heidelberg (2016)
[24] Author, F., Author, S., Author, T.: Book title. 2nd edn. Publisher, Location (1999)
[25] Author, F.: Contribution title. In: 9th International Proceedings on Proceedings, pp. 1–2. Publisher, Location (2010)
[26] LNCS Homepage, http://www.springer.com/lncs, last accessed 2023/10/25