Motivations, Challenges, and Perspectives for the Development of a Deep Learning based Automatic Speech Recognition System for the Under-resourced Ngiemboon Language

Patrice A. Yemmene

School of Engineering
University of Saint Thomas, MN, USA
yemm2299@stthomas.edu

Laurent Besacier

Laboratoire Informatique de Grenoble
University of Grenoble, France
laurent.besacier@univ-grenoble-alpes.fr

Abstract

Nowadays, a broad range of speech recognition technologies (such as Apple Siri and Amazon Alexa) are developed as the user interface has become ever convenient and prevalent. Machine learning algorithms are yielding better training results to support these developments in Automatic Speech Recognition (ASR). However, most of these developments have been in languages with worldwide, political, economic and/or scientific influence such as English, Japanese, German, French, and Spanish, just to name a few. On the other hand, there has been little or no development of ASR systems (or language technologies) in most minority and under-resourced languages of the world, especially those spoken in Sub-Sahara Africa. One of such languages is the Ngiemboon language which is the focus of this paper. The Ngiemboon language is a Grassfield Bantu language spoken in the West Region of Cameroon (Africa) by about 400,000 people. This paper highlights the motivations, challenges and perspectives inherent in a work in progress (speech data collection is underway) to build a Deep Learning based Automatic Speech Recognition System for this minority under-resourced Cameroonian local language. This paper introduces the issues critical to conducting research in Speech Processing in this language

1. Introduction

Automatic Speech Recognition is “the process and the related technology for converting the speech signal into its corresponding sequence of words or other linguistic entities by means of algorithms implemented in a device, a computer, or computer clusters” (Li and O'Shaughnessy, 2003). As an active field of research, Automatic Speech Recognition has told significant stories for a few decades. “Early attempts to design systems for automatic speech recognition were mostly guided by the theory of acoustic-phonetics, which describes the phonetic elements of speech (the basic sounds of the language) and tries to explain how they are acoustically realized in a spoken utterance” (Juang and Rabiner, 2005). These efforts date back to the early 50s. Since then, ASR has yielded incredible development in a broad range of commercial technologies where Speech Recognition as the user interface has become ever useful and pervasive.

However, most of these developments have been in languages with strong scientific, political, and/or economic influences such as English, German, French, and to some extent Japanese and Spanish, just to name a few. Historically, most of these languages have always enjoyed social prestige and their extensive vocabulary has given them prominence in the world of commerce. It is worth noting that ASR research and innovation in these languages are significant and continuous. On the contrary, there has been little or no research and development efforts in ASR and other Human Language Technologies in most minority languages of the world, particularly those spoken in Sub-Sahara Africa. Yet, these languages serve as the main vector for the socio-economic development of communities where they are spoken. In this paper, we highlight the motivations, challenges, and perspectives that must be considered in building Human Language Technologies, more precisely an Automatic Speech Recognition System for the Ngiemboon language.

1.1 Paper objective and contribution

A surge of interest in the development of technologies in African languages is emerging. The African Languages in the Field: speech Fundamentals and Automation (ALFFA)1 project (spearheaded in France by the “Laboratoire Informatique de Grenoble”

[1] http://alffa.imag.fr/

of the Grenoble Alpes University) is a great example and has been leading significant efforts in the automation of languages spoken in sub-Sahara Africa. Researchers interested in African Languages hope to contribute to the history of Language Technologies innovations as it is being written across the continent. The objective of this paper is to contribute to the research on the development of Language Technologies in African Languages. As a pioneer research project on ASR for the Ngiemboon language, this work will provide a guide for work in Natural Language Processing (NLP) in minority and under-resourced language in Cameroon and other Sub-Sahara African languages.

1.2 The choice of the language for ASR

The authors of this paper both share a very strong interest in the automation of minority and under-resourced African languages. In fact, under different circumstances, each of them carried out research on some of these languages and have become aware of the challenges faced when working on the digitization and automation of minority languages of Africa. One of such challenges is "to bridge the gap between language experts (the speakers themselves) and technology experts (system developers). Indeed, it is often almost impossible to find native speakers with the necessary technical skills to develop ASR systems in their native language" (Besacier et al. 2014). It becomes obvious that a degree of collaboration between native speakers and systems developers is essential to addressing this identified challenge. Fortunately, one of the authors of this paper is a native speaker of the Ngiemboon language and a trained linguist who has contributed to the development of an already published trilingual – French – English – Ngiemboon dictionary. The availability of a native speaker explains the authors' preference in exploring ASR for the Ngiemboon language.

2. Motivations

The rationale for ASR research and development in under-resourced, minority languages spoken in Sub-Sahara Africa such as the Ngiemboon language is grounded in a unique sociolinguistic context, an observation of existing literacy gap, a recognition of advances in technology, a paradigm shift in human rights priorities and scientific discoveries as well as an understanding of the implications of these for economic and community development. In this section, we highlight these motivation factors.

2.1 Sociolinguistic considerations

The nation of Cameroon is home to about 247 local languages, two official languages (French and English) and Pidgin English (Echu, 2004). In their linguistic choices, it is estimated that 73% of Cameroonians use their mother tongue (a local Cameroonian language) instead of a foreign language (English and/or French), despite the peaceful coexistence of these Indo-European languages with Cameroonian languages. This linguistic choice is explained by the fact that Cameroonian local languages are spoken either in the village of their native speakers, their homes, and often used for heritage and cultural identification (Ngefac, 2010). In this diverse linguistic landscape, many industries and fields currently access ASR only in the high-resourced languages of French and English where "presently ASR systems find a wide variety of applications in the following domains; Medical Assistance; Industrial Robotics; Forensic and Law enforcement; Defense & Aviation, Telecommunications Industry; Home Automation and security Access Control; I.T. and Consumer Electronics" (Vajpai and Bora, 2016). As vital as these might be, they are still a luxury for speakers of Cameroonian local languages, including Ngiemboon speakers. Speakers of Ngiemboon, as well as speakers of other Cameroonian languages, prefer the use of their mother tongue in daily communication (Echu, 2004). What if vital ASR applications were developed in the Ngiemboon language as well? It would be an opportunity with great excitement for Ngiemboon speakers' economic, social, and community development.

2.2 Literacy gap

Over two decades ago, analysis of the literacy landscape in Cameroon reported that "four million Cameroonians above fifteen years of age are illiterate. This includes people who never went to school and those who have lapsed back into illiteracy. The Cameroon population is about eleven million people. This is a young population. About 60 percent of Cameroonians are below twenty-five years old. The accuracy of literacy rate estimates is doubtful and could be higher" (Tadadjeu, 2004). It is highly likely that the population of Cameroon has grown significantly since then. A 2018 US Federal Government civilian foreign intelligence service report suggested that about 25 million heads were counted in Cameroon with a 75 % literacy rate. This estimate assumes that about 6 million individuals or more living in Cameroon were illiterate as early as last year. We do not have any reason to believe this has changed much during the last few months. In an illiteracy context such as this one, the use of oral communication is preponderate and convenient. Human-Machine interaction via voice has great potential for economic and community development.

2.3 Economic and community development motivations

In recent years, the mobile telephone industry has experienced a significant boom, in this part of the world where cell phone usage has become very pervasive. For example, "the number of subscribers has risen tremendously; in 2012, there were approximately nine million telephone users in Cameroon, a country with a population of twenty million inhabitants. …. These numbers are certainly below the average" (Moraa, 2012). It is believed that these numbers have changed significantly leading to increased opportunities for Human-Machine interaction.

Because mobile phones only require basic literacy, they are accessible to a large segment of the population regardless of their literacy status. In addition to voice communication, they allow for the transfer of data, which can be used in the context of speech applications for the purposes of health, education, commerce and/or governance. Mobile phones can be used as a mechanism to ensure greater participation of different segments of the population in community development efforts. Innovations along these lines will increase the likelihood of connecting Ngiemboon speakers to vital information that they need to enhance the quality of their lives and contribute to the development of their communities.

2.4 Legal motivations

Another motivation to develop ASR in Ngiemboon language is to allow this linguistic community the opportunity to exercise one of their fundamental rights expressed in article 40 of the Universal Declaration of Linguistic Rights, "In the field of information technology, all language communities are entitled to have at their disposal equipment adapted to their linguistic system and tools and products in their language, so as to derive full advantage from the potential offered by such technologies for self-expression, education, communication, publication, translation and information processing and the dissemination of culture in general". Furthermore, in article 47, "All members of a language community are entitled to have at their disposal, in their own language, all the means necessary for the performance of their professional activities, such as documents and works of reference, instructions, forms, and computer equipment, tools and products". The identified linguistic right aligns with 21st-century human rights priority. In addition to giving this linguistic community the opportunity to exercise one of its fundamental rights, it is also a fascinating endeavor to develop an ASR system for the Ngiemboon language.

2.5 Scientific motivations

The Development of a Speech Recognition system in the Ngiemboon language will play a great role in the revitalization and safeguarding of the language. It will also provide a framework for the digital documentation of the Ngiemboon language. Given the linguistic complexities and peculiarities of the Ngiemboon language (as we will be highlighting in the next section) an ASR research in Ngiemboon may yield discoveries that could add to existing and growing scientific knowledge in this exciting and challenging area of ASR in under-resourced languages.

3. A brief overview of the Ngiemboon language

3.1 Sociolinguistic overview

The Ngiemboon language is part of the Bamileke subgroup of the Eastern Grassfields language family, spoken in the West Region of Cameroon (Anderson, 2008). It has an estimated number of 400,000 speakers. Using EGIDS (Expanded Graded Intergenerational Disruption Scale), a tool used to measure the status of a language in terms of endangerment or development, Ethnologue estimates that the Ngiemboon language is developing. In other words, the language is in vigorous use, with literature in a standardized form being used by some, though this is not yet widespread or sustainable. The language has 5 dialectal variations (Batcham, Balessing, Bangang, Bamougong, Balatchi). From personal observations, lexico-statistic variations among these dialects are very minimal, and mutual intelligibility substantially high.

3.2 Linguistic overview

The Ngiemboon language has very complex linguistic characteristics. "Roots consist of the following C(S)V(C)(V)., ie an obligatory root-initial consonant, and optional semi-vowel, an obligatory vowel, and optional final consonants and vowel" (Anderson, 2008). Anderson (2008) further describes this as "an obligatory root-initial consonant, an optional semivowel, and obligatory vowel, an optional consonant, and an optional final vowel". Its nasal prefixes are syllabic. The language has 16 underlying consonants.

		Labials	Coronals	Velars
Stops:	Voiceless		τ	κ
	Voiced	β	δ	γ
Affricates:		πϕ	τσ	
Fricatives:	Voiceless	ϕ	σ	
	Voiced	ϖ	ζ	
Nasals:		m	n	ñ
Semivowels:			j	w

Ngiemboon underlying consonants proposed by Anderson, 2008

These consonants may exhibit variations based on either their position in the root, the vowel that precedes or follows them, resulting in the following phonetic consonant chart.

Ngiemboon phonetic consonants proposed by Anderson, 2008

		Bi	LD	Den	Alv	Ret	A-P	Vel	Uvl	Glo
Stops:	Voiceless	p		t̪	t	ʈ		k	q	ʔ
	Voiced	b		d̪		ɖ		g		
	Unreleased	p̚			t̚				q̚	ʔ̚
Affricates: Voiceless			pf		ts		tʃ			
	Voiced		bv		dz		dʒ			
Fricatives:	Voiceless		f		s		ʃ	x		
	Voiced	β	v		z		ʒ	ɣ	ʁ	
Nasals:		m	ɱ	n̪	n	ɳ	ɲ	ŋ		
	Unreleased	m̚						ŋ̚		
Liquids:				l̪	ɾ	ɭ				
Semivowels:	Unrounded					j		ɰ		
	Rounded					ɥ		w		

Additionally, the Ngiemboon language has a few underlying vowels, a total of seven identified by Anderson (2008). Most of these vowels can be modified by length and/or nasalization. The following chart exhibits the underlying vowels, long and short oral vowels, as well as short and long nasalized vowels in this grammatically rich tonal Bantu language.

Ngiemboon underlying vowels proposed by Anderson, 2008.

This calls for a very complex vowel system. The chart below highlights an overview of this vowel system:

Underlying Vowels	Short oral	Long oral	Short nasalized	Long nasalized
/i/	i	i:	ĩ	ĩ:
/e/	e	e:	ẽ	ẽ:
/ɛ/	ɛ	ɛ:		
/a/	a	a:	ã	
/ɔ/	ɔ	ɔ:		ɒ̃:
/o/	o	o:	õ	õ:
/u/	u	u:	ũ	ũ:

We should also "recognize the four phonetic semivowels [that enrich the Ngiemboon phonology] as underlying units, even though the parallel four high vowels in Ngiemboon are not all underlying (Anderson, 2008). More details on Ngiemboon phonology can be found in Anderson, 1976a.

The complexity of the Ngiemboon language is extended to its tonal system (Anderson, 2008). In other words, the Ngiemboon language is a tonal language. A tonal language is a language that has "morphemes whose surface pitch (acoustically understood as the fundamental frequency with which corresponds articulatory the rate at which the vocal cords vibrate at any point in time) patterns contrast with each other in one or more comparable environment" (Snider, 2017). This tonal system has four main tone melodies on noun stems. "For example, monosyllabic noun stems with a preceding low-tone prefix display the following stem tones in isolation: Rising, Down stepped High, Low, and Low-falling" (Anderson, 2008). Tone perturbations are considered the most complex part of this tonal system, with tones of individual words changing when these are put into sentences, or tone changing in conjugated verbs.

3.3 An under-resourced language

An under-resourced language can be defined as "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). In addition to providing this definition, a quantitative approach that can be used to determine the level of computation/automation of a language is suggested. He assigns a level of criticality, Cκ, to a service or resource available in the language, a grade Nκ, and an average which he calls index-σ. Berment suggests that the weighted average for less-resourced languages should be between 0 – 9.99, the weighted average for resourced languages between 10 – 13.99, and the weighted average for highly resourced languages between 14 – 20. When these criteria are applied to the Ngiemboon language, the results are as follow:

	Services/Resources	**Criticality Ck (0 to 10)**	**Grade Nk (/20)**	**weighted average (Criticality * grade)**
Word processing	Text entry	0	0	0
	Visualization/ printing	0	0	0
	Find/replace	0	0	0
	Text selection	0	0	0
	Sorting	0	0	0
	Lexical Spell check	0	0	0
	Grammatical spell check	0	0	0
	Style	0	0	0
Speech processing	Voice Synthesis	0	0	0
	Speech Recognition	0	0	0

Translation	Automatic translation	0	0	0
OCR	Optical Character Recognition	0	0	0
Resources	Bilingual Dictionary	8	0	8
	Monolingual Dictionary	0	0	0
Total		**8**		8
Average				**8/8 = 1**

The Ngiemboon language is therefore an under-resourced language, based on this approach.

3.4 A minority language

The definition of a minority language is quite complex. However, for this article, we would consider the definition of the European Charter for Regional or Minority Languages which states that " minority languages mean languages that are:

- traditionally used within a given territory of a State by nationals of that State who form a group numerically smaller than the rest of the State's population; and
- different from the official language(s) of that State”

Though the Ngiemboon language is not a European Language, we believe this definition fits it well, because it is spoken by about 400,000 people in a country that claims about 25 million inhabitants today. According to Ngefac (2010), there are over 247 languages spoken in Cameroon. A review of the Cameroon constitution implies that these languages can be classified into two broad categories: national languages and official languages (English and French). Both national and official languages coexist. The Ngiemboon language is one of the many national languages. We may also refer to it in this paper as a Cameroonian local language.

The understanding that the Ngiemboon language is a minority language can be further justified by the fact that though it has a writing system, it is not certain whether this writing system is complete. Although efforts have been made to develop the language in its written form, it still lacks an in-depth written grammatical description. In addition to this, written literature is very limited and is narrowed to Christian literature. Its presence on the web is very insignificant, and the language is spoken mostly in its geo-linguistic area, where it enjoys a lesser social prestige, compared to the French language spoken in the same geo-linguistic area as well as the whole country and beyond.

4. Challenges in developing an ASR system for the Ngiemboon language

The previous section clearly shows that the Ngiemboon language is a minority low-resourced language. There are challenges related to the development of Automatic Speech Recognition Systems in an under-resourced minority language (Besacier et al., 2014). In this section, we highlight challenges that are specific to the Ngiemboon language.

4.1 Lack of adequate data

Deep Learning algorithms used for training ASR acoustic models in high-resourced languages have been yielding very encouraging results with the decrease of the Word Error Rate (WER), the common metric of the performance of a Speech Recognition Systems. “DNN frameworks, however, typically require a very large amount of data, making them less useful for the low-resource scenario typically encountered with endangered languages” (Imerson et al., 2018). Gauthier (2018) seems to agree with this statement. Citing another resource, she highlights three main resources needed for the development of a state-of-the-art Automatic Speech Recognition system:

- A large text corpus (10 to 100 k words)

A large audio corpus (10 to 100 hours)

- A substantial lexical dictionary with a phonetic transcription of words.

This is useful for both language and acoustic modeling.

To the best of our knowledge, none of these resources exist or are available for the Ngiemboon language. The lack of a significant quantity of speech as well as text and audio data limits access to new machine learning algorithms that enable the development of state-of-the-art Speech Recognition Systems. It is important to note that existing corpora and data collection are integral to ASR development in any language (Besacier et al., 2014).

The development of the corpus by itself may present serious challenges in the case of the Ngiemboon language. Although the amount of data available on the web for many languages (high-resource languages as well as some under-resourced languages) is on the rise, under-resources languages like the Ngiemboon language with no existing corpora and data collection do not experience this growth. Our query on the web returned only a text of about 22k of size, collected as part of the Crubadan project carried out by Scannell (2007). Data collection in this case should anticipate higher costs, a significant amount of time, and the availability of innovative techniques and approaches, assuming there is manpower available.

Notwithstanding significant efforts and progress made in speech data collection even for high-resource languages, “we have barely scratched the surface in sampling the many kinds of speech, environments, and channels that people routinely experience. We currently provide our automatic systems only a very small fraction of the number of materials that humans utilize to acquire language. If we want our systems to be more powerful and to understand the nature of speech itself, we need to make more use of it and label more of it. Well-labeled speech corpora have been the cornerstone on which today’s systems have been developed and evolved. However, most of the large quantities of data are not labeled or poorly labeled, and labeling them accurately is costly” (Huang et al., 2014). This further highlights challenges related to the development of large speech corpora, and the quality of the corpus essential to the development of better ASR systems.

4.2 Language typology

The Ngiemboon language, like many other Bamileke languages of Cameroon, has very complex linguistic characteristics. “One of the most complex aspects of Eastern Grassfields languages is the quantity of tonal perturbations … Even more complex are the many tonal morphemes that affect verb roots in the 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). Serious computational challenges are likely to emerge given the complexity of the tonal and grammatical features inherent in the Ngiemboon language. Although there has been successful computation of supra-segmental features (tone for instance) in some Asian languages (Chen et al., 2018), it is still hard to tell if neural network algorithms that yielded these positive results would produce the same level of satisfaction with the Ngiemboon language, because “the tonal system of the Grassfield Bantu Languages, in particular, is known to be among the most complicated in the world” (Anderson, 1991). Furthermore, this tonal complexity may not have told all the stories that it has to tell, despite significant descriptive linguistic studies that have been carried out so far, and it “will likely still be some time until the exact nature of Eastern Grassfields tonal perturbations is fully understood” (Anderson, 2008). This is to suggest that further and in-depth research is needed in this area to help provide answers to questions. This will only make the automation process of the language challenging.

The tonal system is not the only complex linguistic feature of this language. It has a noun class system, and many significant morphological changes stemming from the syllabic nasal prefixes and/or floating tones in the language. The authors of this paper do not know of any language with such morpho-phonological characteristics where natural language processing or speech recognition research has been carried out. They anticipate that these inherent linguistics characteristics in the language might present computational challenges.

4.3 Sociolinguistic challenges

An Ethnologue report suggests that most Ngiemboonphones speak other languages such as French, Pidgin English, or Ngiemboon neighboring languages. Consequently, many of these speakers are multilingual and frequently use code-switching in their regular conversations. The Online Merriam Webster dictionary defines code-switching as “the switching from the linguistic system of one language or dialect to that of another”. This is generally observed between two bilingual speakers that share the same linguistic code. “In voice communication, many East Africans rapidly code-switch (switch between languages). This is usually done multiple times per sentence, throughout an interaction, and usually between English and another language” (Cvitkovic, 2018). This daily interaction of more than one language poses some challenges to the process of building an adequate Speech Recognition System. In fact, “The development of Automatic Speech Recognition (ASR) for code-switched speech is a current research challenge and is constrained by the difficulty of obtaining representative data for acoustic and language model training” (Ewaldvan and Niesler, 2016). This would be true for the Ngiemboon language as well.

4.4 Economic challenges

Finally, Ngiemboon speakers live mostly in rural areas where they make their living by traditional agriculture. Many also live in various towns and cities where their main activity is small trade, and they are involved in several small-scale commercial activities. In other words, the Ngiemboon language lacks an industrialization status, which presupposes that it may not have a major economic impact/advantage at the moment. This could be a discouraging factor in the development of an ASR system for this language.

5. Developing an Automatic Speech Recognition system in Ngiemboon: Perspectives

Despite the challenges discussed in the previous section, advances in machine learning and other technologies have laid down the path needed to build state-of-the-art Automatic Speech Recognition Systems in under-resourced languages, with a reasonable word error rate. The Ngiemboon language could benefit from some of these advancements. An example of such advances is Deep Learning technologies. Below the authors highlight some insightful technological pathways that will be explored as they develop speech technologies in Ngiemboon. We will not however be discussing in this paper Deep Learning architectures that have been successfully applied to ASR in under-resourced languages. Some of these are mentioned here only as a point of reference. A detailed survey of these is explored at length by Sailor et al., (2018)

5.1 LIG-Aikuma application and data collection

Automatic Speech Recognition systems are built around three pillars: acoustic models, language models, and a pronunciation lexicon. There is a strong correlation between the performance of these, and the amount of training data used. The performance quality if better with more data.

For a very long time in the past, most researchers (linguists, computational linguists, phoneticians …) used microphones and/or recorders for collecting speech data. Today there are new opportunities offering scalable networked devices that make the data collection task less tedious and less costly. A great example is the LIG-Aikuma application, an extension of Aikuma, "a mobile app that is designed to put the key language documentation tasks of recording, respeaking, and translating in the hands of a speech community… It collects recordings, respeakings, and interpretations, and organizes them for later synchronization with the cloud and archival storage… Recordings are stored alongside a wealth of metadata, including language, GPS coordinates, speaker, and offsets on time-aligned translations and comments" (Bird et al., 2014). This application is open source, freely downloadable on any android based smartphone or mobile device. In fact, "the application LIG-AIKUMA has been successfully tested on different devices (including Samsung Galaxy SIII, Google Nexus 6, HTC Desire 820 smartphones and a Galaxy Tab 4 tablet), and can be downloaded from a dedicated website … Originally intended for language documentation and data collection in the field, the app has also been useful for collecting speech for technological development purposes targeting under-resourced languages" (Besacier et al., 2019). In addition to its availability for free download, the LIG-Aikuma application is a great innovation in the area of speech data collection and offers different speech collection modes. It is a great tool that has the potential to support the collection of large quantity of data and may also help with long-term archival of the data collected. Large quantity of data is an absolute prerequisite for Deep Learning Architectures. The LIG-Aikuma application is therefore a great data collection tool available for Deep Learning based ASR pioneering research in Ngiemboon.

5.2 Data augmentation

A team of Google Brain Researchers recently made the following observation: "Deep Learning has been applied successfully to Automatic Speech Recognition (ASR), where the focus of research has been designing better network architectures, for example, DNN (Deep Neural Networks), CNNs (Convolutional Neural Networks), RNNs (Recurrent Neural Network) and end-to-end models. However, these models tend to overfit easily and require large amounts of training data. Fortunately, Data augmentation has been proposed as a method to generate additional training data for ASR" (Park et al. 2019). In other words, these new machine learning algorithms (neural networks) offer new perspectives for the development of state-of-the-art Speech recognition systems and have been applied to under-resourced languages with great success. An example includes the LSTM (Long Short-term Memory) models that exhibit better refinements over standard recurrent neural networks (Gauthier, 2018). The development of language technologies in the Ngiemboon language could certainly benefit from these. Even where and when there might be a relative shortage of data, Data Augmentation may prove to be a useful technology in the process. "Data augmentation is a common strategy adopted to increase the quantity of training data" (Povey et al., 2015). Some data augmentation techniques include "augmenting artificial data for low resource speech recognition tasks, adapting vocal tract length normalization, synthesizing noisy audio via superimposing clean sound with noisy audio signals, applying speed perturbation on raw audio for LVSCR tasks, making use of an acoustic room simulator, and studying data augmentation for keyword spotting" (Dossman, 2019). Additionally, the SpecAugment technique developed by Google Brain will also help in the data collection process (Park, et al. 2019).

6. Conclusion

The Ngiemboon language is a minority, under-resourced language spoken in the Western region of Cameroon (Africa). Although it does not enjoy a strong economic, scientific, or political status, there are compelling reasons to carry out Deep Learning-based ASR research and development in this language. As exciting as it might be, this endeavor would undoubtedly be challenging and will face several hurdles, many of which are highlighted in this paper. However, recent developments in Artificial Intelligence and other new technologies are offering new opportunities and perspectives that were not readily available decades ago. It is feasible and exciting to engage in the development of a state-of-the-art Automatic Speech Recognition system in this language. The benefits are significant, and the stakes are high. The authors of this paper encourage continuous research along these lines. They hope to complete this journey towards developing a speech corpus, and full-blown ASR system in this language and watch for its outcomes.

References

Ngefac, Aloys. 2010. Linguistic Choices in Postcolonial Multilingual Cameroon; Nordic Journal of African Studies 19 (3): 149–164.

Anderson, Stephen C. 1976a. A Phonology of Ngyemboon-Bamileke. Yaoundé, Cameroon: SIL.

Anderson, Stephen. (ed.) 1991. Tone in five languages of Cameroon. (SIL Publication in Linguistics 102). Dallas: Summer Institute of Linguistics and the University of Texas at Arlington.

Anderson, Stephen. 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.

Anderson, Stephen. 2008. A phonological Sketch of Ngiemboon _Bamileke. SIL-Cameroon.

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

Besacier, Laurent; Barnard, Etienne; Karpov, Alexey ; Schultz, Tanja. 2014. Automatic Speech recognition for under-resourced languages: a survey. Speech Communication. Volume 56; 85-100, UK.

Besacier, L., Gautier, E., Voisin, S. 2019. Lessons learned after the development and use of a data collection app for language documentation (ligaikuma). International Congress of Phonetic Sciences ICPhS, Melbourne, Australia.

Bird, Steven; Hanke, Florian; Adams, Oliver; Lee, Haejoong. 2014. Aikuma: A Mobile App for Collaborative Language Documentation. Proceedings of the 2014 Workshop on the Use of Computational Methods in the Study of Endangered Languages, pages 1–5.

Chen Charles ; Bunescu, Razvan; Xu, Li ; Liu, Chang. 2018. Tone Classification in Mandarin Chinese Using Convolutional Neural Networks. 10.21437 Interspeech. 2016-528; Wiley-IEEE Press, Hoboken, NJ.

Cvitkovic, Milan. 2018. Some Requests for Machine Learning Research from the East African Tech Scene. Proceedings of NIPS; Workshop on Machine Learning for the Developing World.

Dossman, Christopher. 2019. Google Brain Unveils a Simple Data Augmentation Method for Speech Recognition. https://medium.com.

Echu, George. 2004. The Language Question in Cameroon. Linguistik online. V.18 28(1):114-133.

Ewaldvan der Westhuizen; Thomas Niesler. 2016. Automatic Speech Recognition of English-isiZulu Code-switched Speech from South African Soap Operas; Procedia Computer Science 81; 121 – 127.

Gauthier, Elodie. 2018. Collecter, Transcrire, Analyser : quand la machine assiste le linguiste dans son travail de terrain. Ph.D Dissertation, Grenoble, France.

Huang, Xuedong; Baker, James; Reddy, Raj. 2014. A Historical Perspective of Speech Recognition. Communications of the ACM. 57. 94-103. 10.1145/2500887.

Imerson, Robbie; Simha, Kruthika; Ptucha, Raymond ; Prud'hommeaux, Emily. 2018. Improving ASR Output for Endangered Language Documentation. 182-186. 10.21437/SLTU.2018-38.

Juan, B. and Rabiner, Lawrence. 2005. Automatic Speech Recognition - A Brief History of the Technology Development.

Li, Deng; O'Shaughnessy, Douglas. 2003. Speech Processing: A Dynamic and Optimization-Oriented Approach. CRC Press (2003)

Moraa, Hilda. 2012. How Mobile technology has been used to create an impact in Cameroon. iHUB Internet blog (https://ihub.co.ke/blogs).

Park, Daniel et al. 2019: SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition. arXiv:1904.08779v2 [eess.AS].

Povey, Daniel et al. 2015. Audio Augmentation for Speech Recognition. Interspeech.

Sailor, Hardik; Patil, Ankur; Patil, Hemant. 2018. Advances in Low Resource ASR: A Deep Learning Perspective. 162-166. 10.21437/SLTU.

Scannell, Kevin. 2007. The Crúbadán Project: Corpus building for under-resourced languages. Building and Exploring Web Corpora. Proceedings of the 3rd Web as Corpus Workshop.

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

Tadadjeu, Maurice. 2004. Language, Literacy, and Education in African Development: A Perspective from Cameroon. SIL-Cameroon.

Vajpai, Jayashri; Bora, Avnish. 2016. Industrial Applications of Automatic Speech Recognition Systems. Int. Journal of Engineering Research and Applications Vol. 6, Issue 3, pp.88-