Mechanisms for Translation of Telephone Calls for Inter-Language Communication and the Development of National Languages and Cultures in Cameroon

The Case of Yemba and Fufuldé

Marcellin Nkenlifack(1), Vivien Beyala Kamgang(2), Deris Meh(3), Germain Tegomo(4), Bethin Demsong(5)

(1) LAIA - FV UIT, Bandjoun, University of Dschang, marcellin.nkenlifack@gmail.com
(2) HTTC, University of Bamenda, vivsxx@gmail.com
(3) DLAL - FA, University of Bamenda, mderis2002@gmail.com
(4) FS, University of Dschang, tegomo2lu@gmail.com
(5) LAIA - FV UIT, Bandjoun, University of Dschang, bethdems@yahoo.fr

Abstract: This paper describes the translation mechanisms of a dialogue/conversation, specifically a phone call between two speakers who do not necessarily share the same linguistic code or the same cultural background using Information and communication technology (ICT). Its major concerns dwell on the integration, processing and translation of local languages, that is the establishment of an abstract model capable of attending to a large number of languages. The approach makes translation easier and favours the realisation of software systems in general, by integrating linguistic and cultural considerations into their development. The schema of the semantic analysis of a message, its methodology, the processing and the translation engine are well catered for. An implementation of one of the modules was successfully carried out in the translation of mini-texts in two Cameroonian languages, Yemba and Fulfulde. The end result aims to contribute to raise awareness on the introduction of teaching local languages and cultures in the official training programs in Cameroon as well as the socio-economic development of the peoples by making results of scientific discoveries available to the grassroots.

Keywords: analogue and digital signals; automated translation; ICT; lexical and semantic analyses; national languages and cultures; telephone call.

Résumé: Cette étude décrit les mécanismes de traduction d'un appel téléphonique entre deux locuteurs qui ne partagent ni le même code linguistique ou contexte culturel en utilisant les technologies de l'information et de la communication (TIC). Les préoccupations majeures portent sur l'intégration, le traitement et la traduction de langues locales, donc la conceptualisation d'un modèle abstrait capable de prendre en charge un grand nombre de langues indigènes. L'approche facilite la traduction ainsi que la réalisation de systèmes logiciels en général, en intégrant des considérations linguistiques et culturelles dans leur développement. Le

schéma d'analyse sémantique d'un message, sa méthodologie, le traitement et le moteur de traduction sont bien pris en compte. Une implémentation d'un des modules a été réalisée avec succès s'agissant de traduire des mini-textes dans deux langues camerounaises, le yemba et le foulfouldé. L'objectif est d'accroitre la sensibilisation à l'enseignement de langues et cultures dans les programmes de formation officiels au Cameroun ainsi qu'au développement socio-économique des personnes en mettant à profit les découvertes scientifiques auprès des populations locales.

Mots-clés : analyses lexicales et sémantiques ; appel téléphonique ; langues et cultures nationales ; signaux analogiques et numériques ; tics ; traduction automatique.

1 Introduction

In the current age of globalisation and digitisation, or better still 'ICTisation' [1], we notice the majority of results of scientific discoveries presented on the Internet are in English. This status quo neither promotes cultural diversity nor encourages the representation of our local languages on this all-important medium of communication. Moreover, it creates a communicative barrier between peoples from various linguistic backgrounds, that is, people not sharing the same linguistic code. This difficulty is very noticeable in major projects in Cameroon where a job applicant has to master at least one of the official languages, English or French, in order to be recruited especially on construction projects generally supervised by entrepreneurs of oriental [2] extraction. In a bid to attend to such situations of communication breakdown, an automated phone call translator appears to be a beneficial outcome to palliate the aforementioned language barrier between peoples at the expense of slowing the development of indigenous peoples.

One can legitimately question the relevance of this quasi-monocultural tool (Bronckart, Bain, Schneuwly & Coline, 1985; Dongmo, 2015) which is the Internet, remembering that cultural and linguistic diversity is one of the 11 principles for building the information society as defined by the World Summit on the Information Society (Stauffacher & Kleinwächter, 2005) on the one hand and that mutual understanding is key for human development. Thus, it becomes understandable that the automated translation of phone calls – verbal communication between peoples from different linguistic backgrounds, is one of the core conditions for achieving access to universal knowledge which is of benefit to mankind.

This brings to the fore, the interest of our project which entails the mobilisation of linguists, computer scientists and other researchers geared towards this vital concern – the automated translation of phone calls in a multilingual and multicultural setting such as Cameroon (Bronckart *et al.*, 1985; Izabela, 2009; Nkenlifack *et al.*, 2011a). For this to happen, it implies the indigenous peoples master their language and culture and are themselves acquainted and/or equipped with modern communication tools in the era of ICTs, so that they can, beyond speech making and policy formulation, work to quickly remedy the situation of this spreading monocultural model.

1. Understood as the era of information and communication technologies.
2. Reference is made here to entrepreneurs of Chinese origin found on major construction projects in Cameroon.

Published in : Ndongo-Keller, Justine ; Ntakirutimana, Évariste ; Thierno Cisse, Mame ; Van Campenhoudt, Marc (dir.), *La traduction et l'interprétation en Afrique subsaharienne : les nouveaux défis d'un espace multilingue. Translation and Interpretation in Sub-Saharan Africa: New Challenges in a Multilingual Space*, Editions des archives contemporaines, ISBN: 9782813003898, doi: 10.17184/eac.9782813003898

1.1 Motivations for the project

The interest in this project stems from the merits of introducing ICTs within the school curricula in Cameroon, the benefits of modernising the teaching of national languages [and cultures] (NLCs) in schools as well as the assisting in an intercultural dialogue given our local context. These are visible at the psychological, educational and social levels.

1.1.1 At the Psychological level

The admission of a people into the concert of nations passes inevitably through the use of its language. Intercultural communication will experience its greatest boom as language barriers which separate men are gradually dissolved.

Individual performances and productivity are not only based on simple technical, financial or organisational mechanisms. They are mostly determined by subtle psychological motives for the 'personal equation' of individuals who design projects and realise them. This personal equation is developed gradually and is in concomitance with the mastery of our languages and endogenous cultural skills.

The mastery of our national languages [and cultures] helps to prevent the sociolinguistic uprooting and acculturation of our peoples, youths in particular, and will give them a greater sense of pride by means of their cultural identity and better mutual esteem *via* the knowledge of their common heritage. This will speed up national [and African] cultural and intellectual integration thus creating a true global village.

1.1.2 At the pedagogic level

The automated translation of phone calls and, by extension, communication between people using different local and/or official languages will ease the learning process of our local languages and the dissemination of scientific knowledge or findings to indigenous peoples. This will stimulate the spirit of innovation and creativity in the user population in these target languages. This is because our local languages have a non-negligible role to play in the spread of scientific and technical knowledge. This automated translation of telephone calls will equally be a move in the direction of 'e-learning'.

1.1.3 At the social level

It is incumbent on us, more than ever, to avoid the uprooting and acculturation of the target populations in order to accelerate national and African integration, and include Cameroonian and African languages [and cultures] within the context of globalisation.

1.2 Objectives

Our objective is to develop converters that will enable two speakers, each communicating in their own [local] language, to communicate and understand each other without the need for a translator or an interpreter. This will reduce the degree of communication entropy thereby concentrating on selective information. It will help in better exploring global cultural and linguistic diversity as well as facilitate the

Published in : Ndongo-Keller, Justine ; Ntakirutimana, Évariste ; Thierno Cisse, Mame ; Van Campenhoudt, Marc (dir.), *La traduction et l'interprétation en Afrique subsaharienne : les nouveaux défis d'un espace multilingue. Translation and Interpretation in Sub-Saharan Africa: New Challenges in a Multilingual Space*, Editions des archives contemporaines, ISBN: 9782813003898, doi: 10.17184/eac.9782813003898

spreading and enriching of global and scientific knowledge. In short, it will quicken the development of the populations of Third World countries.

This project is rooted within the different national, sectorial or thematic developmental strategies of Cameroon. It will greatly contribute to the promotion of equal opportunities for all Cameroonians with new forms of ICT-based teaching and multimedia equipment. The introduction of the teaching of local languages and cultures within the official curriculum is at the core of this project (cf 1996 constitution, 1998 Law of orientation on education, decree creating the Department of Cameroonian Languages and Cultures at the Higher Teachers' Training College of the University of Yaoundé 1 etc.).

When all the converters, 'signals', extractors and analysers will be modelled, the various research teams will focus on the full implementation of the overall scheme of this translator. The targeted population of the project includes pupils, students, lecturers and linguists as well as researchers and falls within the framework of the Leon-Grenoble tele-collaboration project (Tomé, 2009) which favours strategies and didactic practices for a better oral comprehension and production with the use of tasks and Web tools for oral production by learners.

The techniques proposed in the study are applied in consecutive stages with concrete examples to highlight their operationalisation. 'Text 1' (T1) resulting from the encoding converter (from the sounds/audio produced to the signal encoded) should be put in the translators which will produce 'Text 2' (T2). The latter will be decoded (from the resulting signal to the output voice) to be listened by the interlocutor. The latter must have previously chosen the target language of communication. The ensuing sections seek to comment on related studies, the methodology we used and an appreciation of the schema of the translator. We will also look at the implementation and application of this technology towards development, the results obtained so far, challenges faced in undertaking the project before our conclusion and research perspectives.

2 Review of related literature

This section identifies some studies and experiments undertaken in the field of automated phone call translations and later identifies various modules within our project which seeks ultimately to present ways of modernising the teaching of our NLCs. It is worth mentioning that the phone call translation project is a module within the global project undertaken by our research team.

2.1 Geolinguistic situation and development of Yemba and Fulfulde

Kouesso (2016) presents Yemba as a Bantu language, spoken in Cameroon which in 1992 had about 300,000 native speakers essentially in the Menoua Division (West Region) and certain localities of the Lebialem Division in the South-West Region. Its dialects include central Yemba (Dschang, Foreke, Fongo-Tongo, Fotetsa), eastern Yemba (Nkongni, Baleveng), southern Yemba (Fokoue, Fontsa Toula, Fotomena), south-eastern Yemba (Baloum, Bamendo) and western Yemba (Fompap, Fondomera).

Published in : Ndongo-Keller, Justine ; Ntakirutimana, Évariste ; Thierno Cisse, Mame ; Van Campenhoudt, Marc (dir.), *La traduction et l'interprétation en Afrique subsaharienne : les nouveaux défis d'un espace multilingue. Translation and Interpretation in Sub-Saharan Africa: New Challenges in a Multilingual Space*, Editions des archives contemporaines, ISBN: 9782813003898, doi: 10.17184/eac.9782813003898

In terms of linguistic development, the Yemba language presents enough arguments to be used for our experimentation and be taught within a formal learning milieu. Some existing works on the language include: *extensions verbale en Yemba* (Harro, 1989); *esquisse phonologique du Yemba* (Haynes, 1989); *ńpɛtí ála'a zímpɛ* (Handbook for reading and writing Yemba – Harro et *al.,* 1990); *Grammar sketch of Yemba* (Harro & Haynes, 1991); *Letɔŋńtí 1* (Read and write – Tadadjeu *et al.*, 1992); Petit dictionnaire yémba-français (Bird & Tadadjeu, 1997); *Menu Mán Plába* (mishaps of Mán Plába – Kouesso, 2000) [3]. The availability of such literature on the language, which we do not claim to be exhaustive, is quite eloquent in favour of its selection for our study and to ascertain the orthography used.

Fulfulde (West Atlantic language), is home to 17 African countries spanning from Senegal to Sudan, Cameroon inclusive. In Cameroon, it is spoken by the Peuls or Fulani, principally found in the Adamawa, North and Far North regions but we find them too in the North West region of Cameroon due to their nomadic nature. Its trans-border characteristic has made it a vehicular language in these northern regions, spoken in 11 out of 15 divisions [4] thereby marginalising Adamawa-Ubangi languages. Common known names include Adamawa Fulani, Fulbe, Nomadic Fulfulde, Mboroore, Peul among others. Some available documentation on the language include: *A reference grammar for Adamawa Funlani* (Stennes, 1967); *Vocabulaire de base en quatre langues : wandala, foulfoulde, français, anglais* (Nouhou, 1991) ; *English-Fula dictionary (Fulfulde, Pulaar, Fulani): a multidialectal approach* (De Wold, 1995); *Alphabet and orthography statement for Fulfulde, Ajamiya* (Scott, 2007); *Ekkitinol ajamiya : Fulfuld windaande bee karfeje Arabiya* (Nelson, 2007); *Fulfulde Adamawa: a language of Cameroon* (SIL, 2018) [5] In addition to these, we have dozens of dissertations on the language in the domains of phonology, morphology (nominal and verbal), syntax which all pieced together to favour the language choice for this study.

2.2 Some experiments

In many countries and societies, the automated translation of communication, be it verbal or oral, has attracted the attention of researchers who have carried out some studies in the domain.

The Japanese mobile operator NTT Docomo operates a vocal telephonic translation system using smartphones and numerical tables. The system is equipped with the android software Google to translate a conversation from Japanese to French, English, German, Spanish, Indonesian, Italian, Portuguese and Thai. The company acknowledges a few shortcomings of the application but the product is quite functional.

Microsoft and Alcatel-Lucent are working on an application coined 'we talk', and forecast it shall be able to render a conversation in twelve languages. Alcatel intends to develop this project for conferences and the simultaneous management of speakers of four different languages.

3. These were obtained from Kouesso (2016) and www.language-archives.org/language.php/ybb.
4. http://www.language-archives.org/language/fub.,
5. http://search.language-archives,org/search.html?q=Adamawa+Fulfulde..

Published in : Ndongo-Keller, Justine ; Ntakirutimana, Évariste ; Thierno Cisse, Mame ; Van Campenhoudt, Marc (dir.), *La traduction et l'interprétation en Afrique subsaharienne : les nouveaux défis d'un espace multilingue. Translation and Interpretation in Sub-Saharan Africa: New Challenges in a Multilingual Space*, Editions des archives contemporaines, ISBN: 9782813003898, doi: 10.17184/eac.9782813003898

Our study seeks to present a translator which will attempt to overcome these shortcomings observed in existing projects.

2.3 National languages and ICTs: State of the art

This section identifies some studies which have contributed to the current project.

Meh & Nkenlifack (2017: 77) opines that 'language teaching methodologies have evolved to fully accommodate ICTs as integral mediation tools in language teaching/learning'. The study lays 'emphasis on learner autonomy and learning environments especially as regards the poor appropriation of our local languages and cultures within the school milieu.' They demonstrate 'that in a context of learning local languages by secondary school students in Cameroon, an adapted approach wherein the language learning methodology is implemented with a learner management system' (LMS) using a content management system (XAMPP) 'which takes into consideration localisation aspects and the beneficiary population, learner autonomy can be attained when viewed through the prism of cultural historical activity theory.'

Beyala (2016) describes the development of a multilingual processing system obeying the classical analysis-transfer-generation model. It is based on an abstract linguistic model favouring generic representations on a software architecture model based on the notion of object-orientedness such as OpenOffice. The languages considered in the study are Fulfulde and Yemba. The problems addressed relate to the integration, processing and translation of our local languages. The study proposes mechanisms for constructing sentences that emphasise semantic and syntactic analysis, as well as rules for translations and instantiations with supporting examples.

Nkenlifack & Demsong (2016) informs us that there is bound to be communicative breakdown between interlocutors from different backgrounds who do not share the same linguistic code. It becomes imperative then to reflect on a system of automatic translation of calls which could palliate this issue. This will diminish the language barrier between peoples which undermines the development of indigenous peoples. A primary criterion for the assessment of a people's level of development stems from the use of digital technologies in general. This study offers favourable opportunities for a country such as Cameroon especially as regards the promotion of social, human, economic and technological development. This will go a long way to reduce the rate of illiteracy vis-à-vis ICT usage and ensure sustainable development ensured by the mastery of our local languages.

'Nkenlifack *et al.* (2011a) proposes a modernisation of the teaching of NLCs in Cameroon schools with the aid of ICTs in a bid to promote the country's cultural diversity and the diffusion of scientific knowledge in these languages. It proposes a multimedia educational platform, not yet available to the public for general consumption but limits itself to presenting the input of ICTs in developing written competences without showcasing how this platform will enhance oral skills in [these languages].' They admit 'that not every technological device is a pedagogic tool.' (Meh & Nkenlifack, 2017: 78).

Published in : Ndongo-Keller, Justine ; Ntakirutimana, Évariste ; Thierno Cisse, Mame ; Van Campenhoudt, Marc (dir.), *La traduction et l'interprétation en Afrique subsaharienne : les nouveaux défis d'un espace multilingue. Translation and Interpretation in Sub-Saharan Africa: New Challenges in a Multilingual Space*, Editions des archives contemporaines, ISBN: 9782813003898, doi: 10.17184/eac.9782813003898

The foregoing studies are all respective modules which provided input to the current project as we made recourse to computer science, automated translation and didactics of NLCs to model this module for the translation of phone calls between two native speakers of different indigenous languages.

2.4 Knowledge areas involved

To undertake computer aided translation (CAT), a combination of techniques has to be considered in order to come up with acquisition systems, treatment, translation and transfer of documents (text type, voice or multimedia).

2.4.1 Computer Science

These techniques enable to ensure efficient coding, integrated treatment and coherent restitution of data:

- Algorithmic techniques: the algorithms developed will enable us solve specific aspects of the problem with precision;
- Computer modelling: will permit the management of communication flux and treatment for better dissemination of information and data through computer networks;
- Databases: will help in organising and stocking useful or treated data as well as our knowledge base;
- Data warehouse: will permit doing more conclusive reports and data analyses;
- Data mining: will permit the extraction of relevant data so as to better treat them.

2.4.2 Mathematics

The mathematical tools employed for the study are needed for:

- Syntax and semantic analysis of the language, as well as the generation of a transferable code (via computers);
- Mathematical models are required for the grammar of languages (system of recurrent uniform equations, graph of any kind);
- Analysis (NP-completeness, graph algorithms, scheduling techniques, Linear Algebra, Shapes Hermite and Smith calculations polyhedral, linear programming);
- Theory of Automata, Petri nets;
- Random generation of automatic tests;
- Matrix and complex functions.

We will equally need mathematical tools for analysis (Fourrier's analysis) and signal processing (determinist digital signals, random digital signals, and determinist analogue signals). These rely on applications such as:

- Compression of sounds (MP3 encoder) and/or of images (JPEG encoders);

Published in : Ndongo-Keller, Justine ; Ntakirutimana, Évariste ; Thierno Cisse, Mame ; Van Campenhoudt, Marc (dir.), *La traduction et l'interprétation en Afrique subsaharienne : les nouveaux défis d'un espace multilingue. Translation and Interpretation in Sub-Saharan Africa: New Challenges in a Multilingual Space*, Editions des archives contemporaines, ISBN: 9782813003898, doi: 10.17184/eac.9782813003898

— Time frequency analysis, application to language;
— Adaptive sampling

Meanwhile, mathematical models for analysing statistical data will enable us interpret and treat data and signals received in order to make informed decisions.

2.4.3 Electronics and telecommunications

This component will enable us ensure the functioning of component apparatus and to code and transport treated signals. These will permit the exchange of information between actors within our verbal communication system, decision makers and the population (consumers).

2.4.4 Linguistics and language didactics

The automated translation of phone calls requires ground work on descriptive linguistics and language didactics. This will enable the preparation of databases and basic knowledge. These databases will be the result of studies on phonetics, structural phonology, morphology, syntax, lexical and semantic analyses of our local languages. The importance of our local languages is acknowledged in many legal instruments (1996 constitution, the 1998 law of orientation on education, the creation of Linguistics departments in Cameroon universities, the creation of the Department of Cameroon Languages and Cultures at the Higher Teachers' Training Colleges of the Universities of Yaoundé 1 and Ngaoundere). Many studies (Tadadjeu, Sadembou & Mba, 2004; Meh & Kayum, 2010; Tadadjeu, 2015) have been carried out as regards the teaching of our local languages. While attempts to modernise the training of teachers (Ndibnu-Messina & Nya, 2014) and for the teaching of these languages have been proposed (Bilounga, 2015; Biloa & Meh, 2016; Meh & Nkenlifack, 2017), we made use of these results as the foundations of our study to develop a cross-cultural communication model applicable to Cameroon's linguistic situation.

We will mention the programme of rural electronic schools in African Languages – ERELA (écoles rurales électroniques en langues africaines) – a Pan-African programme for education in digital multilingual environments put in place under the National Association of Cameroonian Language Committees (NACALCO) in collaboration with the University of Yaoundé I. The concept has been experimented with success in Cameroon between 2006 and 2012 (Meh, 2019) and presents an extensible model on the use of computers and teaching of ICTs in mother tongues at the primary schools. Emphasis is on the local language as a medium for learning ICTs and availing ICT knowledge to the rural Cameroonian thereby contributing to the reduction of the second digital divide that exists between rural and urban Sub-Saharan Africans. It is an avenue to modernise our local languages because their proper appropriation entails major moral and intellectual development among learners.

2.5 Services offered and tools

Most countries of the world make use of ICTs to translate and spread information in order to animate communication strategies. Most used technologies include the radio, television, Internet, collaborative gateway, discussion fora, mobile technologies, geo-

Published in : Ndongo-Keller, Justine ; Ntakirutimana, Évariste ; Thierno Cisse, Mame ; Van Campenhoudt, Marc (dir.), *La traduction et l'interprétation en Afrique subsaharienne : les nouveaux défis d'un espace multilingue. Translation and Interpretation in Sub-Saharan Africa: New Challenges in a Multilingual Space*, Editions des archives contemporaines, ISBN: 9782813003898, doi: 10.17184/eac.9782813003898

graphic information systems, mails, etc. No tool, be it technical, digital or analogue will function on its own if mankind does not guide and instruct it. Technologies help to find solutions to problems but do not work for themselves. This is because the intelligence of creators, the patience of mediators, the ability of animators and the strength of conviction of communicators are highly required to purposefully use ICTs. To make good use of the numerous chances offered by the Internet, mobile phones, digital (smart) television, cloud systems and tools of social networks, it is necessary to integrate these technological apparatuses to traditional ones 'outline' so as to have access to information, consultation and participation of the public to the described processes.

3 Translation process and methodology

We now attend to the methodology involved during the automated translation process.

3.1 Methodology

At the centre of our processes, we have the TICELaCuN [6] multimedia platform (Nkenlifack *et al.*, 2011). Within this platform several steps and modules are catered for:

- Knowledge acquisition modules (second language acquisition experts);
- Structuring training knowledge into scenarios (instructional design experts);
- Restitution of knowledge interfaces to learners;
- Development and implementation of a library of 'open access' services to ensure encoding – extraction – analysis and decoding of the voice, taking into account parameters of specific lexical, syntactic, semantic, contextual and rhetorical language used;
- Specific keyboards in local languages (Tomé, 2009);
- Text editors in local languages with spellcheck (Ngoumamba, 2015; Soh & Kouesso, 2015);
- Generation of non-UNICODE special characters for local languages;
- Development of control drivers under Linux/Windows (virtual keyboard buffer), code managers;
- Translators and digital dictionaries in local languages;
- Voice recognition (pronunciation modules) as well as tutorial workshops; and
- Mobilisation of teams of analysers, implementation and of control.

3.2 Global architecture and technologies used

We relied on a range of techniques and tools offered by new technologies and already echoed by many authors in the field of 'e-learning' and 'Web services' (Nkenlifack *et al.*, 2011b; Matoussi & Simonneaux, 2009; Nkenlifack *et al.*, 2009; Onoholio, 2009; Tomé, 2009; Meunier & Peraya, 1993; Ruzena, 1979). The main technical components of our global learning platform architecture include:

6. Technologies de l'information et de la communication éducatives pour l'enseignement de langues et cultures nationales.

Published in : Ndongo-Keller, Justine ; Ntakirutimana, Évariste ; Thierno Cisse, Mame ; Van Campenhoudt, Marc (dir.), *La traduction et l'interprétation en Afrique subsaharienne : les nouveaux défis d'un espace multilingue. Translation and Interpretation in Sub-Saharan Africa: New Challenges in a Multilingual Space*, Editions des archives contemporaines, ISBN: 9782813003898, doi: 10.17184/eac.9782813003898

— Web access to available information (theoretical and practical learning materials, diagrams, animations and digital video);

— Collaboration using classical communication tools integrated unto our digital platform;

— At the technical level, our strategy is to implement and deploy applications using free software namely:

— Secure Linux servers (access levels, FireWall, backup);

— Some communication and development protocols like MySQL-PHP, telephone software, LDAP, RTP, SIP, SMTP – IMAP [7];

— Harmonisation of authentication from our LDAP directory to centralise resource inputs and the interoperability and portability of the system capacity because the resource platform hosts media in different formats;

— A file server which provides a secure way of storing directories of actors (tutors, learners, speakers, etc.);

— A voice encoder and decoder for each type of signal.

3.3 Global translation procedure of a message

Interlinguistic telephone communication requires sharing the same code. When the latter cannot be shared, the translation engines must be used. The key steps of the process are cursorily presented.

The translation of a message requires the participation of two persons (H1 and H2) positioned at both ends of a line (the geographical distance between them being immaterial) sharing a linguistic code and having a common concern. If the code is not shared by both speakers, the communication will require a translation. The translator should convert the voice of H1 using the mother tongue as the source language (SL) into an analogue signal (S1). The S1 will have to go through a second converter to yield a second signal (S2) which will be digital. After converting S2, a text (T1) will be obtained. T1 will pass through the translation module. The module will consecutively extract the knowledge embodied in T1 based on textual and extra-textual clues. The lexical analyser (C4) will complete the work of the extractor based on lexical indices used and their arrangement to add meaning to T1. Finally, the semantic analyser (C5) will consider tonal, suprasegmental indices and contextual clues to produce the target language text (T2). Stage 3 will be the conversion of T2 into the target language (TL) previously chosen by the listener.

This process is summarised in Figure 1 below.

7. LDAP: Lightweight directory access protocol, RTP: Real-time transport protocol, SIP: Session initiation protocol, SMTP: Simple mail transfer protocol, IMAP: Internet message access protocol.

Published in : Ndongo-Keller, Justine ; Ntakirutimana, Évariste ; Thierno Cisse, Mame ; Van Campenhoudt, Marc (dir.), *La traduction et l'interprétation en Afrique subsaharienne : les nouveaux défis d'un espace multilingue. Translation and Interpretation in Sub-Saharan Africa: New Challenges in a Multilingual Space*, Editions des archives contemporaines, ISBN: 9782813003898, doi: 10.17184/eac.9782813003898

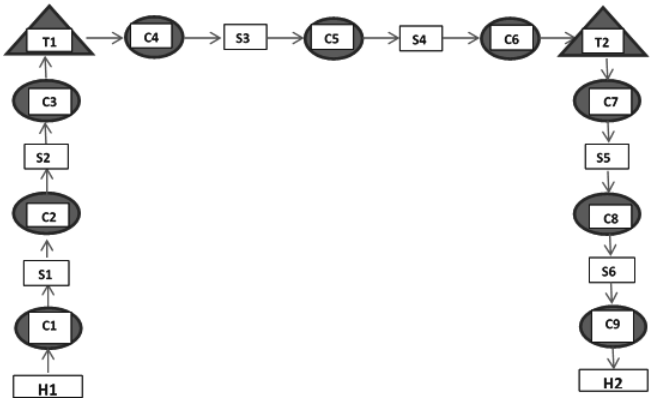


Figure 1 – Schema of our translation process (Nkenlifack & Demsong, 2016:154)

3.4 Semantic analysis of text

Communication is a fundamental element in human relationships. It can be verbal, non-verbal or gestural. The loop of the human verbal communication requires six factors that play complementary functions namely expressive, poetic, conative, metalinguistic, referential and phatic. See Figure 2 below:

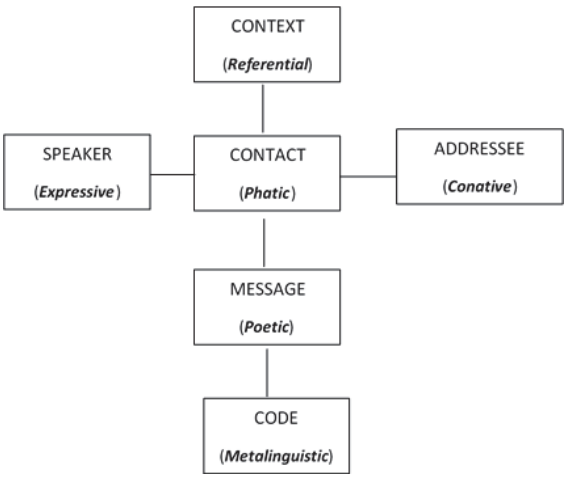


Figure 2 – Schema of human communication showing complementary functions

The interlanguage telephone communication requires prior sharing of the same code. If this is not the case, it will require a translation engine. The translation consists of rendering a text written in an SL to a TL without impeding its meaning. The semantic analysis of T1 will require a number of steps to encode T1 into T2. Figure 3 is an illustration taken from the collaboration among key meta-components and their

Published in : Ndongo-Keller, Justine ; Ntakirutimana, Évariste ; Thierno Cisse, Mame ; Van Campenhoudt, Marc (dir.), *La traduction et l'interprétation en Afrique subsaharienne : les nouveaux défis d'un espace multilingue. Translation and Interpretation in Sub-Saharan Africa: New Challenges in a Multilingual Space*, Editions des archives contemporaines, ISBN: 9782813003898, doi: 10.17184/eac.9782813003898

attributes in a semantic analysis of text. It also takes into account the semantic analysis of a communication text (Izabela, 2009).

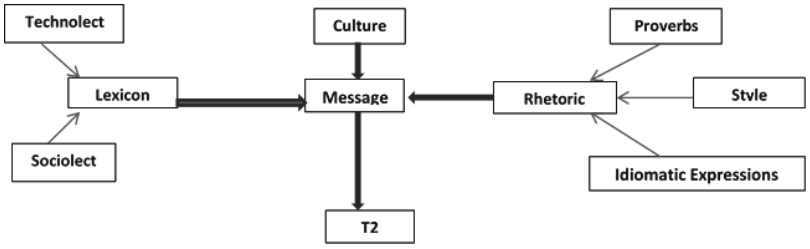


Figure 3 – Collaboration between meta-component analysis

4 Implementation of our module

We now set out to present our approach employed for the conceptualisation of this translation engine.

4.1 The Approach

The project of an automated translation of phone calls will be gradually implemented with some projected achievements namely a comprehensive analysis of the identified TOR (Terms of Reference), modelling of various nodes, encoding the voice of H1, obtaining our T1, extraction of basic information from T1, lexical and syntactic analysis of the T1, semantic analysis of T1, obtaining T2, decoding T2 to obtain the voice H2, development and testing of algorithms, implementation on the operating system of phones, integration unto our digital platform, popularisation and training to use the system.

4.2 Implementation of our translation engine

For an implementation, the linguistic rules were formalised for the test languages (Fulfulde and Yemba). The definition of the constraints and rules of implementation of the translation were based on an approach in conformity with the classic analysis-transfer-generation model. To shed more light on the implementation of the proposed mechanism, we present an extract of one of the components of our general schema which was carried out on an experimental basis, for the step of direct translation of the exchanges by messages *via* a sub-software module implemented for users' phones.

Concretely, the message translation sub-engine is based on a hybrid model, combining both the symbolic analysis approach and the statistical analysis approach to achieve effective translation. The parser (lexical, syntactic, etc.) will be developed with the NLTK (Python) libraries for a recursive and progressive processing of the expressions to be analysed. The functional architecture is described below.

Published in : Ndongo-Keller, Justine ; Ntakirutimana, Évariste ; Thierno Cisse, Mame ; Van Campenhoudt, Marc (dir.), *La traduction et l'interprétation en Afrique subsaharienne : les nouveaux défis d'un espace multilingue. Translation and Interpretation in Sub-Saharan Africa: New Challenges in a Multilingual Space*, Editions des archives contemporaines, ISBN: 9782813003898, doi: 10.17184/eac.9782813003898

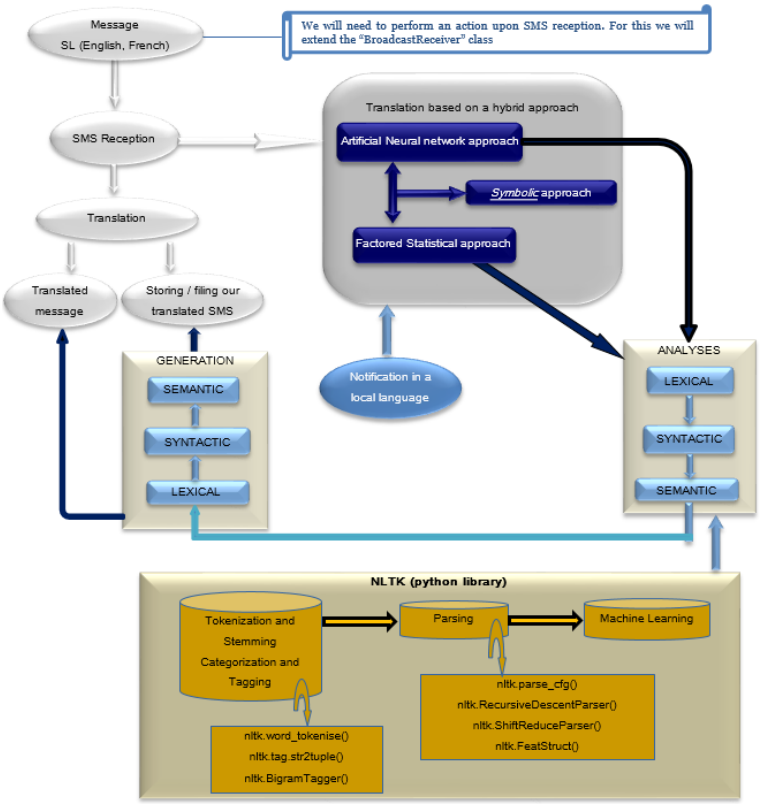


Figure 4 – Functional schema of our translation engine

The functions developed will be compiled as reusable modules. This will facilitate their integration within Android telephone systems. For the reception of messages, the 'BroadcastReceiver' class will be extended in order to generate alerts in the local language chosen by the user. Within the conception, message notifications will be made in a local language. The majority of the solutions either use a classical symbolic analysis approach or a statistical approach. We combined both approaches to avail us a hybrid model for greater efficiency and reliability.

4.3 Exploitation of results and development impact

The coexistence between peoples is ensured by communication. This communication can be intralinguistic, interlinguistic and intersemiotic. Interlinguistic communication requires translation since both the speaker and listener do not share the same linguistic code. Our project is progressing in its implementation (module after module) and some results are already coming through.

An understandable interlinguistic telephone communication offers many advantages that have some bearing directly or indirectly on development namely; sharing of information in real time, assured quality of the information provided, wider dissemination

Published in : Ndongo-Keller, Justine ; Ntakirutimana, Évariste ; Thierno Cisse, Mame ; Van Campenhoudt, Marc (dir.), *La traduction et l'interprétation en Afrique subsaharienne : les nouveaux défis d'un espace multilingue. Translation and Interpretation in Sub-Saharan Africa: New Challenges in a Multilingual Space*, Editions des archives contemporaines, ISBN: 9782813003898, doi: 10.17184/eac.9782813003898

of results of scientific and technological discoveries, expansion of business through some new form of marketing company capable of reaching grassroots' communities, development of contextualised forms of learning (e-learning), concrete development of the people of Third World countries as regards access to ICTs (Osborn, 2010), exploring the cultural values of all peoples, etc.

5 Difficulties encountered or envisaged

Most Sub-Saharan African languages not only have an uneven distribution in terms of the number of words in their corpus but also, often have a limited database of linguistic resources. As such the neural translation engine will not be optimal to overcome these differences that exist between these (local) languages to produce quality translation, although necessary in this process. The call will therefore be made on the statistical approach cascaded to a neural machine translation paradigm, which by their factorised methods uses graphical processing units (GPU) to perform these calculations, and to obtain this GPU is costly.

The different morphological structures of the test languages entail different techniques of translation. Contrary to translation into a less morphologically rich language which does not often require all the morphological information, machine translation into morphologically rich languages is challenging due to their lexical sparsity and large variety of grammatical features with words which depending on the grammatical context can have many different surface forms expressed by these morphemes.

Due to the sparsity of observed word sequences during the translation into morphologically rich languages, the selection of the right surface form often fails when processed by standard statistical machine translation (SMT) models such as phrase-based translation models and n-gram language models (Minkov *et al.*, 2007; Green and DeNero, 2012), keeping in mind that relying only on the word stems hurts the performance of the MT system. The translation of rare words could be improved by using a combination of stemmed and surface word representation in the phrase-based SMT system.

Similarly, although the lexical sparsity to a certain extent is addressed *via* neural language models by projecting word sequences to distributed vector representations, many of the different surface forms possessed by every word stem will be seen rarely, thus still facing the problem of Out-Of-Vocabulary (OOV) or rare words which is worsened in morphologically rich languages (Botha & Blunsom, 2014; Jean *et al.*, 2015; Luong *et al.*, 2015).

To overcome data sparsity in morphologically rich languages, a potential solution is to separate the grammatical aspects of a word, that is the inflectional information from the lexical by using appropriate word representations. Although such word representations exist in the form of morphological lexicons or analysers, no sizeable disambiguated corpus exists for many languages. The necessity of having them for all morphologically rich languages is a great challenge. Due to the issue of ambiguous word analysis, using these resources for statistical language modelling is far from trivial. In order to detect the highlighted error, the grammatical agreement is encouraged

Published in : Ndongo-Keller, Justine ; Ntakirutimana, Évariste ; Thierno Cisse, Mame ; Van Campenhoudt, Marc (dir.), *La traduction et l'interprétation en Afrique subsaharienne : les nouveaux défis d'un espace multilingue. Translation and Interpretation in Sub-Saharan Africa: New Challenges in a Multilingual Space*, Editions des archives contemporaines, ISBN: 9782813003898, doi: 10.17184/eac.9782813003898

by using a morphological analysis that clearly contains useful information. Unfortunately, since each word can have multiple analyses, access to the needed information is difficult.

Due to the fact that large numbers of ungrammatical word sequences are produced by the SMT decoder, an ill-posed problem which introduces spurious ambiguity will thus be the contextual disambiguation and the optimal tag sequence searching. This will require the manual disambiguation of data. The data required to train the disambiguator [8] is costly to produce and is not available in many languages.

6 Research perspectives and conclusion

The use of computers and digital technology in general is becoming the primary criterion for characterising the degree of development of a society. Online education can be fostered with the implementation of translation engines for interlinguistic oral communication, though the role of teachers remains important in this development. The proposed system will be elaborated progressively as per the various sub-modules. At the same time, we will adapt it to other regions and languages of Cameroon in order to further develop and extend the project. The fear of dehumanisation by the computer should not be on the agenda.

In conclusion, this project will ensure sustainable development with the mastery of our local language and cultures as well as the affirmation of our digital identity *via* ICTs. This stems from our conviction that the neo-illiterate, in some years to come, will not only be those who can neither read nor write but those who can neither read nor write in mother tongue and have no basic knowledge of ICTs in today's digital and multilingual era. This is accentuated by the realisation that language learners are entering a world in which communicative competence includes electronic literacies as a crucial component in which communication using registers linked with electronic communication is primordial to achieve success in this information age (Biloa & Meh 2015).

Bibliography

Bengio, Y., Ducharme, R., Vincent, P. & Janvin, C. (2003). 'A neural probabilistic language model'. *Journal of Machine Learning Research.* 3(6), 1137–1155.

Berry, J. W. (2006). 'The World Summit on the Information Society (WSIS): A Global Challenge in the New Millennium'. *Libri.* 56, 11–15.

Biloa, E., & Meh, D. (2015). 'Cameroon languages in computer-assisted language learning: Theoretical considerations for learner autonomy'. In E. Biloa & J. E. Ndibnu-Messina (eds), *Études contrastives, didactique et langues en contact : Enquêtes, pratiques linguistiques et modèles didactiques en Afrique,* Paris : Presses universitaires francophones, 207-214.

Bilounga, M.G.-Y. (2015). *Intégration des TIC dans l'enseignement/apprentissage des langues nationales : le cas du bəti-faŋ dans les classes d'orientation du Lycée Général Leclerc de Yaoundé.* MA dissertation, Yaoundé: University of Yaoundé 1.

Bottou, L. (2012). 'Stochastic gradient descent tricks'. In G. Montavon, G.B. Orr & K.-R. Müller (eds), *Neural networks: Tricks of the trade.* Springer: Berlin, Heidelberg, 421–436.

8. Program that determines the meaning of a term from the context in which it is used, https://en.m.wiktionary.org/wiki/disambiguator.

Published in : Ndongo-Keller, Justine ; Ntakirutimana, Évariste ; Thierno Cisse, Mame ; Van Campenhoudt, Marc (dir.), *La traduction et l'interprétation en Afrique subsaharienne : les nouveaux défis d'un espace multilingue. Translation and Interpretation in Sub-Saharan Africa: New Challenges in a Multilingual Space*, Editions des archives contemporaines, ISBN: 9782813003898, doi: 10.17184/eac.9782813003898

Bronckart, J. P., Bain, D. E., Schneuwly, B., & Coline, D. (1985). *Le fonctionnement des discours : Un modèle psychologique et une méthode d'analyse.* Neuchâtel : Delachaux & Niestlé.

Decree no 18/12/MINESEC/SG/DRH of 27th January 2012 posting the first batch of graduates for the teaching of National Languages and Cultures.

Dongmo K.D.R. (2015). *Mise en place de la téléphonie sur IP avec accès au GSM. Rapport de stage, DUT en génie Informatique.* Bandjoun : Institut Universitaire Fotso Victor.

Duchi, J., Hazan, E., & Singer, Y. (2011). 'Adaptive subgradient methods for online learning and stochastic optimization'. *Journal of Machine Learning and Research*, 12, 2121-2159.

Huang, E. H., Socher, R., Manning, C. D., & Andrew, Y. N. (2012). 'Improving word representations *via* global context and multiple word prototypes'. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics.* Jeju Island, Korea: Association for Computational Linguistics, 1, 873–882.

Izabela, P. (2009). « Analyse sémantique d'un texte de communication ». *Synergies Pologne,* 6, 127–132.

Journal of Machine Learning Research, 3(6), 1137–1155.

Kalchbrenner, N., & Blunsom, P. (2013). 'Recurrent continuous translation models'. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (EMNLP), Seattle, Washington: Association for Computational Linguistics, 1700–1709.

Kamgang, B. V-L. (2016). *An abstract model for localised languages analysis and translation: An application to open-office.* Masters dissertation. Dschang: University of Dschang.

Koum, G., Yekel, A., Tampolla, S., & Sanbong, T (2005). 'Vocal Interaction in Web user interface involving natural language processing'. In A. Akono, E. Tonye, A. Dipanda, Y. Kokou. (eds), *Proceedings of the international conference on Signal & Image Technology and Internet Based Systems* (IEEE SITIS), Yaoundé: Cameroon.

Law N° 08/223/MINESUP/DDES of 3rd September 2008 creating the Department of Cameroon Languages and Cultures at the Higher Teachers' Training College of the University of Yaoundé I.

Law N° 1996/06 of 18th January 1996 revising the June 2, 1972 constitution of Cameroon.

Law N° 98/004 of 14 April 1998 on the orientation on education in Cameroon.

Le H. S., Oparin, I., Allauzen, A., Gauvain, J.-L., & Yvon, F. (2013). 'Structured output layer neural network language models for speech recognition'. *IEEE Transactions on Acoustic, Speech and Signal Processing*, 21(1), 197-206.

Le H. S., Allauzen, A., & Yvon, F. (2012). 'Continuous space translation models with neural networks'. In *Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (NAACL-HLT), Montréal: Association for Computational Linguistics, 39–48.

LeCun, Y., Bottou L., Orr, G., & Muller, K. (1998). « Efficient backprop ». In G. Montavon, G.B. Orr & K.-R. Müller (eds), *Neural networks: Tricks of the trade*, Springer, Berlin: Heidelberg, 9–50.

Mangui, A., & Domche, S. (2008). *Réalisation et mise sur pied d'une plateforme Web de traduction de langues camerounaises.* Mémoire de fin d'études de DUT Informatique de Gestion. Bandjoun : Institut Universitaire Fotso Victor.

Matoussi, F., & Simonneaux, L. (2009). « Didactique et technologies de l'information et de la communication ». *International Journal of Information Sciences for Decision Making – ISDM.* 39, 1–9.

Meh, D. (forthcoming). *An adaptive computer-assisted language learning (CALL)-based model for teaching national languages and cultures in Cameroon secondary schools: A cultural historical activity theory based perspective for learner autonomy.* PhD thesis. Yaoundé: University of Yaoundé 1.

Meh, D., & Kayum, F. C. (2010). 'Towards a revitalization of Cameroon languages *via* formal education: The case study of the Department of Cameroon Languages and Cultures of ENS Yaoundé'. In H. G. Lewis & N. Ostler (eds), *Proceedings of the Foundation for Endangered Languages conference (FEL XIV),* 151-156. Carmarthen, Bath: Foundation for Endangered Languages.

Meh, D., & Nkenlifack, M. (2017). 'Teaching Cameroon Languages and Cultures *via* a Computer-Assisted Language Learning (CALL) Based Perspective for Learner Autonomy: Insights from Activity Theory'. *International Journal of Education, Learning and Development.* 5(9), 77–88.

Meunier, J. P., & Peraya, D. (1993). *Introduction aux théories de la communication.* Bruxelles : De Boeck.

Published in : Ndongo-Keller, Justine ; Ntakirutimana, Évariste ; Thierno Cisse, Mame ; Van Campenhoudt, Marc (dir.), *La traduction et l'interprétation en Afrique subsaharienne : les nouveaux défis d'un espace multilingue. Translation and Interpretation in Sub-Saharan Africa: New Challenges in a Multilingual Space*, Editions des archives contemporaines, ISBN: 9782813003898, doi: 10.17184/eac.9782813003898

Ndibnu-Messina, E. J., & Nya, N. G.-W. (2014). *Usage des tics et développement des compétences professionnelles chez les enseignants en formation initiale au Département d'Informatique et des Technologies Éducatives de l'École Normale Supérieure de l'université de Yaoundé I. www. frantice.net*, 8. [online].

Nkenlifack, M., & Demsong, B. (2016). « Les TIC au service du développement : Traduction automatique des appels téléphoniques inter-linguistiques pour l'essor des langues et cultures nationales ». *Revue Sciences, Technologies et Développement (STD).* 7, 151-156.

Nkenlifack, M., Demsong, B., & Nangue, R. (2011a). « Identité numérique et éducation : enseignement des langues camerounaises grâce aux tics ». *International Journal of Information Sciences for Decision Making – ISDM,* 41, 1-11.

Nkenlifack, M., Nangue, R., Demsong, B., & Fotso, V. K. (2011b). 'ICT for Education: A platform for the modernisation of computer science teaching methods in secondary and high schools in Cameroon'. *International Journal of Advanced Computer Science and Applications*, 2(4), 124-133.

Nkenlifack, M., Nangue, R., Noulamo, T., & Kwonche, A. (2009). « Les TICE au service de la formation ouverte à distance à l'Université de Dschang : Implémentation de SIEL (Système Intégré d'Enseignement en Ligne basé sur Internet) ». *Ethnostylistique et sociolinguistique – revue de communication.* 337-352, ISSN 1560-3407.

Onoholio, S. (2009). « Lobbying ingénieurs polytechniciens : Mobilisation pour l'accès des langues africaines à Internet », *Le Messager* du 7 décembre 2009.

Osborn, D. (2010). African languages in a digital age: Challenges and opportunities for indigenous language computing. Cape Town: HSRC Press.

Ostrá, R. (1979). « L'interprétation sémantique dans la traduction ». *Études Romanes de Brno*, 10, 31–37.

Socher, R., Bauer, J., Manning C. D., & Andrew, Y. N. (2013). 'Parsing with compositional vector grammars'. In *Proceedings of the Annual Meeting on Association for Computational Linguistics* (ACL). Sofia: Bulgaria. 455–465.

Stauffacher, D., & Kleinwächter, W. (eds). (2005). *The world summit on the information society: Moving from the past into the future.* New York: United Nations Information and Communication Technologies Task Force.

Tadadjeu, M. (éd). (2015). *Écoles rurales électronique en langues africaines : Expérimentation au Cameroun et orientation politique panafricaine.* Paris : Harmattan.

Tadadjeu, M., Sadembouo, E., & Mba, G. (2004). *Pédagogie des langues maternelles africaines.* Yaoundé : Les Éditions CLA. Collection PROPELCA n° 144 – 01.

Tomé, M. (2009). « Compétences orales et outils de communication Web dans un projet de télécollaboration pour l'apprentissage du français langue étrangère ». *Journal of Distance Education.* 23(1), 107-126.

Yang, N., Liu S., Li, M., Zhou, M., & Yu, N. (2013). 'Word alignment modeling with context dependent deep neural network'. In *Proceedings of the Annual Meeting on Association for Computational Linguistics* (ACL). Sofia : Bulgaria. 166–175.

Published in : Ndongo-Keller, Justine ; Ntakirutimana, Évariste ; Thierno Cisse, Mame ; Van Campenhoudt, Marc (dir.), *La traduction et l'interprétation en Afrique subsaharienne : les nouveaux défis d'un espace multilingue. Translation and Interpretation in Sub-Saharan Africa : New Challenges in a Multilingual Space*, Editions des archives contemporaines, ISBN : 9782813003898, doi : 10.17184/eac.9782813003898