INTERSPEECH 2016: Show & Tell Contribution
September 8–12, 2016, San Francisco, USA

LIG-AIKUMA: a Mobile App to Collect Parallel Speech for Under-Resourced Language Studies

Elodie Gauthier[1], David Blachon[1], Laurent Besacier[1], Guy-Noel Kouarata[2], Martine Adda-Decker[2], Annie Rialland[2], Gilles Adda[3], Grégoire Bachman[2]

[1]LIG, Univ. Grenoble Alpes, Grenoble, France
[2]LPP, CNRS, Univ Paris 3, France
[3]LIMSI, CNRS, Orsay, France
Elodie.Gauthier@imag.fr

Abstract

This paper reports on our ongoing efforts to collect speech data in under-resourced or endangered languages of Africa. Data collection is carried out using an improved version of the Android application (AIKUMA) developed by Steven Bird and colleagues [1]. Features were added to the app in order to facilitate the collection of parallel speech data in line with the requirements of the French-German ANR/DFG BULB (Breaking the Unwritten Language Barrier) project. The resulting app, called LIG-AIKUMA, runs on various mobile phones and tablets and proposes a range of different speech collection modes (recording, respeaking, translation and elicitation). It was used for field data collections in Congo-Brazzaville resulting in a total of over 80 hours of speech.

Index Terms: speech data collection, mobile app, unwritten languages, fieldwork linguistics, language documentation

1. Introduction

1.1. Context -Bulb project

The BULB (Breaking the Unwritten Language Barrier) project[1] aims at supporting the documentation of unwritten languages with the help of automatic speech recognition and machine translation technologies. The project relies on a strong German-French cooperation involving both linguists and computer scientists. The aim of the project is to design and adapt NLP methods to speed up linguistic analyses of oral, unwritten languages from the Bantu language family. Bantu includes about 500 languages (figures vary from 440 [2] to more than 660 ([3, 4] depending on the author), among those many are exclusively oral and remain unstudied. Towards this aim, we have chosen three languages which are already partly studied and for which a few written resources exist: Basaa (A43), Myene (B11) and Mboshi (C25). In this contribution, we mostly report on our experience in Mboshi data collection using LIG-AIKUMA.

1.2. Live demo scenario

The demo will present LIG-AIKUMA, our mobile app specifically dedicated to fieldwork language documentation. The app works both on Android powered smartphones and tablets. A poster will be also presented to show the recent achievements obtained during the BULB project (updates on the app, recent data collections, etc.).

[1]http://www.bulb-project.org

2. Starting point

2.1. Initial app

The initial smartphone application AIKUMA was developed by Bird et al [1] for the purpose of language documentation with an aim of long-term interpretability. According to the authors, the application is designed for a *future philologist* [1]: it collects enough speech and documentation to allow for a delayed (future) processing by a linguist. Indeed, the authors notice that, in general, language documentation projects lack of resources, especially human resources for the processing of the materials. As a consequence, data may be processed long time after their collection. Moreover, in the case of an endangered language, there is a risk that the study starts after the language has disappeared. This is why the authors extended the concept of respeaking to produce oral translations of the initial recorded material. Theses requirements lead to the development of AIKUMA an Android application which povides the following features: recording of speech, respeaking of audio sources, and oral translation.

2.2. The *respeaking* concept

The concept of *respeaking* was initially introduced by Woodbury [5]. It involves listening to an original recording and repeating what was heard carefully and slowly. This results in a secondary recording that is much easier to transcribe later on (transcription by a linguist or by a machine). The reason is that the initial speech may be too fast, the recording level may be too low, and background noise may degrade the content.

3. From AIKUMA to LIG-AIKUMA

3.1. Motivations and specifications

Some adaptations to the existing application were proposed during the BULB project corresponding to a set of use cases. The first changes to be performed to the application were about user interface that should identify modes associated to each BULB use case. In addition, requirements were made for the application to be quick and easy to use: save and load metadata of the latest recording for saving time of filling them in the form again; give a better feedback on the respeaking once done, etc.

3.2. Recording modes

New developments have focused on the setup of 4 modes, dedicated to specific tasks of speech recording. The following four modes were identified:

- Free recording of spontaneous speech,
- Respeaking a recording (previously recorded with the app or loaded from a wav file): the respeaking allows

Figure 1: Examples of the use of LIG-AIKUMA on Android tablets for data collection: elicited verb conjugations spoken by a native Mboshi woman (left) and free conversations involving several speakers (right).

now to listen (optionally) to the latest recording segment so as to check it and respeak it if needed, before going to the next segment. Also, once the respeaking is finished, a summary view displays the new segments and their corresponding original segments and allows to (optionally) listen to or respeak any of them,

- Translating a recording (previously recorded or loaded): same features as for the respeaking mode except that the source and target languages must be different,
- Elicitate speech from a text file: user loads a text file within the app, then reads the sentence, speaks, listens to the recording for checking and goes to the next sentence, etc. This mode was required for a data collection in Congo-Brazaville during summer 2015.

3.3. Metadata

For every mode, a metadata file is saved with the recording file. Metadata are filled in a form before any recording. Metadata have been enriched with new details (language of the recording, mother tongue of the speaker and other spoken languages) and about the speaker (name, age, gender, region of origin).

3.4. Downloading LIG-AIKUMA

Users who just want to use the app without access to the code can download it directly from the forge direct link[2]. Code is also available on demand. Following the licence of the initial AIKUMA application, the Lig-Aikuma application is provided under the same GNU Affero General Public License[3].

4. First data collections

4.1. Focus on Mboshi

Our objective was to collect large volumes of data from dozens of speakers in different speaking styles. All possible written documents including a 2000-entry Mboshi dictionary [6], traditional tales and biblical texts in Mboshi were gathered. Furthermore, a large part of the 6000 reference sentences for oral language documentation [7] were translated and written in Mboshi by one of the authors (GNK). Figure 1 illustrates the use of LIG-AIKUMA(tablet) to collect Mboshi verb conjugations (left) or more free conversations including several speakers (right).

Practically, two 1-month fieldtrips to Congo Brazzaville were done by one of the authors (GNK) who is a native speaker of Mboshi. The data recording campaign was organised as follows: four *Galaxy Tab 4* tablets were given to four main contact persons who were responsible of the material and who were responsible of recruiting volunteer speakers from their relatives, friends and broader acquaintances. The main contact persons were also asked to contribute to the respeaking process. All speakers were paid and the main contact persons would be given the tablets at the end of the collection process as additional reward. So far, within the BULB project, 48 hours of speech data in Mboshi language were collected with LIG-AIKUMA in Congo-Brazzaville. The corpus is composed of 33 hours of spontaneous speech (mostly debates and stories), 9.5 hours of controlled speech (including conjugations), 2 hours of read sentences (collected using the elicitation mode) and 3.5 hours of read speech. Another collection of 1000 elicitated sentences was performed.

4.2. Re-spoken and translated speech collected

Table 1: *Re-spoken and translated speech collected so far*

Language	Re-spoken	Translated
Mboshi	30h	30h
Myene	soon	soon
Basaa	14h	10h

5. Acknowledgements

ANR-DFG project BULB (ANR-14-CE35-002) and Labex EFL program (ANR-10-LABX-0083)..

6. References

[1] S. Bird, F. R. Hanke, O. Adams, and H. Lee, "Aikuma: A mobile app for collaborative language documentation," *ACL 2014*, p. 1, 2014.

[2] Guthrie, *Comparative Bantu, Volume 2*. Farnborough: Gregg Press, 1971.

[3] M. Mann and D. Dalby, *A thesaurus of African languages*. London: Hans Zell Publishers, 1987.

[4] J. Maho, *A classification of the Bantu languages: an update of Guthrie's referential system*. Routledge, 2003, pp. 639–651.

[5] A. C. Woodbury, *Defining documentary linguistics*. Language Documentation and Description, SOAS, 2003, vol. 1, pp. 35–51.

[6] B. Roch Paulin, R. Chatfield, G. Kouarata, and A. Embengue-Waldschmidt, *Dictionnaire Mbochi-Français*. Congo (Brazzaville): SIL-Congo Publishers, 2000.

[7] T. Bouquiaux and J. Thomas, *Enquête et description des langues à tradition orale*. Paris: SELAF, 1976.

[2] https://forge.imag.fr/frs/download.php/706/MainActivity.apk

[3] https://www.gnu.org/licenses/agpl-3.0.en.html