Using Technology to Bridge the Gap between Speakers, Learners, and Linguists

Elise McClay[1], Erin Olson[1], Carol Little[1], Hisako Noguchi[2], Alan Bale[2], Jessica Coon[1], Gina Cook[3]

[1]McGill University, [2]Concordia University, [3]iLanguage Lab
1086 Ave Docteur Penfield, Montreal, QC H4C 1M7, Canada

[info@migmaq.org]

Abstract

We describe two free, open-source applications developed as language-independent aids for linguists and language revitalization workers of all sorts. The first application, LingSync, is a linguistic database application which facilitates collaboration and encourages openness between linguists and the speakers they work with; the second, Learn [Language] is a language-teaching and language-learning aid geared towards learners of heritage minority languages who may still have speakers in their lives.

Résumé

Nous décrivons deux applications gratuites et open source développées comme auxiliaires indépendants de la langue pour les linguistes et les travailleurs de revitalisation linguistique de toutes sortes. La première application, LingSync, est une application de banque de données linguistique qui facilite la collaboration et encourage l'ouverture entre linguistes et les locuteurs avec lesquels ils travaillent, la seconde, Learn [langue] est une aide à l'enseignement des langues et l'apprentissage des langues orientée vers les apprenants de langues minoritaires qui ont encore des locuteurs dans leur vie.

Background

There is a general tendency for linguistic data collected in the field to stay in databases on university servers, or in notebooks on researchers' bookshelves. Papers published using these data are often only available through academic journals, and written in specialized terminology. The result is that the data are inaccessible to language consultants, teachers, and learners once linguists have collected them. Further, many of the data that linguists work with would be useful to learners; for under-resourced languages, it is crucial to have all materials widely available to learners and teachers alike.

Our project—a collaboration between Concordia and McGill students and researchers, and programming linguists at iLanguage Lab—built two tools to help correct this one-way flow of information. The first, LingSync, is an application which aids collaborative database creation and improves accessibility; the second, Learn [Language], is the prototype for an app which serves as an easy way for students to create their own language lessons, and eventually import material directly from LingSync into a lesson format.

LingSync

While it is not the first linguistic database to cater to teams working at a distance (cf. the Online Linguistic Database), LingSync improves on existing technology in several ways. First, LingSync works both online and offline. Second, the database is version-controlled, making it is impossible to lose information unintentionally. Third, as LingSync is free and works on any device with Google Chrome, it is adaptable to teams with any type of equipment and resources.

LingSync subscribes to best practices in data management, making it easy for community members, consultants, and linguists to create, correct, and otherwise interact with the data. Further, all changes to data are tracked in the database, so there is a record of who made which alterations. LingSync provides multiple levels of data encryption. These levels mean that a given database can be public or private as a whole, but that individual pieces of data can be also encrypted or shared independently of the database's privacy settings. Also, in order to make the data more accessible to more people, public corpora are searchable by search engines such as Google, Bing, and Yahoo.

A final area in which LingSync innovates is its machine learning component. As data are entered, the app "learns" the morphemes and glosses, and streamlines the data entry process by suggesting options as the user works. As a user's analysis of a given morpheme may change, the information that the app learns may also be edited to reflect these changes in analysis.

Learn [Language]

Though there are many existing apps for language-teaching, and for indigenous languages specifically (see Petersen 2013), Learn [Language] is exceptional in combining the following five aspects. First, it is created to be language-independent, allowing a great deal of variation in types of lessons and materials. Second, it focuses on spoken ability in the target language, putting oral communicative skills front and centre. Third, the app and the technology behind it are free and open-source, and will remain so. Fourth, as a language-teaching aide programmed specifically for the needs of learners of minority languages, it provides a framework for keeping the conversation in the target language despite operating in the environment of a different dominant language. Fifth, Learn [Language] allows the users to collaborate with speakers in their lives to generate their own material, tailored specifically to their own learning interests and needs.

Use Cases

LingSync has been used by a variety of organizations, with various goals. Following are three LingSync usage examples, and one proposed use for Learn [Language].

LingSync was used by students at the University of Ottawa for a 2013 field methods class, to record, sort, and share their data and discoveries with each other. The Mi'gmaq Research Partnership based in Montréal and Listuguj, QC, uses LingSync to store, organize, and clean linguistic information learned from consultants. The Myaamia Center in Miami University (Ohio) is developing an Optical Character Recognition system linked to LingSync so that the Gravier Dictionary (a set of 15th-century French Jesuit manuscripts written in Miami-Illinois) can be read into a database format. As a database, this manuscript will become more easily useful for creating Miami-Illinois language resources.

Learn [Language] is a potentially helpful tool for Hinton et al. (2002)'s Master-Apprentice Program (MAP), an oral immersion-based way to learn a heritage language. The MAP is a method for a "master" speaker and an "apprentice" learner; programs based on MAP have had success in the USA and Australia (Linn et al, Florey & Olawsky 2013). Learn [Language]'s build-your-own-lesson feature is designed for apprentices, allowing them to record their sessions with their masters and study it on their own, without monopolizing their teachers' time.

Future work

LingSync is continually growing—having open source code means that any user with programming experience can contribute to improvements. Among the planned modules, we intend to fully integrate the Prosodylab Aligner, a tool allowing users to automatically align audio data to text—this will be helpful to linguists working on segmental analysis, and to language learners who want to pinpoint spellings and pronunciations of words outside the context of a sentence. Another planned module is a file system for narratives and conversations and other longer kinds of speech. This would again be useful to both linguists and learners, as linguists will be able to do more detailed semantic and syntactic analysis, and learners will be able to model their own use of the language off actual speakers.

In the case of Learn [Language], the application is still a prototype; there is a lot of work remaining to be done. Our top priorities are to make it networkable between social media sites and LingSync database, and to implement ways to share lessons between different accounts, respecting the privacy of all involved. A comprehensive evaluation system should also be drawn up, so that students have the option of measuring their successes against that if they so choose. In the long term, we would like to game-ify the learning process, adding points and other incentives to help tap into the fun side of language-learning.

References

Dunham, Joel. *The Online Linguistic Database*. http://www.onlinelinguisticdatabase.org/

Florey, Margaret, and Knut Olawsky. (2013). *Developing a regional Master-Apprentice training network in Australia.* Presentation at 3rd International Conference on Language Documentation and Conservation, Honolulu, HI.

Hinton, Leanne, Matt Vera, and Nancy Steele. (2002). *How to keep your language alive: A commonsense approach to one-on-one language learning*. Berkeley, CA: Heyday Books.

Linn, Mary, Marcellino Berardo, Akira Y. Yamamoto. (2009). Creating language teams in Oklahoma Native American communities. *International Journal of the Sociology of Language*. 132(1), 61–78.

McIvor, Ofelia. (2005). Building the Nests: Indigenous Language Revitalization in Canada Through Early Childhood Immersion Programs. Master's thesis, University of Victoria.

Myaamia Center. *Jesuit Document Translations (webpage)*. Miami University, OH. http://myaamiacenter.org/?page_id=331

Petersen, Rachael. *iDecolonize: Indigenous language-learning mobile apps (round-up)*. http://globalnativenetworks.com/2013/06/18/idecolonize-indigenous-language-learning-mobile-apps/

Rice, Karen. (2009). Must there be two solitudes? Language activists and linguists working together. In J.A. Reyhner and L. Lockard (Eds.), *Indigenous Language Revitalization: Encouragement, Guidance & Lessons Learned* (pp. 37–59). Flagstaff, AZ: Northern Arizona University.

Sarkar, Mela and Mary Ann Metallic. (2009). Indigenizing the Structural Syllabus: The Challenge of Revitalizing Mi'gmaq in Listuguj. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 66(1), 49–71.