LingSync

Dunham, J., Coon, J. and Noguchi, H.
We propose to present LingSync, a suite of open source software tools designed to facilitate collaborative and computer-assisted linguistic fieldwork, language documentation, and linguistic research. LingSync is being developed by linguists, field workers and software developers at Concordia and McGill. It has been used in numerous linguistics field methods courses, it includes features specifically designed to assist with linguistic research, and its architecture and in-development components aim to contribute to efforts in language pedagogy and revitalization. The presentation will consist of three parts: a description of LingSync and its more salient features; example cases where LingSync has been used to support research and teaching initiatives; and discussion on how its use in such initiatives may be extended.

LingSync (www.lingsync.org) is a collection of web services and graphical user interfaces (GUIs) that allow language documenters to build data sets collaboratively. Using LingSync, a language documenter can create a server-stored corpus and configure access settings to it; at this point, the users with access to the corpus can concurrently curate and add to it wherever an Internet connection is available. The software is all open source licensed so that it is freely available and so that it may be freely modified, reused or contributed to. The primary LingSync GUIs are single-page applications that run in modern web browsers, which ensures cross-platform access, an essential prerequisite for a collaborative software application.

LingSync has been used to support the teaching of at least twelve linguistic field methods courses at universities in Canada and elsewhere. Assistant professor Jessica Coon has used LingSync to teach a field methods class on Dharamsala Tibetan at McGill University. The tool gives the instructor and the students access to each other’s data; the benefits of this include accountability to the instructor, opportunities to learn from the fieldwork data of student peers, greater resources for formulating and testing hypotheses, and a more stimulating experience for the language consultant since she is not asked the same questions repeatedly. By using LingSync to collaboratively build corpora on a language that is often endangered and/or under-resourced, the students are helping to ensure that valuable data are not lost once the class ends but are instead potentially available to future researchers and community language activists. In the Dharamsala field methods case, students took turns transcribing and analyzing the consultant’s speech and judgements into the LingSync interface (which was displayed via an LCD projector) during in-class elicitation; this meant that the students could observe and learn from one another and contribute to the collaborative documentation project in a social setting.

LingSync includes a number of features designed to support linguistic research, salient among which are consistency-supporting features, advanced search, text creation features, and features for implementing morpho-phonological models. LingSync allows users to define what consistency means, assists them in attaining it, helps them to search for research-relevant data points in their consistent data sets, and allows them to easily embed their formatted data points in exportable texts for the production of drafts of research papers. While LingSync currently possesses a feature which attempts to identify the morphological analyses during data entry, we are working on a GUI for the finite-state morphological parser creator of Dunham (2014) through which users will be able to create arbitrarily many morphological parsers and use these to a) automate the creation of morphological segmentations and glosses and b) automate the evaluation of the grammatical models that underly

1 LingSync’s source code is version controlled and available on GitHub at https://github.com/OpenSourceFieldlinguistics/FieldDB.
2 Some of these features are currently under development; however, we expect significant advances to have been made by the time of presentation.
the parsers.

LingSync’s high-level architecture crucially involves the use of web services with consistent and standards-adherent application programming interfaces (APIs). This design decision means that the data and functional components (e.g., parsers) of LingSync can be easily reused in other applications such as language learning software or talking dictionaries. We are interested in discussing possible uses of LingSync and its components with a variety of stakeholders in the linguistics-related domains (e.g. language teachers, language revitalization activists) and would be happy with the presentation and/or the demonstration format.