

ICAMPO: UN APLICACIÓN DE COLECCIÓN DE DATOS LINGÜÍSTICOS*

MARYELLEN CATHCART^{a,b}

GINA COOK^b

THERESA DEERING^{b,c}

YULIYA MANYAKINA^{d,e}

GRETCHEN MCCULLOCH^d

HISAKO NOGUCHI^e

^a*University of Delaware* ^b*Language Lab* ^c*Visit Scotland* ^d*McGill University* ^e*Concordia University*

Spanish Abstract - ME

English Abstract

LingSync is an OpenSource database system that allows language researchers to securely enter, store, organize, annotate, and share linguistic data. The application is accessible on any device; as it runs in a HTML5 browser, it runs on laptops (Mac 10.5 and above, Linux, Windows, ChromeBooks) as well as on mobile devices (Android and iPhone/iPad). It is suitable for both online and *offline* use. Furthermore, the application is created with collaborative goals in mind; data is syncable and sharable with other researchers. Researchers can form teams that contribute to a single corpus, where team members use the application to modify and discuss the data. The system also has a simple and friendly user interface, allowing users to drag and drop data (audio, video, text), or record audio/video directly into the database when using the Android App. In addition, the application has import and export capabilities for multiple file types. The application is designed from the ground up to conform to E-MELD and DataOne data management best practices, an important requirement for any database which will house data funded by granting agencies. Most importantly, the application is designed intuitively and theory free, so it is not necessary to be a field linguist or programmer to figure out how it works. The application is hosted on cloud servers so that users can use it without knowing how to set up its servers. The application also has an installation guide for linguistics department server administrators so that they can set up unlimited data usage on their own department servers.

*LingSync: A Free Tool for Creating and Maintaining a Shared Database For Communities, Linguists and Language Learners

We would like to thank everybody who has contributed to this work.

1 Introduction

1.1 Why LingSync was created

1.1.1 The need - GRETCHEN

LingSync was conceived out of the needs of language researchers doing fieldwork or other large scale data collection. Linguistic fieldwork often requires researchers to travel to places where a stable connection to the internet is not guaranteed. Also, it often involves a group of researchers contributing to building a single database. An ideal linguistic database should therefore work both online and offline as well as making it easy to share and integrate data.

1.1.2 Other programs: pros and cons - GRETCHEN

There are existing programs/software used for linguistic fieldwork, however, they fall short in providing features necessary for collaborative fieldwork, while keeping the data confidential if needed. For example, some web-based databases (<http://linguistics.berkeley.edu/karuk/links.php> Karuk Dictionary and Texts , The Washo Project <http://washo.uchicago.edu/dictionary/dictionary.php>) work only online, hence it is impossible for researchers to enter new data or search the database while in the field where the internet is unavailable.

Non-web-based software such as Toolbox (<http://www.sil.org/computing/toolbox/>) and FLEx/FieldWorks (<http://fieldworks.sil.org/flex/>) are excellent in terms of annotating data and organizing data into various formats (corpus, grammar or lexicon). Nonetheless, integration of data taken by individual researchers is not well-automated and takes considerable work by researchers. Moreover, they run only on a single platform (either PC or Linux, but not both). These tools therefore cause further difficulty in data sharing among researchers using different platforms.

General purpose database software such as FileMaker Pro can be customized for the purpose of language research. However, it demands researchers to learn the software, and research teams often need to hire a programmer to customize the software for their research purposes. The stand-alone nature of such software makes data sharing and integration difficult as well.

The existing linguistic database programs, although useful, have various shortfalls that would hinder collection and integration of data. Some of them are constrained by the internet accessibility or by the computer platform types. Some others demand extra human work in order to integrate data. Data entry can require hours of a researchers time. All linguistic database programs surveyed did not provide a good user experience. The number of clicks required and the delay between actions did not meet current software engineering best practices. In addition to core functionalities, a good user experience is necessary to ensure quality data management. The present project grows out of discussion with a number of fieldworkers dissatisfied with currently available options.

1.1.3 Technological background: why we can make this now - GINA

1.2 Design principles - HISAKO

1.2.1 Open (open source, open access)

1.2.2 Standards-compliant (Unicode, EMELD, Leipzig, XML)

1.2.3 User-friendly

2 What is iCampo/LingSync?

2.1 Current functions -GINA

2.1.1 Data entry and import

2.1.2 Auto-glosser

2.1.3 Search

2.1.4 Sharing corpora, activity feed

2.1.5 Custom settings

2.1.6 Export

2.2 Desired future modules -GINA

2.2.1 Wiki-dictionary format export

2.2.2 Language learning module for android

2.2.3 Integration with prosodylab aligner

2.3 Collaborators - GRETCHEN

3 How is iCampo/LingSync used so far?

3.1 McGill-Listuguj partnership

3.2 Field methods classes

3.3 Future users

4 Conclusion - YULIYA

(1) The postman was rung by the door bell.

4.1 Subsection

Lorem ipsum dolor sit amet, lectus faucibus, vitae tortor velit ut dolor integer, nunc sapiente mi, mauris nam magna, gravida et eu quisque. Etiam nibh mi tincidunt, libero euismod sed

consectetuer amet, nam eget donec scelerisque sem sodales, curabitur et ut imperdiet libero leo pulvinar. Eros ipsum consectetur donec imperdiet, dictum aliquet ac sollicitudin dui. Amet magna. Nunc nam. Pellentesque aut vestibulum, neque neque lacus quisque, curabitur bibendum fringilla pede magna non sociis, tempus gravida orci neque dictumst donec, etiam sit. (Postal, 2005)

4.1.1 Subsubsection

Lorem ipsum dolor sit amet, lectus faucibus, vitae tortor velit ut dolor integer, nunc sapiente mi, mauris nam magna, gravida et eu quisque. Etiam nibh mi tincidunt, libero euismod sed consectetur amet, nam eget donec scelerisque sem sodales, curabitur et ut imperdiet libero leo pulvinar. Eros ipsum consectetur donec imperdiet, dictum aliquet ac sollicitudin dui. Amet magna. Nunc nam. Pellentesque aut vestibulum, neque neque lacus quisque, curabitur bibendum fringilla pede magna non sociis, tempus gravida orci neque dictumst donec, etiam sit.

5 Introduction

Lorem ipsum dolor sit amet, lectus faucibus, vitae tortor velit ut dolor integer, nunc sapiente mi, mauris nam magna, gravida et eu quisque. Etiam nibh mi tincidunt, libero euismod sed consectetur amet, nam eget donec scelerisque sem sodales, curabitur et ut imperdiet libero leo pulvinar. Eros ipsum consectetur donec imperdiet, dictum aliquet ac sollicitudin dui. Amet magna. Nunc nam. Pellentesque aut vestibulum, neque neque lacus quisque, curabitur bibendum fringilla pede magna non sociis, tempus gravida orci neque dictumst donec, etiam sit.

- (2) The postman was rung by the door bell.

5.1 Subsection

Lorem ipsum dolor sit amet, lectus faucibus, vitae tortor velit ut dolor integer, nunc sapiente mi, mauris nam magna, gravida et eu quisque. Etiam nibh mi tincidunt, libero euismod sed consectetur amet, nam eget donec scelerisque sem sodales, curabitur et ut imperdiet libero leo pulvinar. Eros ipsum consectetur donec imperdiet, dictum aliquet ac sollicitudin dui. Amet magna. Nunc nam. Pellentesque aut vestibulum, neque neque lacus quisque, curabitur bibendum fringilla pede magna non sociis, tempus gravida orci neque dictumst donec, etiam sit. (Postal, 2005)

5.1.1 Subsubsection

Lorem ipsum dolor sit amet, lectus faucibus, vitae tortor velit ut dolor integer, nunc sapiente mi, mauris nam magna, gravida et eu quisque. Etiam nibh mi tincidunt, libero euismod sed consectetur amet, nam eget donec scelerisque sem sodales, curabitur et ut imperdiet libero leo pulvinar. Eros ipsum consectetur donec imperdiet, dictum aliquet ac sollicitudin dui. Amet magna. Nunc nam. Pellentesque aut vestibulum, neque neque lacus quisque, curabitur bibendum fringilla pede magna non sociis, tempus gravida orci neque dictumst donec, etiam sit.

6 Introduction

Lorem ipsum dolor sit amet, lectus faucibus, vitae tortor velit ut dolor integer, nunc sapiente mi, mauris nam magna, gravida et eu quisque. Etiam nibh mi tincidunt, libero euismod sed consectetur amet, nam eget donec scelerisque sem sodales, curabitur et ut imperdiet libero leo pulvinar. Eros ipsum consectetur donec imperdiet, dictum aliquet ac sollicitudin dui. Amet magna. Nunc nam. Pellentesque aut vestibulum, neque neque lacus quisque, curabitur bibendum fringilla pede magna non sociis, tempus gravida orci neque dictumst donec, etiam sit.

(3) The postman was rung by the door bell.

6.1 Subsection

Lorem ipsum dolor sit amet, lectus faucibus, vitae tortor velit ut dolor integer, nunc sapiente mi, mauris nam magna, gravida et eu quisque. Etiam nibh mi tincidunt, libero euismod sed consectetur amet, nam eget donec scelerisque sem sodales, curabitur et ut imperdiet libero leo pulvinar. Eros ipsum consectetur donec imperdiet, dictum aliquet ac sollicitudin dui. Amet magna. Nunc nam. Pellentesque aut vestibulum, neque neque lacus quisque, curabitur bibendum fringilla pede magna non sociis, tempus gravida orci neque dictumst donec, etiam sit. (Postal, 2005)

6.1.1 Subsubsection

Lorem ipsum dolor sit amet, lectus faucibus, vitae tortor velit ut dolor integer, nunc sapiente mi, mauris nam magna, gravida et eu quisque. Etiam nibh mi tincidunt, libero euismod sed consectetur amet, nam eget donec scelerisque sem sodales, curabitur et ut imperdiet libero leo pulvinar. Eros ipsum consectetur donec imperdiet, dictum aliquet ac sollicitudin dui. Amet magna. Nunc nam. Pellentesque aut vestibulum, neque neque lacus quisque, curabitur bibendum fringilla pede magna non sociis, tempus gravida orci neque dictumst donec, etiam sit.

7 Introduction

Lorem ipsum dolor sit amet, lectus faucibus, vitae tortor velit ut dolor integer, nunc sapiente mi, mauris nam magna, gravida et eu quisque. Etiam nibh mi tincidunt, libero euismod sed consectetur amet, nam eget donec scelerisque sem sodales, curabitur et ut imperdiet libero leo pulvinar. Eros ipsum consectetur donec imperdiet, dictum aliquet ac sollicitudin dui. Amet magna. Nunc nam. Pellentesque aut vestibulum, neque neque lacus quisque, curabitur bibendum fringilla pede magna non sociis, tempus gravida orci neque dictumst donec, etiam sit.

(4) The postman was rung by the door bell.

7.1 Subsection

Lorem ipsum dolor sit amet, lectus faucibus, vitae tortor velit ut dolor integer, nunc sapiente mi, mauris nam magna, gravida et eu quisque. Etiam nibh mi tincidunt, libero euismod sed consectetur amet, nam eget donec scelerisque sem sodales, curabitur et ut imperdiet libero leo

pulvinar. Eros ipsum consectetur donec imperdiet, dictum aliquet ac sollicitudin dui. Amet magna. Nunc nam. Pellentesque aut vestibulum, neque neque lacus quisque, curabitur bibendum fringilla pede magna non sociis, tempus gravida orci neque dictumst donec, etiam sit. (Postal, 2005)

7.1.1 Subsubsection

Lorem ipsum dolor sit amet, lectus faucibus, vitae tortor velit ut dolor integer, nunc sapiente mi, mauris nam magna, gravida et eu quisque. Etiam nibh mi tincidunt, libero euismod sed consectetur amet, nam eget donec scelerisque sem sodales, curabitur et ut imperdiet libero leo pulvinar. Eros ipsum consectetur donec imperdiet, dictum aliquet ac sollicitudin dui. Amet magna. Nunc nam. Pellentesque aut vestibulum, neque neque lacus quisque, curabitur bibendum fringilla pede magna non sociis, tempus gravida orci neque dictumst donec, etiam sit.

8 Introduction

Lorem ipsum dolor sit amet, lectus faucibus, vitae tortor velit ut dolor integer, nunc sapiente mi, mauris nam magna, gravida et eu quisque. Etiam nibh mi tincidunt, libero euismod sed consectetur amet, nam eget donec scelerisque sem sodales, curabitur et ut imperdiet libero leo pulvinar. Eros ipsum consectetur donec imperdiet, dictum aliquet ac sollicitudin dui. Amet magna. Nunc nam. Pellentesque aut vestibulum, neque neque lacus quisque, curabitur bibendum fringilla pede magna non sociis, tempus gravida orci neque dictumst donec, etiam sit.

(5) The postman was rung by the door bell.

8.1 Subsection

Lorem ipsum dolor sit amet, lectus faucibus, vitae tortor velit ut dolor integer, nunc sapiente mi, mauris nam magna, gravida et eu quisque. Etiam nibh mi tincidunt, libero euismod sed consectetur amet, nam eget donec scelerisque sem sodales, curabitur et ut imperdiet libero leo pulvinar. Eros ipsum consectetur donec imperdiet, dictum aliquet ac sollicitudin dui. Amet magna. Nunc nam. Pellentesque aut vestibulum, neque neque lacus quisque, curabitur bibendum fringilla pede magna non sociis, tempus gravida orci neque dictumst donec, etiam sit. (Postal, 2005)

8.1.1 Subsubsection

Lorem ipsum dolor sit amet, lectus faucibus, vitae tortor velit ut dolor integer, nunc sapiente mi, mauris nam magna, gravida et eu quisque. Etiam nibh mi tincidunt, libero euismod sed consectetur amet, nam eget donec scelerisque sem sodales, curabitur et ut imperdiet libero leo pulvinar. Eros ipsum consectetur donec imperdiet, dictum aliquet ac sollicitudin dui. Amet magna. Nunc nam. Pellentesque aut vestibulum, neque neque lacus quisque, curabitur bibendum fringilla pede magna non sociis, tempus gravida orci neque dictumst donec, etiam sit.

References

Postal, Paul. 2005. *Contrasting extraction types*. Cambridge, MA: MIT Press.