

Technology Session
November 1, 2012 @ ZAS, Berlin, Germany

Building a database for syntax- phonology interface research: using FileMaker Pro to organize data

Emily Elfner (McGill University)
Seunghun J. Lee (ZAS & Central Connecticut State University)

Goals

- What role can databases play in linguistic research?
- More specifically, what can databases (e.g. FileMaker Pro) do for syntax-phonology interface research?
- To generate discussions about using databases in research (NOT a tutorial of how to use FileMaker Pro)

Database for Syntax-Phonology Interface Research

2

Background

- NSF Grant "The effects of syntactic constituency on the phonology and phonetics of tone" (BCS-1147083) to Elisabeth Selkirk (PI), Gorka Elordieta, Seunghun Lee (Co-PIs)
- Four typologically distinct languages
 - Basque (Gorka Elordieta)
 - Irish (Emily Elfner)
 - Luganda (Elisabeth Selkirk & Yelena Fainleib)
 - Xitsonga (Seunghun Lee & Elisabeth Selkirk)

Database for Syntax-Phonology Interface Research

3

Background

- Empirical goals
 - Fieldwork on syntax-phonology interface in the 4 languages
 - Collecting and archiving comparable materials from the 4 languages
- Theoretical goals
 - To compare predictions of Match Theory (Selkirk 2011) with previous theories of syntax-phonology interface

Database for Syntax-Phonology Interface Research 4

Why do we need a database for interface research?

- To organize and keep track of **linguistic data** such as sentence types and phonological properties as we collect it.
- To keep track of various kinds of **metadata** such as people, recordings, etc. across the 4 languages.
- To keep track of **hypotheses** and **evolving analyses**.
- To keep track of **archiving** of recordings and **dissemination** of research findings.

Database for Syntax-Phonology Interface Research 5

Examples of linguistic data - specific to our goals

- Sentence types (i.e. different syntactic structures)
- Phonological properties of sentences
 - Number of words in a sentence & in a phrase
 - branching, non-branching, binarity...
 - Choice and combination of words
- Properties of words
 - Type and number of syllables
 - Lexical tone, stress and accent
 - Segmental content (voiced, sonorants...)

Database for Syntax-Phonology Interface Research 6

How can we manage this data?

- Flat databases vs. Relational databases
- Flat databases
 - Excel spreadsheets
 - Word documents
- Relational databases
 - A database that keeps track of student registration
 - A database that keeps track of customer orders for small businesses

Database for Syntax-Phonology Interface Research

7

Flat databases - An example using Excel

- Create multiple spreadsheets
 - One sheet for sentence types/structures
 - One sheet for phonological properties of words
 - One sheet for information about recordings
 - One sheet for people metadata
 - Etc. etc.
- Challenge
 - Using these sheets, it would not be a simple task to find the following set of data
 "All sentences that have a 3 syllable word in the head of the subject DP for a speaker X of the language Y"

Database for Syntax-Phonology Interface Research

8

Relational databases - Why is it worth looking into them?

- Using a relational database program such as FileMaker Pro has advantages over Excel because:
 - Entering data is easy and automatic (fewer errors and repetition)
 - It is possible to embed various file types and view them easily: audio, video, pdf, jpeg and others
 - It is straightforward to search for complex queries such as "All sentences that have a 3 syllable word in the head of the subject DP for a speaker X of the language Y"
 - Results of complex queries are displayed in a way useful for analysis

Database for Syntax-Phonology Interface Research

9

Filemaker Pro 12

- Disadvantages
 - Cost (occasional need for upgrades)
 - Proprietary software (not open source)
 - May need training to use all the features
- Advantages
 - Exportable into XML and other types of files
 - Relatively intuitive to use (does not involve knowledge of programming languages)
 - Possibility of web-publishing

Database for Syntax-Phonology Interface Research 10

Filemaker Pro 12 DEMO

Database for Syntax-Phonology Interface Research 11

Designing a database for syntax-phonology interface
An Entity Relationship Diagram (ERD)

Database for Syntax-Phonology Interface Research 12

Where do we start?

- What do we want to search for?
 - Sentence types (i.e. different syntactic structures)
 - Phonological properties of sentences
 - Whether phrases are branching, non-branching, binary...
 - Choice and combination of words (especially in tone languages)
 - Properties of words
 - Number and type of syllables
 - Lexical tone, stress and accent
 - Segmental content (voiced, sonorants...)

Database for Syntax-Phonology Interface Research

13

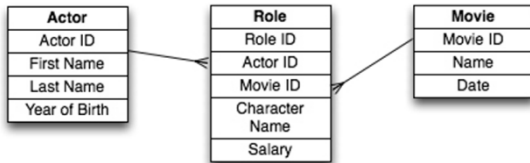
Also, we may want ...

- To restrict (narrow down) the search in some way:
 - *A specific speaker producing sentences with a binary branching object DP vs. all speakers*
- To combine search criteria in some way:
 - *All the recordings of a sentence with a non-branching object DP in all 4 languages.*

Database for Syntax-Phonology Interface Research

14

A example of a part of an ERD *many-to-many relationship*

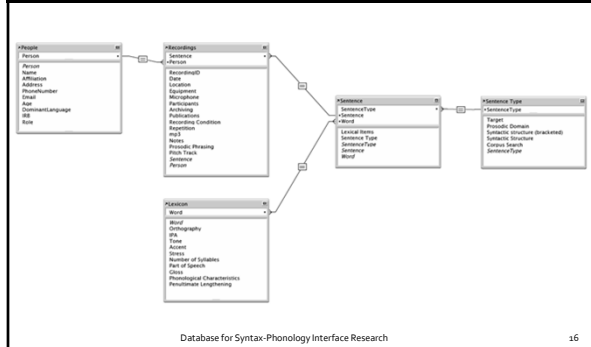


- **Entity:** *type or class of thing that has a distinct existence — something that one wishes to track in a consistent fashion*
- **Attribute:** *a characteristic of an entity, rather than an entity in its own right*

Database for Syntax-Phonology Interface Research

15

A sample of an ERD for syntax-phonology interface research



Discussion

- What other attributes would we need to build an effective database for syntax-phonology interface research?
- Any other comments or questions?

Acknowledgements

- Special thanks to:
 - Volkswagen Foundation
 - Andrew W. Mellon Foundation
 - NSF grant #BCS-1147083 "The Effects of Syntactic Constituency on the Phonology and Phonetics of Tone"
