# The SLAY Database:
# A Meta-Analytic Database of Sign Languages

Rachael Tatman
The University of Washington

# Outline

1) Overview
2) Motivations
3) Organization
4) Data and Input Analysis
5) Distribution
6) Future Work

# Outline

1) **Overview**
2) Motivations
3) Organization
4) Data and Input Analysis
5) Distribution
6) Future Work

# Overview

- SLAY:
  - Sign Language AnalYses
  - Database of grammatical information on signed languages designed to answer modality-specific questions
- Sign languages:
  - Visually-encoded languages used primarily by deaf populations
  - Almost 140 extant sign languages

# Outline

1) Overview
2) **Motivations**
   1) Database as a whole
   2) Parameters
3) Organization
4) Data and Input Analysis
5) Distribution
6) Future Work

# Motivations for SLAY

- There are databases of grammatical structures for spoken language
  - The World Atlas of Linguistic Structure (Dryer & Haspelmath 2013)
    - Zeshen (2013) looked non modality-specific questions
  - The Phonetics Information Base and Lexicon (Moran 2012).
- And corpora/databases of signed languages
  - DGS Corpus (Hanke et. al 2010)
  - Corpus NGT (Crasborn & Zwitserlood 2008)
- But (previously) no databases of grammatical structures only of signed languages

# Parameters

- Motivated by a specific desire to know what parameters (sub-lexical units of sign) are used by signed languages
- Proposed parameters (examples on next side) include:
  - Movement, handshape and location (Stokoe 1960)
  - Number of hands (Bellugi & Fischer 1972)
  - Palm orientation (Friedman 1975)
  - Non-manual component (facial expression, body position, etc.)  (Lidell 1978)
  - Added during second pass: Contact (Klima & Bellugi 1979)

# Parameters (2)

- Examples (all from American Sign Language):
  - Movement: SIT vs CHAIR
  - Handshape: I vs MY
  - Location: SUMMER vs UGLY
  - Number of hands: LIKE vs INTERESTING
  - Palm orientation: MAYBE vs BALANCE
  - Non-manual: LATE vs NOT-YET
  - Contact:  no minimal pairs in ASL (minor parameter)

# Parameters (3)

- To answer this question, the database needs to be hold information about:
  - Individual signed languages
  - Individual grammars/sources
  - The parameters proposed by each source for each language
- In addition, there needs to be room for growth/ways to investigate other questions

# Outline

1) Overview
2) Motivations
3) **Organization**
   1) Design considerations
   2) Current structure (Schema)
   3) Current content
4) Data and Input Analysis
5) Distribution
6) Future Work

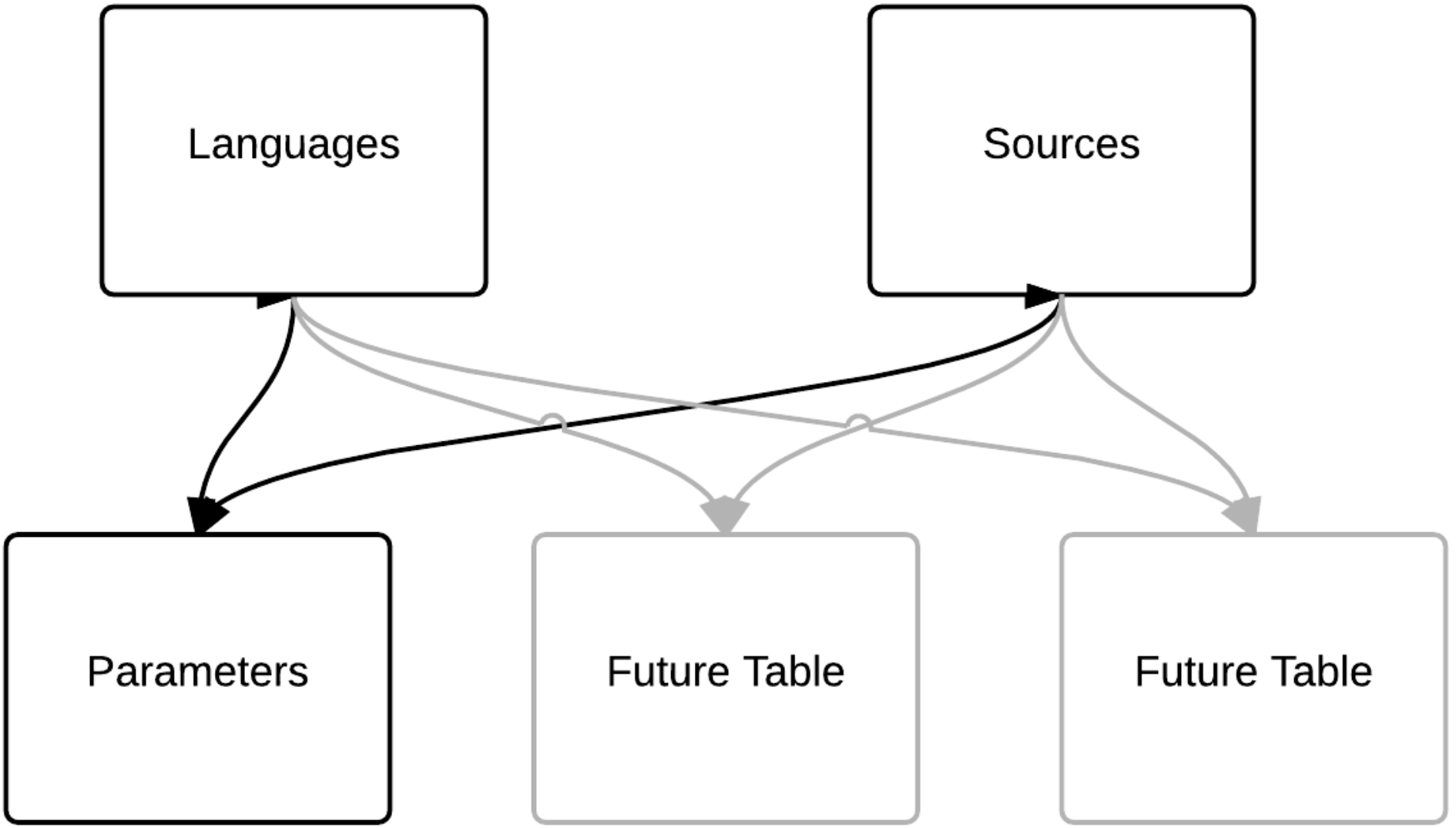
# Design Considerations

- Problem: Need to have three types of information: language, source and information from source about language
  - Solution: Relational database with separate tables for each type of information tied together by foreign keys
  - Bonuses:
    - Modular structure paves the way for future growth
    - Smaller tables make for better data normalization

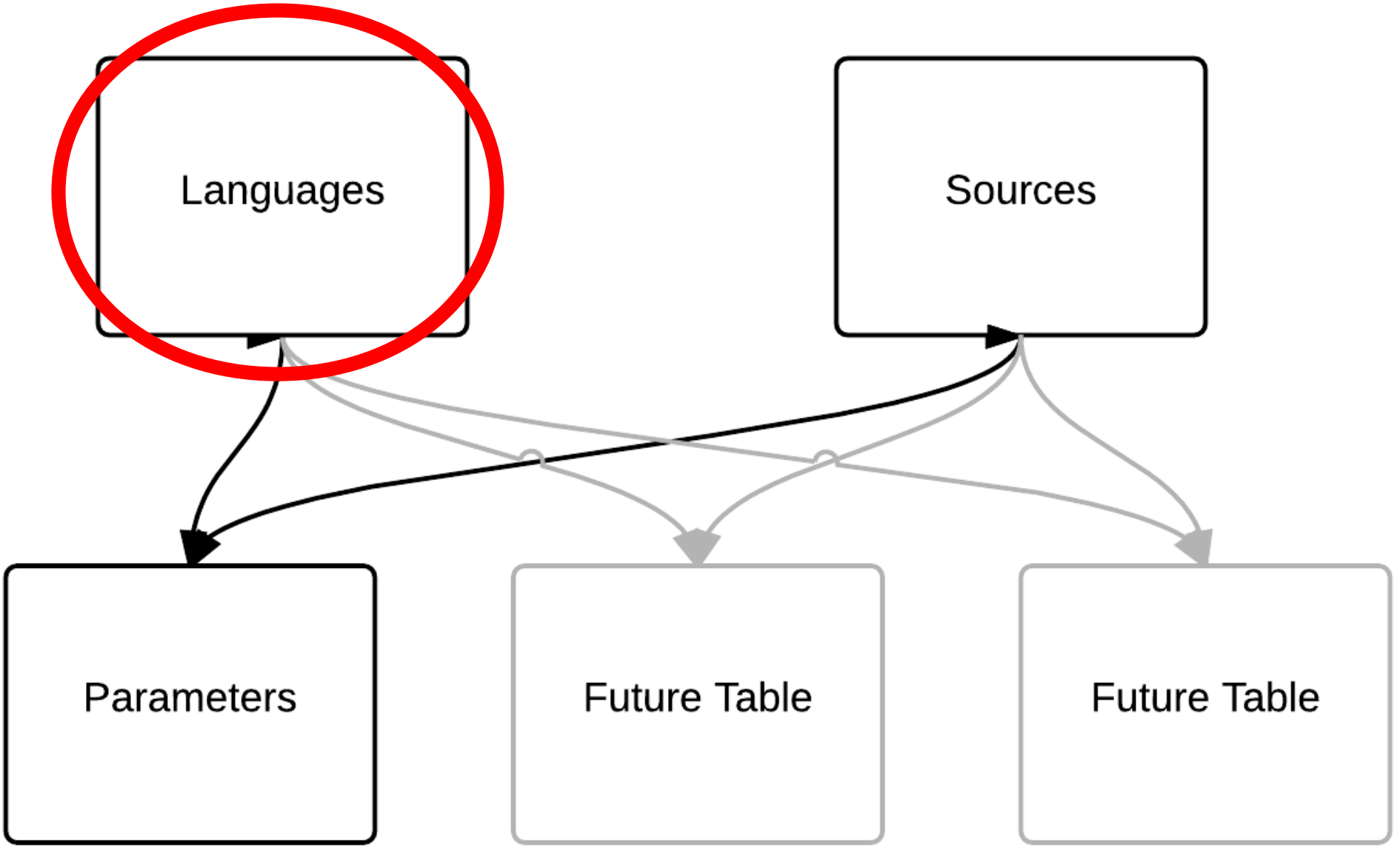
# Design Considerations (2)

- Problem: Need to allow for future growth
  - Solutions: Modular structure and automatically iterated numeric keys (i.e. not using ISO codes for languages)
- Problem: Need an inexpensive cross-platform way to work on/add to the database
  - Solution: Use MySQL Community Edition (Dubois 2005) and Python scripts for data-entry: both free and open source

# Current Structure



# Current Structure



# Current Structure:
## Languages

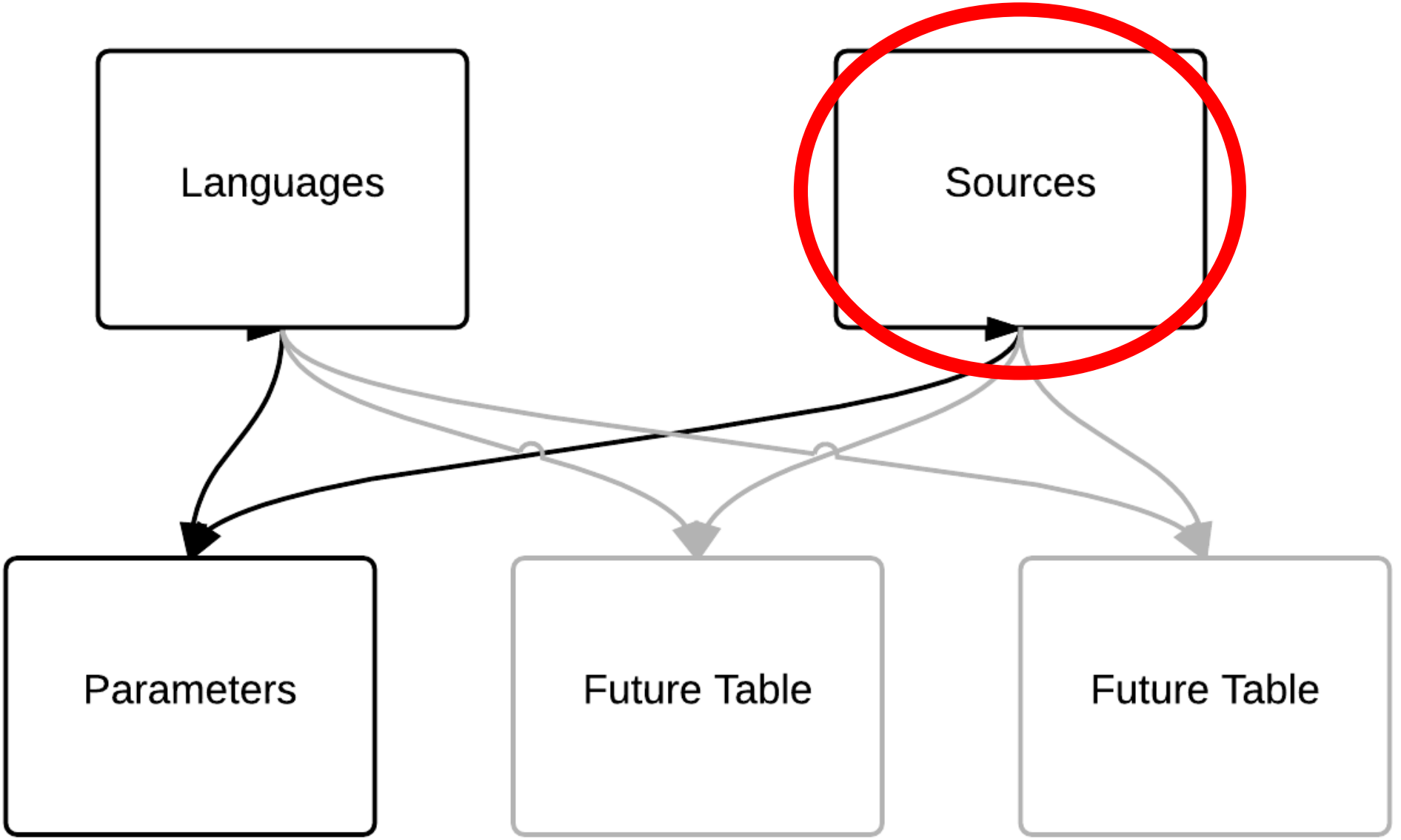
**Structure:**

| Column | DataType | Key? |
|---|---|---|
| idLanguages | INT | Primary |
| LanguageName | VARCHAR(45) | no |
| EthnologueID | VARCHAR(3) | no |
| Country | VARCHAR(45) | no |

**Example rows:**

| idLanguages | LanguageName | EthnologueID | Country |
|---|---|---|---|
| 1 | Adamorobe Sign Language | ads | Ghana |
| 2 | Afghan Sign Language | afg | Afghanistan |

# Current Structure



# Current Structure:
## References

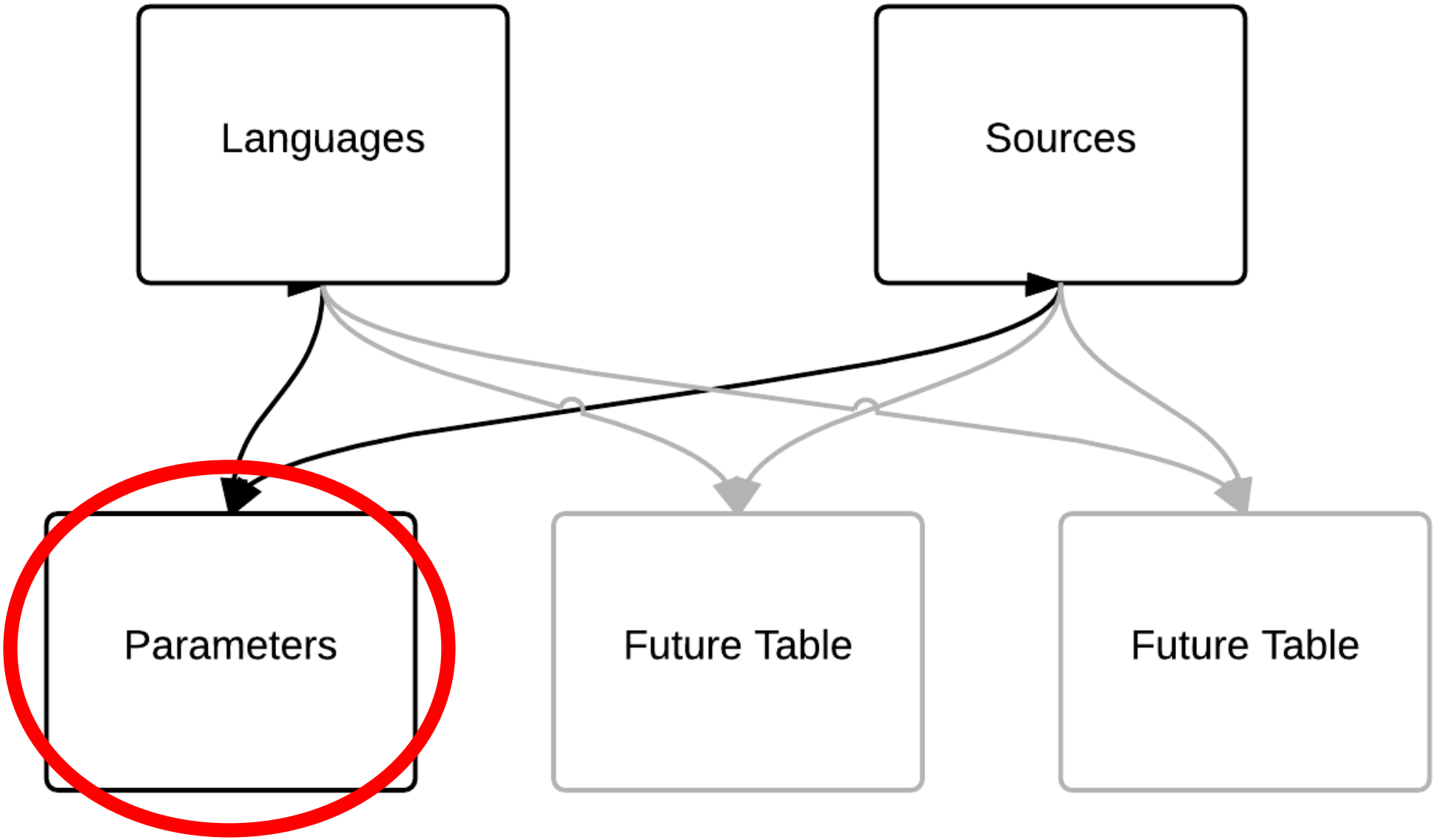
**Structure:**

| Column | DataType | Key? |
|---|---|---|
| IdReferences | INT | Primary |
| Title | MEDIUMTEXT | no |
| Author | MEDIUMTEXT | no |
| YearPublished | YEAR | no |
| AppearedIn | MEDIUMTEXT | no |
| URL | VARCHAR(150) | no |
| Bibtex | LONGTEXT | no |

**Example row:**

| Id References | Title | Author | Year | Appeared In | URL | Bibtex |
|---|---|---|---|---|---|---|
| 1 | Visual imagery and visual-spatial language | Emmorey, Kosslyn & Bellugi | 1993 | Cognition | | @article{emmorey1993visual, title={Visual imagery and visual-spatial language… } |

# Current Structure



# Current Structure:
## Parameters

**Structure:**

| Column | Datatype | Key |
|---|---|---|
| IdParameters | INT | Primary |
| LanguageParam | INT | Foreign |
| ReferenceParam | INT | Foregin |
| Handshape | BOOLEAN | no |
| Movement | BOOLEAN | no |
| Location | BOOLEAN | no |
| NonManualMarker | BOOLEAN | no |
| PalmOrientation | BOOLEAN | no |
| NumberOfHands | BOOLEAN | no |
| OtherParameters | LONGTEXT | no |
| Notes | LONGTEXT | no |

# Current Structure:
## Parameters (2)

**Example:**

| Id | Lang | Ref | Shape | Move | Loc | NMM | PO | NOH | Other Parameters | Notes |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 4 | 1 | TRUE | TRUE | TRUE | | FALSE | | Contact | |
| 2 | 6 | 3 | TRUE | TRUE | TRUE | TRUE | | TRUE | | |
| 3 | 7 | 4 | TRUE | TRUE | TRUE | | | | Speed | |

# Current Contents

- Languages table:
  - 135 languages
  - Ethnologue (Gordon 2004) was used as a starting place but is not entirely correct:
    - lacks some languages (e.g. Ghardaiaest Sign Language, Caucasian Sign Language)
    - lists dialects separately (e.g. Malagasy Sign Language and Norwegian Sign Language)
    - at least one language may be an idiolect (e.g. "Rennellese Sign Language" )
- 87 different signed languages in the parameters table
- 84 sources for parameters table; does not currently include more than one grammar per language

# Outline

1) Overview
2) Motivations
3) Organization
4) **Data and Input Analysis**
   1) Sources of Data
   2) Input Analysis (Parameters table only)
5) Distribution
6) Future Work

# Sources of Data

- Descriptive grammars
  - Pros: Linguistically-motivated, usually explicitly used named parameters
  - Cons: Not all in same scholarly tradition, many in other languages (e.g. French)
- Automatic sign recognition work
  - Pros: Usually in English, very short
  - Cons: Used feature-based systems, not linguistically motivated, did not engage with sign language research
- Misc: dictionary which listed minimal pairs, storybook in SignWriting (Sutton 1995)

# Input Analysis
## (Parameters table only)

- Three-way distinction for each parameter in Parameter table
  - NULL: Source did not discuss given parameter, cell left blank
  - TRUE: Source explicitly argued for parameter
  - FALSE: Source explicitly argued against parameter
- Not all sources used the same terminology/ scholarly tradition so some additional analysis was called for

# Input Analysis (2)
(Parameters table only)

- Examples:
  - “Place of articulation” in Danish sign language (Engberg-Pederse 1993) judged to be the same as location
  - “Point of articulation” and “hand configuration” in Irish Sign Language (LeMaster 1997) considered equivalent to location and handshape
- Provides room for potential introduction of bias or error

# Input Analysis (3)
## (Parameters table only)

- Minimizing bias and errors:
  - Base categories on previous review of the literature
  - Take detailed notes during analysis and review them later
  - Make a second pass including emergent themes taken from content analysis of notes (in this case, addition of Contact parameter)
- Other possible techniques:
  - Multiple researchers/inter-rater reliability measures

# Outline

1) Overview
2) Motivations
3) Organization
4) Data and Input Analysis
5) **Distribution**
   1) Current Distribution
   2) Long-term Solutions
6) Future Work

# Current Distribution

- Currently available via SQL Share (Howe et. al 2011)
  - Pros:
    - Available on-line for free
    - Can run SQL-based queries on-line (nothing to download)
  - Cons:
    - Does not actually support **relational** databases
    - Each table uploaded individually with no way to include foreign keys

# Long Term Distribution

- Hosting by the UW Linguistics department
  - Pros:
    - Can host full database
  - Cons:
    - Difficult to maintain once affiliation ends
- Other, long-term/archival solutions?
  - Commercial hosting (Amazon Web Services or Google)
  - WEDG (small personal server with cloud storage)

# Outline

1) Overview
2) Motivations
3) Organization
4) Data and Input Analysis
5) Distribution
6) **Future Work**

# Future Work

- Find a more permanent distribution solution
- Additional questions
  - Distribution of specific morphological, syntactic, semantic features
  - Relation of signed languages
  - Framework provided for other scholars to build on
- Additional Sources
  - Ongoing descriptive work
    - Example: On-going work on Ethiopian sign language coordinated by Dr. Binyam Sisay at Addis Ababa University
  - Include additional analyses for already included languages

# Outline

1) Overview
2) Motivations
3) Organization
4) Data and Input Analysis
5) Distribution
6) Future Work

# Citations


Hanke, T., König, L., Wagner, S., & Matthes, S. (2010, May). *DGS Corpus & Dicta-Sign: The Hamburg Studio Setup*. In 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT 2010), Valletta, Malta (pp. 106-110).

Crasborn, O., & Zwitserlood, I. (2008). *The Corpus NGT: an online corpus for professionals and laymen. In Construction and exploitation of sign language corpora*. 3rd Workshop on the Representation and Processing of Sign Languages (pp. 44-49). ELDA, Parijs.

DuBois, P. (2005). MySQL (Developer's Library). Sams.

Engberg-Pedersen, E. (1993). *Space in Danish Sign Language: The semantics and morphosyntax of the use of space in a visual language*. SIGNUM-Press.

Gordon, R. G. (2005). *Ethnologue: Languages of the world* (Vol. 15). Dallas, TX: SIL international.

Haspelmath, M. (Ed.). (2005). *The world atlas of language structures*. Oxford University Press.

Howe, B., Cole, G., Souroush, E., Koutris, P., Key, A., Khoussainova, N., & Battle, L. (2011, January). *Database-as-a-service for long-tail science*. In Scientific and Statistical Database Management (pp. 480-489). Springer Berlin Heidelberg.

LeMaster, B. (1997). *Sex differences in Irish sign language*. The Life of Language: Papers in Linguistics in Honor of William Bright, 67-85.

Moran, S. (2012). *Using Linked Data to create a typological knowledge base.* In Linked Data in Linguistics (pp. 129-138). Springer Berlin Heidelberg.

Sutton, V. (1995). *Lessons in sign writing*. SignWriting.

Zeshan, U. (2013). *Sign Languages*. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info/chapter/s9, Accessed on 2014-10-13.)