Toolbox 1

Susan Gehr
susan@gehr.info
Cell/text (707) 599-2719
With gratitude

- Albert Bickford, Toolbox instructor for InField 2008, 2010 & CoLang 2012
- Neil Brinneman, Shoebox instructor 2003
- Shoebox/Toolbox Field Linguist’s Toolbox Google Group
- Schoolmates at University of Oregon, especially Connie Dickinson
Course Goals

- Review of Toolbox for use in creating lexical databases
- Introduction to connections to the topics related to dictionary projects (workflow, lexicography, community involvement)
- Information about how to learn more and get help after this class
- Study of how Toolbox is part of the Karuk dictionary project
- Other related software

COLANG 2012
Institute for Collaborative Language Research
Getting to know each other

- What languages and projects are you working on with Toolbox?
- Did you bring your own computer and your own data to work with?
- Will you be using Toolbox on the computers in Watson 419? If so, I need to ask Jari to install the New Project Package.
- What are your current Toolbox-related questions or learning goals?
- Will someone else be setting up your project for you, or do you need to do it yourself?
Logistics

- Main class: Tuesday June 19 to Friday June 22
  2:15 to 3:45 PM
- Location of Class: 419 Watson Library
- Possibilities for consultation outside of class
June 19, Tuesday

- Introductions to our Toolbox projects
- Discussion of the parts and preparation behind every Toolbox Project
  - Language Encoding (orthography, sort order)
  - Database Type development
  - Workflow
  - Project Management

- Toolbox life cycle, considerations for migrating data
Course Outline 2

• June 20, Wednesday
  • Database types, continued
    • Establishing a data structure / cheatsheet
      • What are your fields? What rules do they require?
      • Documenting all decisions that you make about your workflow, whether you work alone or as part of a team
  • Getting to work – Using Toolbox

• June 21, Thursday
  • More on Using Toolbox – Based on class need & interest
Course Outline 3

- Friday, June 22
  - Related software (operating systems, Microsoft Word, Windows virtualization software for Mac users, backup software, etc.)
  - Outputting data
    - For online and print use (dictionaries, wordlists made with filters)
    - Working with a publisher or a printer
Karuk Toolbox Project

- Started in November 2003
- Took 2-week course at JAARS
- William Bright data and participation
- Nailing down the language encoding and revising it in 2009
- Publication of a printed dictionary (2005) and an online dictionary (ongoing with UC Berkeley Linguistics)
Toolbox Project Preparation

Before data entry can start, you need to ‘teach’ Toolbox some things about your project.

- Language Encoding
  - Orthography
  - Sort Order
  - Unicode compatibility
Language Encoding Features

• The order for sorting (using that script).
• Upper and lower case forms of the characters (if any).
• Special groupings of characters, variables, which are useful in examining or searching the data.
• A font to represent the character shapes.
• Often, a special keyboard to facilitate entering the characters.
Exercise: Creating your sort order

• What is the order of your writing system?
• Will your audience expect it to sort as English does, or will they expect some other sort order?
Language Encoding
Karuk_2009

Sort Order Properties

Name: Karuk_2009

Description: Sort order with default properties.

Note: characters are separated by spaces.

Primary characters (one group per line):
A A A a a a a
Ch ch
F F F e e e e
F f
H h
I I I I I I
K k
M m
N n
O O O o o o o

Secondary characters ordered before unmarked primaries

Secondary characters ordered after

Secondary characters follow their corresponding primaries
Ignore characters
- ! () '

OK
Cancel
Help
Toolbox Project Preparation

(Database type)
- What types of information do you want to keep track of?
- MDF – the Multi-Dictionary Formatter
  - What is it?
  - A database type that can function as a data structure standard and a data content standard
  - Read
    - MDFFields Toolbox Project and then read recommended parts of
    - MDF_2000.pdf
Toolbox Project Preparation

- Database type
  - What is your data structure?
  - What are your rules about entering data?
  - Do you keep a cheatsheet data structure and a notebook?
  - Karuk examples in Word and in Notebook
What are databases & database types?

In Toolbox, databases can be:
- dictionaries
- one text or a collection of texts

In Toolbox, a database type is a file that:
- includes a collection of properties that defines various fields of the database and some of the methods used for manipulating records.
More about database types

• A Dictionary database type might contain:
  – A recommended set of Field Markers
  – Filters used for finding particular records in a lexicon (e.g., all nouns, a particular morpheme, words with homonyms, etc)
  – The Date Stamp field marker `\dt`
Making your own Dictionary type

• You can & should make a database type for your language by copying the MDF type and modifying the copy (see Karuk MDF)
Fields in a Dictionary Database
What fields might I use?

- See the list of all MDF fields pp. 13-39, *Making Dictionaries, a guide to lexicography and MDF*.
- Take time before you start making lots of entries into your database to, otherwise you might have to either correct them by hand or have someone write a CC Table to correct all the records.
Exercise: Choose field markers & make a cheatsheet

• Basic minimum set
• \lx - Lexeme
• \a - Alternate form
• \u - Underlying form
• \ps - Part of speech
• \ge - Gloss
• \de - Definition
• \sd - Semantic domain
• \nt - Notes
• \dt - Date Last Edited
An example data structure

<table>
<thead>
<tr>
<th>Key</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>lx</td>
<td>Karuk Lexeme</td>
<td></td>
</tr>
<tr>
<td>lhn</td>
<td>Homonym number</td>
<td></td>
</tr>
<tr>
<td>lc</td>
<td>Citation form</td>
<td></td>
</tr>
<tr>
<td>la</td>
<td>Alternate form</td>
<td></td>
</tr>
<tr>
<td>lu</td>
<td>Underlying form</td>
<td></td>
</tr>
<tr>
<td>lps</td>
<td>Part of speech</td>
<td></td>
</tr>
<tr>
<td>lva</td>
<td>Variant form(s)</td>
<td></td>
</tr>
<tr>
<td>lge</td>
<td>Gloss (E)</td>
<td></td>
</tr>
<tr>
<td>lnr</td>
<td>Reversal (E)</td>
<td></td>
</tr>
<tr>
<td>lde</td>
<td>Definition (E)</td>
<td></td>
</tr>
<tr>
<td>lit</td>
<td>Literally</td>
<td></td>
</tr>
<tr>
<td>lsc</td>
<td>Scientific name</td>
<td></td>
</tr>
<tr>
<td>lso</td>
<td>Source</td>
<td></td>
</tr>
<tr>
<td>lfr</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>lex</td>
<td>Example (v)</td>
<td></td>
</tr>
<tr>
<td>lee</td>
<td>Example free trans. (E)</td>
<td></td>
</tr>
<tr>
<td>lfr</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>lex</td>
<td>Example (v)</td>
<td></td>
</tr>
<tr>
<td>lee</td>
<td>Example free trans. (E)</td>
<td></td>
</tr>
<tr>
<td>lee</td>
<td>Usage (E)</td>
<td></td>
</tr>
<tr>
<td>lcf</td>
<td>Cross-reference</td>
<td></td>
</tr>
<tr>
<td>lce</td>
<td>Cross-ref. gloss (E)</td>
<td></td>
</tr>
<tr>
<td>lsd</td>
<td>Semantic domain</td>
<td></td>
</tr>
<tr>
<td>lnt</td>
<td>Notes (general)</td>
<td></td>
</tr>
<tr>
<td>lng</td>
<td>Notes (grammar)</td>
<td></td>
</tr>
<tr>
<td>ldt</td>
<td>Date (last edited)</td>
<td>30/Jan/2010</td>
</tr>
<tr>
<td>lnum</td>
<td>ID number</td>
<td></td>
</tr>
</tbody>
</table>
Toolbox Project Preparation

• Workflow matters
  • Do you work by yourself?
  • Are you working on a team?
Toolbox Project Preparation

- Project management
  - Printed dictionaries have front and backmatter
  - Online dictionaries have websites with supplemental text, graphics and sound files
  - How will you keep track of master files, working files, passwords, contact information, and other valuable records?
Toolbox Project Preparation

- Toolbox life cycle
  - “Toolbox is nearing the end of its life-cycle.” Albert Bickford, SIL, June 2008.
  - So Toolbox must be really near the end here in 2012?
    - “the [Toolbox] programmer is still active…working on a kind of large new feature which is taking its own sweet time…Our assignment for the foreseeable future is Toolbox. All programs and people are mortal, but we are doing our best to keep ourselves and Toolbox among the living.” Toolbox Support. May 1, 2012
      http://groups.google.com/group/shoeboxtoolbox-field-linguistics-toolbox/
  - For now, I will keep using Toolbox, keeping my eye out for the day that it can’t be run on contemporary computers/operating systems and support dries up.
End of Day 1?

- Suggestions & adjustments for following days based on student need and preparation for Toolbox 2.
- Work on Language Encodings and Database Types
Day 2

- Work on Language Encodings and Database Types
- Database types & workflow matters, continued
  - Establishing a data structure / cheatsheet
    - What are your fields? What rules do they require?
    - What does the MDF book say about your chosen fields?
    - What if, 5,000 records in, you want to add new fields?
    - Documenting all decisions that you make about your workflow, whether you work alone or as part of a team
- Dictionary work Google Doc
On the Computer

• Go to the Project menu > Language Encodings…

• Select the vernacular Language Encoding and click the Copy button.

• Type your sort order into the Primary Characters box.
What to include in your Toolbox data structure

Information in a dictionary entry

- **Lexical/Semantic** – about the word and its meaning, how the word relates to other words in the language
- **Phonological/Phonetic** – pronunciation information
- **Grammatical** – paradigm forms, suppletion, gender/class information, etc
- **Social** – usage contexts, register, dialect, etc
- **Encyclopedic** – information about the item in the real world (e.g. how it’s made, where it lives, etc)
- **Historical** – etymology of the word, is it a loan, etc
- **Source** – where the words came from
Helpful Information

• Scientific Names [http://www.itis.gov/]
• Toolbox [http://www.sil.org/computing/toolbox/]
• (Carbon Copy Cloner) another backup solution for Mac
Day 3

June 21, Thursday

- Filtering data
  - Helps with analysis and problem solving
  - Let’s set some up and use them
- Collaboration documents
- Data structure matters continued
- Related software (operating systems, Microsoft Word, Windows virtualization software for Mac users, backup software, etc.)
Filters, Finding and Searching

- Find & Search are two different things in Toolbox

<table>
<thead>
<tr>
<th>Search</th>
<th>Find</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locates only at the beginning of a field.</td>
<td>Locates anywhere in a field</td>
</tr>
<tr>
<td>Locates only in the field sorted by.</td>
<td>Any sorting is possible, but it is best to sort by the first field of the entry.</td>
</tr>
<tr>
<td>Searches the entire database.</td>
<td>Hunts through the data only from the cursor forward (or back). Can be restricted to look within a single record.</td>
</tr>
<tr>
<td>Displays a list when it matches more than one item.</td>
<td>Travels linearly through the data, displaying matches as it comes to them.</td>
</tr>
<tr>
<td>Gives the option to match the whole field.</td>
<td>Gives the option to match the whole word.</td>
</tr>
</tbody>
</table>
Search

Match whole field

Match characters
- By primary grouping only (loosest)
- By secondary ordering, disregarding
- Exactly by secondary ordering
- Even those normally ignored (strictest)
Filter

Database Type Properties: Karuk MDF

Markers | Filters | Jump Path | Options | Interlinear
---|---|---|---|---
Name | Condition |

- **Certain Date**: \dt = 03/Jun/2008
- **Check these**: English [chk]
- **FF**: English FF 21 And \sd name ; n
- **Perhaps**: \nq Perh

Note: Cancel does not undo Add, Copy, Modify, or Delete.
Workflow Filter: All Records after a certain date
High Interest (and multi-part) Filter: All Personal Names

Filter Properties

Filter Name: names both

Filter Elements:
- Marker Text...
- Marker Date...
- Language Text...
- And
- Or
- Not
- ( )
- With
- Non-Unique

Element Text:
- \
d  name ; female
- Or
- \sd  name ; male

- Match whole field
- Match characters:
  - disregarding Case
Day 4

Friday, June 22

- Outputting data
  - For online and print use (dictionaries, wordlists made with filters)
  - Working with a publisher or a printer
- User testing
  - Does your main audience like your work?
Outputting Data: Print

• Word list – using a filter
  – Choose a \( \text{sd} \) – Semantic Domain filter and output a word list.

• Whole database
  – Karuk to English
  – English to Karuk
Outputting Data: Filtered Word List
Outputting Data: Filtered Word List

Multi-Dictionary Formatter Options

Dictionary
- Select Fields to be Excluded...
- Sort by citation form
- Include lexeme with citation form
- Include example sentences
- Include notes
- Include regional with national
- Include non-MDF fields

Gloss index (finderlist)
- Include part of speech

Marker hierarchy:

Document formatting
- Headwords in the header
- Date in the footer
- Total number of entries

Web page copyright:

Next process in chain (optional)
Outputting Data: Filtered Word List

[Image of a window with file save options: File name: household, Save as type: Rich Text Format (*.rtf)]
Outputting Data:
Filtered Word List

MDF Reminder

The file has now been exported as Rich Text Format, but the formatting is not yet complete.

To complete the formatting process in Word, choose Finish MDF Export.

After formatting is complete, you can view and print the document normally.

☐ Don't show me this reminder again

OK  Help
Outputting Data: Filtered Word List
Outputting Data:
Filtered Word List
Outputting Data:
Filtered Word List

To adjust the styles and formatting:
Outputting Data: English to Karuk
### Outputting Data: English to Karuk

<table>
<thead>
<tr>
<th>Letter</th>
<th>English</th>
<th>Karuk</th>
</tr>
</thead>
<tbody>
<tr>
<td>A - a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B - b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C - c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D - d</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- A - a
- B - b
- C - c
- D - d

**Additional Translation Examples**
- boat for washing hands: N'pakmuyvaxa / N'pakmuyvaxa
- bottle, glass: N'nah'sip
- broom: N'fuykmyvaxa / N'fuykmyvaxa
- basket: N'xishvaxa / N'xishvaxa
- chimney: LOC ablyyaxa
- comb: N'apthix
- comforter: N'matunyaxa
- corner of house pit: N'Mooyvaxa
- desk, cupboard: N'xishvaxa / N'xishvaxa
- doctor's office: N'turwa
Outputting Data: XML format

[Diagram of an XML export process properties window]

- Process Name: XML Export
- Fields to export:
  - All fields
- CC Table applied to XML output (optional)
- XSL Stylesheet (optional)
- Check consistency before exporting
- Output File: Z:\Desktop\karuk_dictionary2012_06_21.xml
- Overwrite file without asking
- Add alphabetic dividers
Outputting Data: XML format

<table>
<thead>
<tr>
<th>Name</th>
<th>Date Modified</th>
<th>Size</th>
<th>Kind</th>
</tr>
</thead>
</table>
Outputting Data: XML format
Outputting Data: XML to online

http://dictionary.karuk.org/

<table>
<thead>
<tr>
<th>Karuk Dictionary</th>
</tr>
</thead>
<tbody>
<tr>
<td>by William Bright and Susan Gehr</td>
</tr>
<tr>
<td>© Karuk Tribe</td>
</tr>
</tbody>
</table>

Your search: Karuk aachichha (including aachichha) | New search

**Search Index**
2 result(s)

- aachichha / aachichhi- be happy
- aachichhar / aachichhara- happy

**Dictionary Entry**
lexicon ID #2 | revised 18 Aug 2005

**aachichha / aachichhi- • v • be happy.**

Source: WB 8, p.313

Grammatical note: G622.1

**Text examples**

Display mode: sentence | word | gloss

- xás tu’aachichha patóo pma pamú’aramah.
  And he was happy when he saw his child.
  Source: Julia Starrett, “Coyote Marries His Own Daughter” (WB16) | read full text

- té ni’aachichha patá na’êe pamilápûrikh.
  I’m glad that you gave me your song.
  Source: Nettie Reuben, “Coyote Trades Songs” (WB7) | read full text
Working with a publisher or printer

• Find out their technical requirements
• Provide them with the format they require. In the case of my first printer, it was hard copy.
• See also “10. Completing the dictionary” from *Making dictionaries: a guide to lexicography and MDF.*
User testing


- Dictionaries can have advisory boards, which could be a community’s language committee or a subset of that committee.
  - Reviewing sections of the dictionary, handling questions or disagreements, discussing ‘new words’
Yôotva!

Susan Gehr

susan@gehr.info
Cell/text (707) 599-2719

http://dictionary.karuk.org/