

Working Group 2: Case Studies and Exhibit Hall

Andrea Berez (on occasion)

Jessica Boynton

Laura Buzsard-Welcher

Lyle Campbell

Östen Dahl

Naomi Fox

Johanna Nichols

Arienne Dwyer, chair

Case Studies to incl. Lang sketches (a.k.a. Exhibit Hall)

- Given that the primary purpose of the School of Best Practice is pedagogical (i.e. “This is how some people did it, and you can do it too”) we propose:
 - subsuming the language sketches and analysis found in the Exhibit Hall into Case Studies (renamed “Conversion Examples”?)
 - Entails deleting “Exhibit Hall” from Entrance Hall, and
 - replacing Case Studies pop-up text with an unambiguous introduction “Six studies of successful conversion of legacy data. Choose a project...” .

Case studies start page

- “Another purpose of the School is to introduce users to the languages & communities. You will find quick links to language data in the right sidebars of each Case study, under 'About the language'. ”
 - (This is the info formerly in the Exhibit Hall.)
- “You can access the language data directly in the right sidebar, or you can follow the steps needed to convert data from each language to better or best-practice formats.”

Case Studies tart page, para 2: “About this language”

- [para 2] incorporate text from the old Exhibit hall start page:
- “Please note that the 'About this language...' pages provide examples of language documentation in various stages of completion, which have been prepared in accordance with b.p. recommendations”
- The navigation of Lang sketches should be more standardized across the various language pages.

General desiderata

- In the fullness of time, aim for more geographical diversity in case studies
 - consider adding esp. Central America, PNG, Melanesia
- Include more detailed maps
- For each case study, make the balance of pedagogical material (how-to) vs. linguistic analysis explicit
- some found the 3-deep left toolbar difficult to use

general desiderata, II: clarify terms

- “presentation format”
- “portability”
- “access version” (=presentation version: re: two different image versions)

New Pages (1): Classroom

- in Classroom's Annotation and Transcription area
 - Transcription rendering issues: transcription vs. encoding; practical orthographies, IPA vs. other transcription systems
 - move SAMPA page here and label it as such (from Ega)
- In Classroom under new section “Practicalities” (alternatively, under “Ethics”):
 - Human Subject Review Committees – how to deal with them

[new page] To OCR or to Keyboard?

- (should be linked to ...notecards.html AND the Classroom as a separate drop-down menu item
- The page should make explicit the decision-making process (in prose or as a flow chart):
 - Is my text handwritten? (Y --> Keyboard)
 - N ---> Is it cleanly printed?
 - Does it contain nonstandard characters, especially if there are handwritten annotations or modifications on typewritten text?
 - (...--> OCR)
 - link to what programs are available, advantages of each .. etc.

[new pg] Refereed Publication Credit

- Have EMELD's Board of Advisors and specialist linguists referee lexicons, grammars, etc. on the site.
 - Explain this process in a new page in the Rdg. Room
- Top of each refereed page should say, “part of the EMELD publication series No. xx. This publication has been peer-reviewed.”
- Rationale: To encourage researchers to contribute data, and to encourage academia to formally recognize this analytic work as a publication
 - The EMELD site must communicate that all steps of this process are *analytical* in nature. (We know this, but it is not at all obvious to outsiders)

Biao Min (1)

- <http://emeld.org/school/case/notecards.html>
 - [Add before Para 1] [link] “Jump to the path of the Biao Min data”(and delete the right hand “pg. index”
 - [Para1]Add intro S: “ This project converted paper notecards to an XML table by scanning and archiving the notecards , and keyboarding the data.”
 - change “Mr. Fanggui” > “Mr. Deng” (Ditto for Monguor (Dr. Dechun > Dr. Li)
 - [Para2] Add a header: “To OCR or to Keyboard?”
 - After “over a decade”: “Then the researchers faced a decision that anyone with legacy print data has: can the scanned material be directly processed digitally, or must it be painstakingly re-entered by hand? For help in this decision-making process for your project, click here:
 - [hyperlink] “To OCR or to Keyboard” decision process

Biao Min (2)

- should put the available digital resources front and center
- should put the language sketches in right sidebar
- /Biao-Min/lexicon/, should make clear that “more” means “download tone file now”>> standardize links to A/C

Biao Min (3)

- <http://emeld.org/school/case/notecards.html>
- [para 2 cont'd] change S 2: “... manually entering and editing all the data... using the FIELD tool. [delete S 3]. (The FIELD tool allows data to be entered in IPA using a clickable input keyboard, as explained in the previous link.)
- [links] Search... About... should be moved to the bottom of the page, AND to a sidebar.
- [paras 3 &4] need header “Digitizing notecards as images”
-

Case studies, intro page

- Information available is a piece of the language, examples of some aspects of these languages. On the Case Studies pages, there may be links to more detailed description of these languages.
- (each page should include a step-by-step how-to – e.g. Saliba “time-linked annotation” without mentioning a tool)
- The PATH to CONVERSION
- Right sidebar would repeat this and show Gramm sketch

Maps & Lg families trees

- HSC statements can be linked to CSs
- Set off style sheets in sidebar and in text
- All pages should be heavily laden with examples
- need Search button at top so that you don't have to scroll down
- Maps – should have more detail
 - general overview map: should contain cities and rivers
 - detail map – should indicate where lg. spoken
 - (detail as large inset of general map)
 - Alternatively – wait for Atlas of Lg Families....

Follow the path....

- needs to be customized to each language; right now
- Mocoví style sheets (for comparative data): subgroupings include some unrelated subgroups, not explained
- need to specify sources for comparative data
- only focuses on treating the lexicon

Biao min (4)

- <http://emeld.org/school/case/notecards.html> (cont'd)
- under “Image details”, hyperlink each technical term (e.g. Bit Depth, Compression...) to the Glossary, and after these detail, add “ (For recommendations on image settings, see [link] Digitization of Existing Images)”
- R sidebar “Related links” : add ALL Biao Min ling data now under Exhibit hall under a link named “Language Sketch”

Mocoví

- On line lexicon: when you click on it, it takes you to a page where then you have to further hunt for the lexicon, which here, and in other languages (Saliva), Put clearer indication of how to get to the lexicon on this page.
- The dictionary needs to be highlighted and easier to find.
- Comparative word lists: (1) Not easy to find how to get to it; it and dictionary needs direct link on the page index. (2) There is no indication of where these lists come from – needs mention of the sources
- On the family tree link: says Guaykuran languages, but lists also Matacoan languages. Also, There are several Matacoan varieties listed as equal with the other Matacoan languages – these should be removed, or at least branching under Wichi/Mataco. Also, it is now known that the Matacoan languages branch first in two branches, with each branch further branching. We recommend removing Matacoan entirely, but if not removed, then tree should be Wichi and Chorote on one branch, and then Chulupi and Maka on the other branch.
- Map: since several other languages are represented in the comparative lists, it would be good to add their names/locations to the map.

Mocovi (2)

Editing/Textual Changes

- “bilingual helper” = change to “bilingual assistant”
- Change to read: Colonia “El Pastoril” near the town of Villa Angela in Chaco Province, Argentina (check spelling of names here).
- “3000 items” = “3000 lexical items”
- “long-term intelligibility” = “long-term accessibility”
- Change “[gamma] replaces as x” to “[gamma] is replaced by x”, or “x replaces [gamma]”
- “General rules for editing Mocovi data” = change to: “character conversion chart”
- Make explicit that only lexical data are available here for interaction.
- Can you add some lines of texts, so it is not limited only to lexical information.

Potawatomi (1)

- school\case\legacy.html
- Define: “flat file”, “depricated in db design”
- adjust tone in all paragraphs of this page to address multiple levels of expertise.
- dead link on page: #
- Lexicon: Lang=Potawatomi comes up on every page, search pulls up infl + other forms as 'no gloss available'

Monguor (1)

- <http://cf.linguistlist.org/cfdocs/emeld/school/case/tasx.html>
- map should be linked to the words in text, e.g. “Qinghai”
- “Monguor culture” rename as “Monguor sociolx.”
- Specify that the current data is one Monguor dialect, namely Minhe Monguor
- video and audio hung up on java applet
- Lexicon
C:\school\exhibits\monguor\lexicon.html:
 - data is too raw to be displayed: tokens shown as lemmas
 - a search for 'hand' did not produce this basic lexem

Monguor (2)

- add Chinese characters:
 - Li Dechun (Limusishiden) 李德春, Wang Xianzhen 王献珍
 - Qinghai 青海, Monguor 蒙古儿, 土
- Add modified Ethnologue codes used by the project:
 - MJGm Minhe county Monguor 民和蒙古儿
 - MJGh Huzhu county Monguor 互助蒙古儿
 - MJGt Tongren county Monguor 同仁蒙古尔

Monguor (3)

- Bibliography incomplete -
 - delete Slater diss ref and add: Slater, Keith W. 2003. A Grammar of Mangghuer. Curzon.
 - Change: Zhu et al. 1995 (add full ref)

Ega (1)

- Add [link] “Jump to the path of the Ega data”
- add publication info to the ref.
- standardize presentation format (in line with other pages)
 - sketch grammar + morphosyntax outline should be in R sidebar under [link] “Language sketch”
 - Under “scenario”: October when?

Ega (2)

- Exhibit Hall
<http://cf.linguistlist.org/cfdocs/emeld/school/exhibits>
- move technical data right into Case studies (most of the Exhibit Hall)
- Add the language data into the sidebar “Language sketch”
 - <http://cf.linguistlist.org/cfdocs/emeld/school/exhibits/ega>
 - http://cf.linguistlist.org/cfdocs/emeld/school/exhibits/ega/xsltransform.cfm?xmlfile=walds_lex_ega_07_b-tidied.xml&xslfile=lexicon-user.xsl

Sáliba

- Process of conversion not discussed until the bottom of the page. Should start with the conversion process.
- Keep naming consistent : care needs to be taken to keep names consistent or explain why, e.g. “Sáliva”, “Sáliba” , also “Salivan” -- recommend add comment on variant names, and, avoid anything suggestive of “saliva”, that is, keep the accent religiously.
- Line 1-2: Based on standardized 15-page wordlist – needs to be identified (something of Swadesh – Rowe) Where does the 'standard' come from?
- Lines 1-3 up: Says “created XML file to capture the details of the recording” ... sounds unclear about whether this is metadata, needs clarification ... though later says “and created a metadata description”, so is apparently not about metadata, though originally seems to suggest that.

Sáliba

Word lists and audio recordings Index page:

- Line 2: Again mentions “standardized wordlist” (of 375 words). Needs clarification of what this list is.
- Bullet 2: (1) What does “archival quality” mean? (2) For audio record could have a hyperlink to ...
- Paragr after bullet point 7, and elsewhere: Where mentions “recommended best practice” link to this Link to reading and audio-visual
- Same paragr: “access set” -- clarify meaning.
- Same paragr: Time-aligned transcriptions = need some indication of how they did this.
- Same paragr and elsewhere: they mention goals, but not so much on what they did, how did it;
- need clearer delineation of what they did, and bring this nearer the top.
- Follow path to Sáliva data = only goes to general procedures, not on Saliva itself. Fix.
- Exhibita Hall page for Saliva repeats / has the stuff on digitizing ... better to be in the Case studies
- Presentation Format takes you to word list, need direct link.