Contributing Data

Sharon Correll

2017 © SIL International

Types of data to contribute

- Character exemplars
- Sort specifications (collation tailoring)
- Word lists
- Phonograms (mapping of graphemes to phonemes)

2017 © SIL International

Character exemplars

- Must have a ScriptSource log-in
  - scriptsource.org
- Search for your language and go to writing system
- Click on Symbols & Characters tab
- Click on Add characters button
- Fill in character lists
  - Main – basic set of word-forming characters, including diacritics
  - Auxiliary – eg, for loan words
  - Index – eg, sections of a dictionary
  - Punctuation

2017 © SIL International

Character exemplars

- Fill in character lists
  - Main – basic set of word-forming characters
  - Auxiliary – e.g., for loan words
  - Index – e.g., sections of a dictionary
  - Punctuation
- Example is given on the page
  - Surround with [ ]
  - Include vowel marks and diacritics as separate characters
  - Can use either characters or USVs (e.g. \u0628)
  - Use curly braces around multigraphs

2017 © SIL International

Character exemplars

- Check that language and script are in the association boxes
- Click Submit
- Click I agree to these statements, and submit this content
- After contribution is approved:
  - Characters will be shown on writing system page
  - Character exemplars will be added to LDML file in the SLDR
    - LDML file will be created if necessary

2017 © SIL International

Sort specifications

- Get put as text directly into the LDML file

```
<?xml version="1.0" encoding="utf-8"?>
<ldml xmlns:sil="urn://www.sil.org/ldml/0.1">
...
<collation type=“standard”>
        <cr><![CDATA
                &B<t<<<T<s<<<S<e<<<
                &C<k<<<K<x<<<X<i<<<I
                &D<q<<<Q<r<<<R
                &G<o<<<O
        ]]></cr>
</collation>
...
```

- Email to Scriptsource (moderator@scriptsource.org)

2017 © SIL International

Phonograms

- Phonograms are mappings between graphemes (symbols) and phonemes (sounds), for a given writing system
- Richer information than a simple list of character exemplars
  - "What symbols are used to write a given sound?"
  - "What sounds are represented by a given symbol?"
- Imported into ScriptSource

2017 © SIL International

Phonogram

Combination of:

- Writing system
  - Languages with multiple writing systems need multiple phonogram data sets
- Phoneme
  - Phonemic not phonetic!
- Grapheme – recognizable character or sequence of characters.
  - "What people need to know in order to read"
  - Multiple characters = multigraph
  - Written with angle brackets <>

2017 © SIL International

Phonograms

- en: <ch> used to write /tʃ/ – ‘church’
- en: <ch> used to write /ʃ/ – ‘machine’
- en: <ch> used to write /k/ – ‘school’
- en: <tch> used to write /tʃ/ – ‘match’
- en: <sh> used to write /ʃ/ – ‘ship’
- es: <ch> used to write /tʃ/ – ‘mucho’
- fr: <ch> used to write /ʃ/ – ‘chien’
- de: <ch> used to write /x/ – ‘bach’
- ha-Latn: <c> used to write /tʃ/ – ‘cika’

2017 © SIL International

Phonograms

- en: <ch> used to write /tʃ/ – ‘church’
- en: <ch> used to write /ʃ/ – ‘machine’
- en: <ch> used to write /k/ – ‘school’
- en: <tch> used to write /tʃ/ – ‘match’
- en: <sh> used to write /ʃ/ – ‘ship’
- es: <ch> used to write /tʃ/ – ‘mucho’
- fr: <ch> used to write /ʃ/ – ‘chien’
- de: <ch> used to write /x/ – ‘bach’
- ha-Latn: <c> used to write /tʃ/ – ‘cika’

2017 © SIL International

Phonograms

- en: <ch> used to write /tʃ/ – ‘church’
- en: <ch> used to write /ʃ/ – ‘machine’
- en: <ch> used to write /k/ – ‘school’
- en: <tch> used to write /tʃ/ – ‘match’
- en: <sh> used to write /ʃ/ – ‘ship’
- es: <ch> used to write /tʃ/ – ‘mucho’
- fr: <ch> used to write /ʃ/ – ‘chien’
- de: <ch> used to write /x/ – ‘bach’
- ha-Latn: <c> used to write /tʃ/ – ‘cika’

2017 © SIL International

Phonogram template spreadsheet

- Introductory text
  - Sources, assumptions, simplifications, dialects, etc.
- Phonemes and graphemes
  - Phonemes are shown in the two left-hand columns
    - DON'T change the first column!
  - Add one line for each grapheme that can be used to write the sound
  - Graphemes – enter characters or \u1234
  - Status
  - Give an example and gloss for each combination; eg: chien 'dog'
  - Add comments in the notes column; eg "Only in such-and-such dialect"
  - Delete rows that are not needed

2017 © SIL International

Phonogram template file

- Nulls
  - Null grapheme – sound that is not written: use <>
  - Null phoneme – silent letters: use ∅ (cons_null or vwl_null)
- Diphthongs
  - A list is given; more can be added
  - Keep in mind the difference between rising and falling diphthongs
- Sequences – multiple sounds per grapheme
  - Eg, /ks/ written as <x> in English
- Possible to add "modified" forms of phonemes (eg, nasalized, aspirated, geminate, etc.) if needed
- Double articulations
  - A few are given; more can be added

2017 © SIL International

Phonograms

- Files in Language Data.zip:
  - PhonogramTemplate.ots / .xlsx
  - Example files – English and Spanish
- Dig in and if you get stuck we'll help!

2017 © SIL International

Word lists

- Used to create a spell-checking module for LibreOffice 5.3
  - Version 5.3 was recently released!
- Export word list from Paratext or Flex
- Utility to create a .oxt file from the word list
  - https://github.com/silnrsi/oxttools
- Install in LibreOffice: open the .oxt file.
- To check installation: look in Format | Character dialog
- Mark your text with the language and the spell checker should be activated
- Uninstall using extensions manager

2017 © SIL International