Orthography Development: the ‘midwife’ approach

Colleen Fitzgerald    Gwen Hyslop    Keren Rice    Kristine Stenzel

CoLANG 2012 INFIELD 2008/2010
Some background

- These slides were developed for a course at InField, taught in 2008 by Keren Rice and Kristine Stenzel and in 2010 by Gwen Hyslop and Keren Rice and in 2012 by Colleen Fitzgerald and Keren Rice.
- Gwen has been involved in orthography development in Bhutan, Keren in northern Canada, Kris in Brazil, and Colleen has worked with tribal writing systems in the Southwestern US.
- July 2012
Our Goals:

1. Discuss important questions and parameters (socio-political, technical-linguistic, pscho-cognitive) related to orthography development

2. Consider an approach to orthography development (o.d.) based on community involvement, writing practice, and analysis

3. Provide opportunities for hands-on analysis

4. Exchange experiences, brainstorm, expand resources
Contents of Power Point

- Introductory discussion (slides 5-9)
- Introduction to ‘Midwife’ approach (slides 10-28)
- Overview of linguistic issues (slide 29)
- Dealing with allophones (slides 30-38)
- Dealing with allomorphs (slides 39-44)
- Suprasegmental problems (slides 45-48)
- “New” sounds: Dene and Kurtöp (slides 49-77)
- Variation and standardization (slides 78-92)
- Review of Methodology (slides 94-111)
- Summary (slides 112-113)
- References/Contact info (slides 114-115)
Initial Discussion Questions
(for those interested in o.d.)

- What is an orthography?
- How would you define the role of the linguist in the process of orthography development?
- What do you think a language community expects from the linguist and from the orthography development process in general?
- What are the features of a ‘good’ orthography and what kinds of things do we need to know in order to develop one?
What is an orthography? Some thoughts for discussion

- Agreed upon system to represent sounds/words/concepts of a language
- Practical tool for communication
- …
What is the role of the linguist?

- Facilitator
- Mediator
- …
What does the community expect from the linguist?

- Intervention around different spellings and competing orthographies
- Expertise and connections that are not present in the community
- Legitimacy of the language
- ...

...
What are features of a ‘good’ orthography?

- Easy to learn and to produce
- Minimize number of characters, maximize what they represent
- Suitable to typology of the language
- Culturally relevant
- Transfer from matrix language
- Visually contrastive
- ...
- ...
The ‘Midwife’ approach

What is it?
The ‘midwife’ approach to the development of an orthography

- overall goal: to approach o.d. as a *process*
  - based on exchange and integration of knowledge and experiences of linguist and language community (LC)
  - with LC as active participant, sharing ‘joint responsibility’ for final outcome

- methodology: practice of writing and analysis of the language feed into each other

- linguist’s role: facilitator/guide in the practice - analysis dialectic
  - What kind of practice can help identify and focus the issues so that the analysis becomes more clear?
  - What kinds of appropriate metaphors can be useful tools?
Basic principles of the approach

- notion of o.d. as a **process** whereby members of a language community (LC) come to analyze aspects of their own language and develop a new practice: writing during the process (which may continue over an extended period of time), **orthographic variation** is ok
- **continuous and reflective practice** (LC writing and reading) is always the primary input to language-analysis activities
- LC **linguistic knowledge** and **social interpretations** are also a fundamental input
An overview of the ‘midwife’ approach to orthography development

Getting started 1: Discussion with LC
Getting started 2: Types of writing systems
Getting started 3: Learnability
Getting started in developing a writing system – 1: discussion with the LC

Why do we need to study our own language in order to think about writing it?

Discussion:

How are oral language and written language similar and how are they different?
### Oral Language . . .

- Communication between people in same place/time (immediate).
- Allows for reductions, use of body language and abbreviated deictic references, because misunderstandings / doubts can be resolved then and there.
- Is where innovations and change appear first.
- Always includes more types of variation, which may show different origins, group affiliations, or contexts requiring different registers (e.g. formal/informal).

### Written Language . . .

- Communication between people in different places/times (extended).
- Requires more complete forms and additional symbols to aid understanding – tools to make sure that the writer’s message will reach the reader intact.
- Tends to reflect stable forms, changes more slowly.
- May include (or not) variation that represents differences between groups of speakers of the same language, especially during initial phases;
- May be unified (or not) as a result of process of practice, analysis (discussion of variations and what they represent), and political decision-making.
Getting started – 2: presentation/discussion of types of writing systems and the symbols they use

What do symbols represent in different types of writing systems?

1. ‘Morphographic’ / ‘Logographic’ representations of words or morphemes

Pictographs

女 子 口 日 月 山 川 汉 目 心 雨 田 树 龟
woman child mouth sun moon mountain river pig eye heart rain field tree turtle

Compound Pictographs / Compound Ideographs

好 安 明 男 家 思 里 雷

<table>
<thead>
<tr>
<th>Compound</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>good (woman + child)</td>
<td>peaceful (woman under a roof)</td>
</tr>
</tbody>
</table>
2. ‘Phonographic’ systems: representations of **syllabic** combinations

**Cree**
3. ‘Alphabetic’ representations of individual sounds

The current thought in o.d. is that each symbol in an alphabet should represent a phonological segment, (ideally) corresponding (as directly as possible) to the phonemes of the language.

Consonantal alphabets: symbols represent consonants

Full alphabets: symbols represent consonants and vowels (e.g. Greek and Latin alphabets)
Our God Itzamna made his fire using friction.

‘Our God Itzamna made his fire using friction.’
All orthographies change over time

Roman Alphabet (2,600+ years old)

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>[G]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
<td>c</td>
<td>d</td>
<td>e</td>
<td>v</td>
<td>z</td>
</tr>
<tr>
<td>h</td>
<td>i</td>
<td>k</td>
<td>l</td>
<td>m</td>
<td>n</td>
<td>o</td>
</tr>
<tr>
<td>p</td>
<td>q</td>
<td>r</td>
<td>s</td>
<td>t</td>
<td>u</td>
<td>ks</td>
</tr>
</tbody>
</table>
Getting started – 3: discussion of ‘Learnability’

Who is the writing system for? Will it be used primarily by native speakers? Learners of the language?

What kinds of orthographic features might help increase ‘learnability’ for each of these target groups?
An important assumption

- There is a writing system to begin from. (For instance, learners are literate in another language such as English or Spanish.)

- In such cases, familiarity with an existing system will probably lead the LC to adopt a similar type of orthographic representation, but will require analysis so that they can recognize where ‘adjustments’ need to be made.
A ‘getting started’ exercise

This type of exercise works well in workshop-type situations, and will likely provide activities for many days of work. It is a good way to get a large variety of members of the LC involved in the discussions. If activities are organized in groups, literate and non-literate individuals and speakers with varying degrees of fluency can have input.
A. LC participants choose a theme (or themes) and write short texts (individually or in groups)

B. Participants exchange texts to read, making lists of doubts they encounter or alternative ways to write specific words
C. Participants present their doubts and suggestions to the entire group – this is the data that will guide the analysis and inform decision-making.
What kinds of information are likely to be revealed by this initial exercise?

In terms of orthographic symbols, that:

• Various symbols are being used for the same sound

• No symbol is available for a sound in the language

Both cases may result from the effect of literacy in a different language or from alternate existing orthographies. Recognition of where the problems lie is a first step in analysis.
In terms of phonology:

- Sets of examples of important phonological elements and indications as to their ‘functional loads’
- Evidence of allophonic variation
- Indication of variation between speakers of different ages or from different regions

In terms of morphology:

- Questions as to word boundaries and other morphological issues such as what to do about compound words or complex constructions
Continuing the exercise . . .

D. As the participants present the results of these activities, the linguist should be able to recognize and group together the different categories of ‘doubts’ and begin to think about how to work on them with the LC

E. Subsequent activities should focus on individual issues, analyzing them with the LC so that informed decisions can be made collectively
Linguistic issues: what to do about . . .

- Allophones
- Allomorphs
- Suprasegmentals
- Sounds in the language that are not represented in a known writing system
Representing allophones
Allophones in English

pit \quad \text{[p^h]}
spit \quad \text{[p]}

Allophones share a representation in the orthography.
An example of allophones and their representation in the orthography in o.d.:

Kotiria (Eastern Tukanoan) [d] and [r]

In this language, as onset consonants, these sounds occur in complementary distribution: [d] word-initially and [r] word-internally

dukuri ‘manioc roots’

duhire ‘you/he/she/they sat’

diero ‘a dog’

What decision was made in this case?
Analysis with the LC

a) Participants in the language workshop compiled a list of words containing the two sounds from their own written texts

b) All occurrences of the sounds were highlighted, so that participants could visually observe their distribution

c) Participants were asked if they could think of other words with different sounds in the positions of [d] and [r] (in other words, to find minimal pairs), leading to analysis and recognition of /d/ as a ‘basic sound’ (phoneme) and [r] as a ‘variant’ (allophone)
d) Once speakers had observed and analyzed for themselves that [r] was a variant of /d/ in a specific position, it was possible to discuss whether or not to represent it with a different symbol

e) Collectively, several of the texts were re-written using only ‘d’ and speakers were asked to evaluate how they felt, as writers and readers, about the use of a single symbol
Coming to conclusions

f) While recognizing /d/ as underlying sound, use of the symbol ‘d’ in both positions felt uncomfortable to the participants. They argued it contradicted a well-established surface distribution of sounds, making the written and spoken versions of the language look too different. Additionally, use of ‘d’ in word-internal position made the written texts look like they represented the pronunciation of closely-related languages in the family, in which the d-r distribution does not occur.

g) Thus, the LC has opted to use different symbols for the ‘d’ and ‘r’ sounds in the orthography, a decision informed by linguistic analysis but respectful of input from the LC as the end users of the system.
Another Allophone Example

- Choctaw, a Native American language in Mississippi and Oklahoma has had a number of orthographies.
- The language has three vowel phonemes, /i o a/, which can be short, long, or nasalized.
- However, the writing systems often use six symbols, following how the language was written in the 19th century, associated with Cyrus Byington.
Choctaw Vowels
(from Alphabet links at Choctaw Language School online)

Two of the allophones of phoneme /a/ get represented in the writing system. Allophone [a] tends to appear in open syllables, written as a.

- chaha  'tall'
- taloowa  'sing'

The other allophone, [ə], tends to appear in short closed syllables and is written using a symbol not used as a vowel in English, v.

- anvmpa  'word, language'
- kvllo  'hard'
Conclusions from the Choctaw allophones

- This example using allophones shows that sometimes the choice is made to write allophones.
- Not all Choctaw vowel allophones are represented with unique orthographic symbols, but some are.
- We will see some parallels in the upcoming allomorph examples from English, where some of the variation can be chosen to be written overtly in the writing system.
Questions of allomorphy
Problems of allomorphy

Shallow vs. deep orthographies

Shallow: close to pronunciation

Deep: preserves graphic identity of meaningful elements
English allomorphy

A combination of deep:
  cats [s]       dogs [z]

And shallow:
  intangible [n]   impossible [m]
# Allomorphs: Dene voicing alternations

<table>
<thead>
<tr>
<th>Shallow orthography?</th>
<th>sa</th>
<th>sezá</th>
</tr>
</thead>
<tbody>
<tr>
<td>(“phonetic”)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deep orthography?</td>
<td>sa</td>
<td>sesá</td>
</tr>
<tr>
<td>(“phonemic”)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sa</th>
<th>‘watch’</th>
<th>sezá</th>
<th>‘my watch’</th>
</tr>
</thead>
<tbody>
<tr>
<td>xa</td>
<td>‘hair’</td>
<td>seghá</td>
<td>‘my hair’</td>
</tr>
<tr>
<td>shá</td>
<td>‘knot’</td>
<td>sezhá</td>
<td>‘my knot’</td>
</tr>
</tbody>
</table>
The process 1

- An orthography standardization committee was established to make decisions about the orthography.
- A few decisions involved symbols; most involved spelling conventions.
- The committee considered basic principles – audience, goals of writing, transfer from English, …
The process, continued

- The committee identified areas of concern with the different choices.
- People experimented with the different ways of writing words with these alternations.
- Decision: shallow orthography
- Why? Easier to figure out from the pronunciation
Beyond the segment:
suprasegmental problems
Nasalization in Tukanoan languages

In Tukanoan languages, nasalization is a property of the morpheme rather than of individual segments, thus it functions as a suprasegment and the question quickly arises as to how it should be represented in the orthography of these languages.
Analysis of nasalization with the LC:

Finding metaphors to help speakers understand how nasalization works in different kinds of languages . . .

‘umbrella’ nasalization – ‘covers’ the entire morpheme (Tukanoan languages)

‘raincoat’ nasalization – ‘covers’ individual segments (e.g. in Spanish / Portuguese)
Some nasalization proposals

Over time, after analyzing and understanding how nasalization operates in Tukanoan languages, a number of different proposals for how to mark nasalization were ‘tested’ by participants in language workshops. In each case, writing and reading exercises utilizing the different possibilities were proposed, practiced, and then evaluated. Eventually, it was collectively decided that:

- Morphemes with nasal consonants (m, n, ñ) require no further marking, the nasal C being sufficient to identify the morpheme as +nasal

- In morphemes with no nasal C, the first vowel is marked with a tilda: ñ to indicate the morpheme as +nasal
‘New’ sounds: sounds not distinctly represented in a known writing system
‘New’ sounds 1: Dene mid front open and closed vowels

- Early orthography used the symbol {e} for both an open and closed front vowel.
- Both these vowels exist in some dialects.
The process

- A question: Should these vowels be differentiated?
- The answer: yes!
- Why?
  - Accurate representation of sounds of the dialect
  - Ease of reading
The process: What symbol to use?

- A new symbol is needed.
- The open vowel is more common than the closed vowel.
- Choice: symbol \{e\} for the open vowel; schwa (‘upside down e’) for the closed vowel.
- This choice was made because the open vowel is more common, and it meant fewer changes in how people were already writing.
- This decision was a surprise for some of the linguists involved, but people liked it because they knew that schwa was used in linguistics.
‘New’ sounds 2: The case of Kurtöp

- Kurtöp is a previously undescribed Tibeto-Burman language of Bhutan
- About 15,000 speakers
- Speaker who are literate are usually familiar with 1) English and 2) Dzongkha
- Roman orthography was a natural product and ’Ucen was suggested by community and Dzongkha Development Commission
Which versions of ‘Ucen?

• We opted to begin with *joyi*, since it was what children learned and was purely Bhutanese (as opposed to *tshui*, which is shared with Tibetan).
The ’Ucen syllable

In the Classical Tibetan Orthography, an abugida derived from Brahmi, and devised in 632 AD, syllables are represented according to this diagram.

The “R” represents a simple onset, or in the case of an onset-less syllable, the vowel. C1, C2, and C4 may be used to add consonants to the onset, making it complex. The V slots are for vowels (i, e, o go above; u goes below). C3 represents a single coda (if present) and C5 makes a complex coda (rarely occurs).
The ’Ucen syllable

For example, this is how the Classical Tibetan word /bsgrubs/ was written. The complex onset is represented by <b> in C1 position, <s> in the C2 position, <g> in the root position, and <r> in the C4 position. The vowel /u/ is represented below the C4. <b> in C3 and <s> in C5 indicate the complex coda.
The ’Ucen syllable

Traditionally, there is a fixed number of symbols available for each slot. C1 may be one of five symbols.; R may be one of 30; C2 may be one three (one of which is modified from its occurrence elsewhere); C3 draws from ten possible symbols; C4 draws from a set of five (mainly) ‘half’ symbols; and C5 may be one of two. The top V may be one of three vowel diacritics and the lower V is reserved for one diacritic.

In Joyi, various combinations of C2 with R, or C4 with R, lead to unique symbols reserved for the exclusive representation of the combination, similar to ‘conjucts’ in devanagari.
’Ucen and Tibetan

- Classical Tibetan phonology had around 28 consonants (labial, dental, palatal velar).
- And complex onsets
- And five vowels
- No tone

’Ucen was designed for this phonology
’Ucen and Tibetan

- However, after almost 1,400 years of change, Lhasa Tibetan (the prescribed standard) has:
  - A new series of retroflex consonants
  - Two new vowels (front high and mid rounded)
  - High and low tonal registers; level and falling tonal contours
  - Changes in voicing/aspiration contrasts
  - Simplified onsets
’Ucen and Bhutan

- The modern use of ’Ucen assumes the 1400 years of change from Classical Tibetan to modern Lhasa Tibetan.
- ’Ucen is used this way in Bhutan; for example, words with complex onsets in Classical Tibetan are still written as such in modern Tibetan/Dzongkha, but not pronounced as such.
- Representing any pronunciation using ’Ucen entails the reader to infer the sound change.
- There is no way to represent various aspects of the phonology – such as the complex onsets – in the history of Bhutanese education.
’Ucen and Tibetan

• For example, the spelling <bsgrubs> is pronounced: qùp
’Ucen and Kurtöp

- Kurtöp is not a descendent of Classical Tibetan.
- The phonology of Kurtöp is different from the phonology of Classical Tibetan or Dzongkha.
- Kurtöp tone, vowel length, and complex onsets are particularly difficult to represent.
- The following is an illustration of how we chose to represent complex onsets.
Kurtöp phonology

Kurtöp complex onsets
The problem

<pr-> is pronounced as a voiceless retroflex, but in Kurtöp /pra/ = ‘monkey’
Midwife process

- We presented ideas to a small group of literate Kurtöp speakers;
- Consulted local teachers
- Consulted highly educated speakers of related languages with similar phonologies
Midwife process

Idea 1: Use ’Ucen in a way similar to Roman.

But the following problem developed:
How to represent vowels other than /a/?
Midwife process

This would be confused with /lé/ in Dzongkha/Tibetan conventions

This leads people to tend to pronounce the word correctly, but does not follow the traditional conventions and is unattractive.
Midwife process

- In 2009 we organized a workshop with the Dzongkha Development Commission, Scott DeLancey, local leaders and interested community members to address all the issues.
2009 DDC workshop
2009 DDC workshop
DDC 2009 workshop
Proposed solution

• We will add ‘half’ letters to be used directly below the root consonant.
• Based on existing (but rarely used) conventions established in Tibetan to represent different languages.
• Should not affect Dzongkha transference issues
• Aesthetically pleasing
• Kurtöp speakers find it intuitive and easy to read
Proposed solution

- Existing computer fonts do not allow the needed combinations
- Chris Fynn, DDC font developer, agreed to adapt the Bhutan ’ Ucen fonts (joyi and tshui) to accommodate the new combinations
- In addition to the complex onsets, the adapted fonts will be able to mark tone
Proposed solution

• Tshui font is finished but the Joyi font has been held up indefinitely for unknown reasons.
• In addition to handling the ‘new’ complex onsets, we also have a way (marks above the other symbols in top row) to mark tone, another ‘new’ sound.
Moving forward (the midwife process continues)

- The Kurtöp/English/Dzongkha dictionary is expected to be published in 2013.

- Kurtöp entries will use the new font and proposed combinations, in Joyi if it is made soon, or else using Tshui.

- Testing will continue…
What have we seen – and not

- Linguistically, we looked at orthography choices with respect to the implications of representing:
  - allomorphs, allophones, suprasegmentals, and sounds in the language that are not represented in a known writing system

- With sounds that are not represented in a known writing system, different choices might be made, with different pros and cons to each choice.

- Let's consider Choctaw, which has a voiceless lateral, IPA symbol /ɬ/. 
Considering implications of symbol choice and the language's phonology

- Using IPA:
  - Pro: linguistic representation, new representation for unfamiliar sound
  - Con: font, no transference

- Adapt English symbols:
  - Pro: familiar symbols
  - Con: symbols used in unfamiliar ways

- We could imagine *lh* and *hl*. Choctaw uses both, in different environments.
  - *lh* (**pvlhki** 'fast') before a consonant and *hl* before a vowel (**hlampko** 'strong')

- Pro: uses familiar symbols, no font challenges
- Con: Confusion with phonemes [h] and [l] with words like (**mahli** 'wind', **asil.hah** 'to request')
The realities of language: variation, and standardization
Orthographic Variation

Discussion questions:

- Is orthographic variation a problem and if so, why?
- What kinds of variation are we likely to encounter?
- What kinds of things can variation represent?
Standardization

- What are some of the advantages and disadvantages of ‘standardization’ or ‘unification’ of an orthography?
Kinds of variation

- Variation at a regional level
- Variation at a local level
- How can these be dealt with?
**Between community variation: an example from Dene dialects**

<table>
<thead>
<tr>
<th>South Slavey</th>
<th>Mountain</th>
<th>Déline</th>
<th>Hare</th>
</tr>
</thead>
<tbody>
<tr>
<td>-tthí</td>
<td>pí</td>
<td>kwí</td>
<td>fí     ‘head’</td>
</tr>
<tr>
<td>tth’ ih</td>
<td>p’ ih</td>
<td>kw’ ih</td>
<td>w’ i ‘mosquito’</td>
</tr>
<tr>
<td>tha</td>
<td>fa</td>
<td>wha</td>
<td>wa ‘sand’</td>
</tr>
<tr>
<td>-dhe</td>
<td>-ve</td>
<td>-we</td>
<td>-we ‘belt’</td>
</tr>
</tbody>
</table>

Should there be a common spelling for the different dialects?
The process

- Discussion of dialects: systematic differences
- Discussion of spelling possibilities
  - one spelling for all dialects?
  - different spellings for each dialect?
The decision

Write each dialect with its own symbols (e.g., \textit{tth ‘ih} in South Slavey and \textit{w ‘i} in Hare)

Reasons

transferability from English dialect identity
Within community variation: an example from Dene

- zha ya ‘snow’
- zhú yú ‘clothing’
- zhíi yíi ‘inside’

Some questions to ask

What might underlie this variation?
Is the variation really free?
The first decision

We began with a discussion of variation and the different ways of dealing with it.

The first decision: standardization
- Write zh if it is ever used in that word.
- If only $y$ is used, write $y$. 
And the development over time

This did not work in practice
  - variation among individuals
  - no resource materials

Consequence: Both \textit{zh} and \textit{y} are used.

Lesson: Early decisions might have to be changed based on practice.
From related dialects to related languages

- The Dene example shows how different sounds are treated in closely related varieties.
- What choices might be made in representing similar sounds in closely related languages?
  - One possibility would be to choose the same symbol.
  - Another would be to represent the same sound in different ways. This is what has happened in Muskogean languages.
How to represent similar sounds in closely related languages?

- The Muskogean languages include Muscogee (Creek), Seminole Creek, Choctaw, Chickasaw, Alabama and Coushatta/Koasati.

- All have a phoneme /ɶ/, a voiceless lateral, but the languages make different orthographic choices.
  - Choctaw uses lh (pvlhki 'fast') before a vowel and hl before a consonant (hlampko 'strong')
  - Chickasaw uses lh consistently (hilha 'dance')
  - Muscogee (Creek) uses r (rvrō fish)
  - Alabama uses Ʉ (ɄaɄo 'fish')
  - Coushatta uses th (thatho 'fish')
  - Linguists vary in documentation, mostly lh or Ʉ
Writing and Variation in O'odham

- The O'odham varieties include Tohono O'odham (formerly Papago), 'Akimel O'odham (formerly Pima), and the Mexican variety, Sonoran O'otam.

- Multiple writing systems in use, which were developed in a variety of contexts.
  - Tohono O'odham Nation and the Salt River community use the Alvarez and Hale orthography, which was developed as a linguist-native speaker collaboration.
  - The Saxton orthography is a practical orthography and was tested out with native speakers, and is use in the Gila River Indian Community.
  - The influence of Spanish as a transfer is leading the Sonoran O'otam to consider another option.
  - Linguist Madeleine Mathiot uses yet another system in linguistic documentation.
Some differences in the four writing systems for shared sounds

<table>
<thead>
<tr>
<th></th>
<th>A&amp;H</th>
<th>Saxton</th>
<th>Mathiot</th>
<th>Sonoran proposal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long vowels</td>
<td>a:</td>
<td>ah</td>
<td>aa</td>
<td>aa</td>
</tr>
<tr>
<td>Palatals</td>
<td>ň</td>
<td>ni</td>
<td>ň</td>
<td>ň</td>
</tr>
<tr>
<td>Retroflexes</td>
<td>ḍ ṣ</td>
<td>d sh</td>
<td>ḍ x</td>
<td>sh th</td>
</tr>
<tr>
<td>Voiceless vowels</td>
<td>ĕ</td>
<td>n/a</td>
<td>ĕ</td>
<td>n/a</td>
</tr>
<tr>
<td>Palatal affricate</td>
<td>c</td>
<td>ch</td>
<td>c</td>
<td>ch</td>
</tr>
<tr>
<td>Glottal stop</td>
<td>'</td>
<td>'</td>
<td>?</td>
<td>'</td>
</tr>
<tr>
<td>Lateral flap</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>r</td>
</tr>
</tbody>
</table>
Sounds which vary across dialects

- [w] vs. [v]/[u] –
  - Alvarez and Hale goes with w
  - Saxton goes with w
  - Mathiot goes with v
  - Sonoran proposal goes with v

- Dialect variation within Tohono O'odham dialects for certain vowel sequences, like io or eo hiosig vs. heosig 'flower'

- These are acknowledged and both end up used.
Standardization

Standardization often emerges as the writing system is used; it may not be the best starting point.

What do potential users want from writing?
A review of the method
Methodology: a review

The ‘midwife’ approach views input from LC as fundamental, this input consisting of:

- **practice** (written material produced by the LC that concretely reveals issues for analysis, discussion, and decision-making)

- **LC insights** (about the language itself, socio-political issues, and their experiences)
  - Do members of the LC regularly write/read in any language? Are writing/reading themselves new experiences for them? How can these new practices be expanded and reinforced?
The approach also relies on interwoven activities of analysis leading to periods of experimentation of whatever ‘decisions’ have been agreed upon, with ongoing evaluation by the LC in both the roles of writers and readers.

The LC may be viewed more broadly, as in Bhutan, in which the government is necessarily involved.

Throughout the O.D. process, the linguist should build ongoing written record with explanations and examples of the analysis and discussion that went into each decision.
The role of the Linguist

interpret LC feedback, looking for clues as to:
- the functional loads of phonol. features
- other important cognitive issues
- interference issues
- socio-political issues

monitor and interpret written input

practice

analysis

evaluation

LC

organize analysis and discuss options

suggest further practice and record decisions
Some further issues

functional load
cognitive needs
socio-political issues
technological issues
who is the audience?
Complex issues in orthography development

- Evaluating the functional loads of phonological features
- Recognizing the cognitive needs of the users of the orthography
- Discovering socio-political issues: what else the writing system represents for the LC and the wider community (if relevant).
Evaluating the ‘functional load’ of suprasegmental features: examples from Kotiria

In Kotiria, three suprasegmentals are associated to root morphemes: *nasalization*, *glottalization*, and *tone*

- Minimal pairs are found for all three:

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>waha</strong> ‘drag/row’</td>
<td><strong>wãhã</strong> ‘kill’</td>
<td><strong>doa</strong> ‘envy’</td>
<td><strong>do’a</strong> ‘cook’</td>
<td><strong>kóró</strong> ‘rain’</td>
<td><strong>kórò</strong> ‘umbrella’</td>
</tr>
<tr>
<td><strong>hu</strong> ‘smoke’</td>
<td><strong>hũ</strong> ‘worm’</td>
<td><strong>maa</strong> ‘stream’</td>
<td><strong>ma’a</strong> ‘be small’</td>
<td><strong>khòá</strong> ‘leave’</td>
<td><strong>khóá</strong> ‘part/half’</td>
</tr>
<tr>
<td><strong>sa’a</strong> ‘dig’</td>
<td><strong>sã’ã</strong> ‘electric eel’</td>
<td><strong>waa</strong> ‘give’</td>
<td><strong>wa’a</strong> ‘go’</td>
<td><strong>sóà</strong> ‘grind’</td>
<td><strong>sóá</strong> ‘rest’</td>
</tr>
<tr>
<td><strong>kha</strong> ‘hawk’</td>
<td><strong>khã</strong> ‘chop’</td>
<td><strong>wa’ma</strong> ‘young/new’</td>
<td><strong>wama</strong> ‘name’</td>
<td><strong>báa</strong> ‘decompose’</td>
<td><strong>baá</strong> ‘swim’</td>
</tr>
</tbody>
</table>
However, despite shared phonemic status, each suprasegmental feature has a different functional load. This variation is manifested in spontaneous writing and has been discussed throughout the o.d. process.

**nasalization**

***

- ++ salient
- (Roots / Suffixes)
- ++Min.Pairs
- value unaffected by morphological processes
- always marked in spontaneous writing

**glottalization**

**

- + salient
- (Roots only)
- ++M.Ps
- reductions occur in morphological processes
- marked most, but not all, of the time in spontaneous writing

**tone**

*

- + salient
- (Roots, few Suffixes)
- +M.Ps
- melody variable in morphological processes
- not marked in spontaneous writing
Recognizing cognitive issues:

- In Kotiria root morphemes, internal voiceless Cs are *always* pre-aspirated, a regular allophonic variation. From the purely linguistic perspective, this aspiration would not need to be represented in the orthography.

- Thus, the words could be written as:

  | [daʰpo] ‘head’ | dapo |
  | [maʰsa] ‘people/beings’ | masa |
  | [tuʰti] ‘to bark’ | tuti |
  | [puʰka] ‘blowgun’ | puka |
  | [daʰfo] ‘day’ | dacho |
However, given the salience of this aspiration and the fact that when written, it helps readers identify the root morpheme in a word, the decision was made to represent this pre-aspiration in the orthography.

+ articulatory salience
+ root recognition in reading

Thus, the words

[daⁿp'o] ‘head’
[maⁿsa] ‘people/beings’
[tuⁿti] ‘to bark’
[puⁿka] ‘blowgun’
[daⁿt'o] ‘day’

are written as:

daⁿp'o
maⁿsa
tuⁿti
puⁿka
daⁿt'o
Examples of symbolic-political choices in O.D. in Kotiria

- Use of the symbol ‘k’ over ‘c/q(ui/e)’ – a macro-level choice, to distinguish the writing system of the indigenous language from those of the national languages (Spanish/Portuguese).

- Use of the symbol ‘ʉ’ over ‘ɨ’ – a regional-level choice, to differentiate the orthography of a minority indigenous language from that of the locally dominant indigenous language (Tukano proper);

- Variation between use of the symbols ‘w/v’ among the Kotiria from different regions – a group-internal choice distinguishing sub-groups within the Kotiria population.
Cognitive/social issues: Phonemic-based systems might prove unpopular

Choosing English-based writing systems over phonemic systems

Navajo code talkers

<table>
<thead>
<tr>
<th>Wol-la-chee</th>
<th>‘ant’</th>
<th>Wóláchíí’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shush</td>
<td>‘bear’</td>
<td>Shash</td>
</tr>
<tr>
<td>Moa-si</td>
<td>‘cat’</td>
<td>Mósí</td>
</tr>
<tr>
<td>Klizzie</td>
<td>‘goat’</td>
<td>Tliízí</td>
</tr>
</tbody>
</table>

Young and Morgan dictionary
Familiarity with English-based systems

**Eastern Pomo**

<table>
<thead>
<tr>
<th>Phonemic Orthography</th>
<th>Local English Orthography</th>
<th>Pronunciation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>káli</td>
<td>caw lee</td>
<td><code>dole</code></td>
<td>‘one’</td>
</tr>
<tr>
<td>do:l</td>
<td>dole</td>
<td><code>four</code></td>
<td></td>
</tr>
<tr>
<td>lé:ma</td>
<td>leh ma</td>
<td><code>five</code></td>
<td></td>
</tr>
</tbody>
</table>
One more factor

- What is the writing system for? What does a writer/reader want from it?
  - primacy or written text?
  - valuable information about the speaker?
  - symbolic system?
  - something else?
Writing Systems for Endangered Language Communities

- Issues when literacy is used for second language teaching because of transfer effects.
  - O’odham has a high central vowel, IPA /i/. All the writing systems in the U.S. use the symbol e to represent this. The language uses l to represent a flap (IPA /l/), another possible point of confusion.
  - Muscogee (Creek) uses r for the voiceless lateral (IPA /Ʉ/).

- Can hinder learner awareness of the unique sounds of the endangered language because of literacy in the majority language.
Parameters: Socio-political

- need for community involvement in o.d. process
- acceptability of orthography (locally and in larger context)
- relationship with dominant language – use of conventions
  - symbolic issues (±differentiation)
  - literacy transference issues (±learnability)
- standardization / variation
Parameters: Techno-linguistic

Representation: what to represent, how to represent it, where to represent it

- choice of script, symbols, conventions
- identification of phonemes/allophonic processes/other phonological processes/morphological processes
- evaluation of functional loads
- evaluation of resources where information can be registered, if not in the orthography itself (practical grammar, dictionary, etc.)
Parameters: Psycho-cognitive

‘Learnability’ (Orthographic depth)

- **shallow O:** (close to pronunciation)
  - + learnability for beginners and non-(fluent) speakers
  - - readability (may obscure morpheme identities)
  - harder to standardize dialect variation

- **deep O:** (preserves graphic ID of meaningful elements)
  - - learnability for beginners and non-(fluent) speakers
  - + readability
  - easier to standardize dialect variation
Summary: the goals
End goals of the ‘midwife’ approach

For the LC:
- a practical orthography that is a comfortable tool for both writers and readers
- a new means of expression developed collectively, with their own input
- empowerment, incorporating skills and resources for future decision-making

For the Linguist:
- an experience where some of the ‘heat’ is taken off, but where creativity is crucial
- a richer analysis, the result of L’s technical knowledge + LC input
Some references

Good starting places:


A more exhaustive list can be obtained from our CoLang 2012 website:

http://idrh.ku.edu/colang2012/workshops/
About us

- Colleen Fitzgerald (cmfitz@uta.edu) has been working on Tohono O'odham for nearly 2 decades, and on Native languages of Oklahoma since 2009, in the areas of phonology and revitalization.

- Gwen Hyslop (gwendolyn.hyslop@anu.edu.au) has been working on languages in Bhutan since 2006, including collaborative (language community and government) development of ’Ucen orthographies for Bhutan’s endangered languages.

- Keren Rice (rice@chass.utoronto.ca) has been working on Dene languages in northern Canada since the 1970’s, and served on an orthography standardization committee in the 1980’s.

- Kris Stenzel (kris.stenzel@gmail.com) has been working on the description and documentation of Kotiria and Wa’ ikhana, two Eastern Tukanoan languages spoken in northwestern amazonia (Brazil/Colombia) since 2000.

- We welcome your feedback/comments/questions!