Introduction

What information do you need?
Where do you get it?

An Orthography Statement
- Written by the linguist
- Describes the orthography
- Difficult to track down
  - Front of dictionary

Different scripts, different hot issues

Encoding

What is the character set?
- Not just ‘the alphabet’
- Vowels, tone marks, punctuation, numbers

Do they behave like other Unicode characters?
- Is casing the same?
- Sequence or unit for á?
Rendering: 1
What is the basic glyph set
  • Variant forms of characters
  • Decide in conjunction with font designer

How do the characters behave?
  • What combines with what how?

Rendering: 2
Line breaking
  • Special characters? Space, punctuation, etc.
  • Hyphenation issues
  • Using a syllable statement
    • Usually poorly written – just a phonological list
    • Needs to be written in terms of orthography

Analysis
Sort order
  • Any special behaviours?
  • Any special re-ordering?
  • Secondary ordering issues
Keyboarding
Not part of an orthography statement
Made to order

Conversion
Relationship between writing systems
• Through the phonology
• Under specification in the source script
• Over specification in the target script
Multi-script encodings
• Which base script

The Orthographic Block
Different bases
• Orthographic Syllable
• Cluster
• Word
Different Uses
• Cursor placement and unity – cluster
• Re-ordering – syllable
• Line breaking – word or syllable or cluster
Difficult to resolve
Script Specific Issues

Roman
  • Hyphenation
  • Diacritic placement (Vietnamese vs IPA)

Arabic/Hebrew
  • Character forms
  • Bi-directional properties
  • Diacritic collision

Indic
  • Orthographic syllable
  • Diacritic positioning and interaction

Contact Information

For More Information Contact:
  • Non-Roman Script Initiative
  SIL International
  7500 West Camp Wisdom Rd.
  Dallas, TX 75236
  (972) 708-7440
  fonts@sil.org

This presentation is Copyright ©2001 SIL International,
and may not be reproduced without permission