



Keyboards for Indic Languages

Gihan Dias

*University of Moratuwa
Sri Lanka*



Keyboards

- Remains by far the most common text input method
- Easy to learn and use
 - “You can use it right away. Just search for the letter you want and push the key.”
 - not quite... How do you type Ø or å?
- Standardised for most European languages
 - applications “know” how text is entered
- Selected using *locale*



Keyboards (cont.)

- **Based on manual typewriters**
 - have inherited many legacy features
 - difficult to change – e.g. QWERTY
- **Keyboard layout**
 - the letter assigned to of each key (shown on the keycap)
- **Key sequence**
 - the sequence of keys which generate a given output



Keyboards should be...

- **Intuitive and easy to learn**
 - should follow a user's internal model of text
 - should follow “do what I mean” principle
- **Efficient and easy to use**
 - minimise keystrokes
 - common letters on “strong” keys
- **Complete**
 - all letters and symbols should be typeable
- **Otherwise users will get discouraged**



Need for Standard Keyboards

- If no standard keyboard ...
- Users and developers must deal with multiple keyboards
 - must be addressed in manuals, help files, etc.
 - users are confused



Indic Scripts

- Used to write the languages of South and South-East Asia
- Are classified as **abugidas**
- A consonant with a specified vowel is represented by a single symbol
- A consonant without a vowel (pure consonant) or with another vowel shown by a modified consonant symbol
- A leading vowel shown as independent symbol



Example

- In Tamil, the consonant **p** followed by the vowel **a** is represented by **பா**.
- The pure consonant **p** is shown by adding a dot (pulli) above the base symbol - **பு**
- **p** with the vowel **i** is represented by adding a modifier to the base symbol:
p + i = பி
- The vowel **i** at the beginning of a word is represented by **இ**



Example (cont.)

- Modifiers may appear on various sides of the base symbol, e.g.:
 - $p + ai = \text{൬} \sqcup$ - Before,
 - $p + aa = \sqcup \pi$ - After,
 - $p + u = \sqcup$ - Below
- Some modifiers may be on both sides of the base, e.g. $p + au = \text{൬} \sqcup \pi$.
- Sometimes the base letter changes:
 - $k + a = \text{ඝ}$; $k + uu = \text{ඝූ}$



Consonant clusters

- In some scripts, e.g., Devanagari, a pure consonant (i.e., without a vowel) combines with the following consonant to form a cluster.
- in Devanagari:
 $sa = स ; s = स् ; va = व ; s + va = स्व$
- Some conjuncts are different from either of the constituents - e.g.:
 $k + ssa = क् + ष = क्ष$



Keyboards for Indic Scripts

- Typewriter
- Consonant-Vowel
- Romanised
- Transliteration



Typewriter Keyboards

- Based on manual typewriters
- Each letter is entered using one or more keys which produce **parts** of the letter
 - carriage does not shift when some symbols (dead keys) are typed
- Symbols are based on **shape**, not linguistics
- Output is an **approximation** of the “correct” shape



Consonant-Vowel Keyboards

- **Consonant typed first, then associated vowel**
 - typing is linguistic
 - may be different from **visual** order
 - may be different from **writing** order
 - corresponds to **pronunciation**
- e.g. In Sinhala, **කෙ** is typed as **ක + ඌ**



Inscript Keyboards

- Standardised by the Indian Govt.
- **Similar layouts** for all Indian scripts
 - a person can type even in an unfamiliar script if he knows the Inscript layout
- Follow consonant-vowel model
- Vowels on the left, consonants on the right



The Malayalam Inscript Keyboard





Romanised Keyboards

- The output of a key is based on the English letter printed on it
 - convenient for those with only English keyboards
- e.g. On a Sinhala romanised keyboard, the key **p** produces the letter **ආ** (pa)
- Generally has one-to-one correspondence between keys and display symbols
- Problem: English and Indic scripts do not map one-to-one



Transliteration Keyboards

- **An approximation of the text is typed in English characters**
 - each Indic letter may use one or more keys
 - converted to correct output by keyboard driver



Romanised and Transliteration Keyboards

- Romanised keyboards map a key(s) to a display symbol
- Transliteration keyboards convert key sequences into character(s)
- e.g. The Sinhala word චන්ඞ
- Typed

c	n	z	n
---	---	---	---

 (ච ඞ ඞ) on a romanised keyboard
- Typed

c	h	a	n	n	a
---	---	---	---	---	---

 on a transliteration keyboard - **cha** = ච ; **n** = ඞ ; **na** = ඞ



Standardising the Sinhala and Tamil Keyboards



The Sinhala Script

- Used by 15 million people in Sri Lanka
- South-Indic Script
- Letters are not joined together
- Uses a mark (*al-lakuna*) above base symbol to indicate a pure consonant
- Vowel modifiers may occur on any side of the base, and some modifiers are split to two sides



Existing Sinhala Keyboards

- **Wijesekera-based keyboard layouts**
 - based on the typewriter keyboard
 - one key per visual symbol
- **“Phonetic” layouts**
 - called “Romanised” in other languages
 - popular among casual users
- **Transliteration schemes**
 - not popular
- **Consonant-vowel sequence keyboards.**
 - not used



Development of the Standard Sinhala Keyboard

- The Inscript-based consonant-vowel keyboard did not get user support
 - users did not accept the concept
 - not intuitive
- Transliteration schemes were considered too complicated and ambiguous
- Need for phonetic (romanised) keyboard identified, but left for a later date
- Decided to standardise the Wijesekera keyboard



Standardisation Objectives

- **Compatibility with the Wijesekera typewriter keyboard**
- **Compatibility with the English (US-ASCII) keyboard**
 - **as most users are bil-lingual**



Design Principles

- **Common letters as on typewriter keyboard**
- **1st-row numbers and symbols as in US-ASCII keyboard**
- **One key for each modifier**
 - the typewriter keyboard has separate keys for each different form of each modifier
- **No “half letters” on the keyboard**
 - Conjuncts typed using *join* key
- **Typing sequence same as writing sequence**



The Standard Sinhala Typewriter Keyboard

- Most letters retained on same key as typewriter
- Some letters typed using right-alt (alt-gr) key (as in European keyboards)
- Keys assigned to common symbols yansaya- ි ʒ rakaransaya- ි ʒ etc.
- Punctuation mostly as in typewriter or US-ASCII



Evaluation of Sinhala Keyboard

- Accepted by typists
- Several brands of physical keyboards manufactured
- Methods of producing *sangyaka* letters and conjuncts are not intuitive
 - need more awareness and training
- Should have placed common punctuation (comma, period) on same key as US-ASCII



The Tamil Script

- A South-Indic Script
- Separated letters
- Explicit “pulli” for pure consonants
- Much **smaller number of letters** than other Indic scripts
- Includes some *Grantha* letters for representing non-Tamil words



Standardisation of Tamil Keyboard

- **Renganathan**
 - typewriter-based keyboard
 - very popular in Sri Lanka
- **Inscript-based keyboard**
 - Standardised by Indian Govt.
 - not optimised for Tamil
 - not accepted by Tamil users
- **Romanised keyboards**
 - widely used



Tamil 99 Keyboard

- Introduced at the TamilNet conference in 1999
- Adopted by the Gov. of Tamil Nadu
- A consonant-vowel keyboard
 - same key used for independent vowels and vowel modifiers
 - All *Tamil* letters are on unshifted keys
- Adopted by ICT Agency of Sri Lanka in 2004



Evaluation of Tamil 99 Keyboard

- **Endorsed by users and successfully piloted**
- **Did not gain acceptance**
 - lack of awareness and training
- **Reported shortcomings**
 - Text is typed differently from how it is written.
 - Key placements are totally different from the typewriter layout
 - Lack of vowel symbols on keyboard is disconcerting



Sri Lanka Tamil Keyboard - 2007

- ICTA held two consultations in 2006
- Consensus that Tamil 99 keyboard is not acceptable
- Users preferred a Renganathan-based keyboard as it is more familiar
- Requirements:
 - be close to the Renganathan / Bamini layout
 - be uniform and logical and
 - be compatible with the English keyboard.

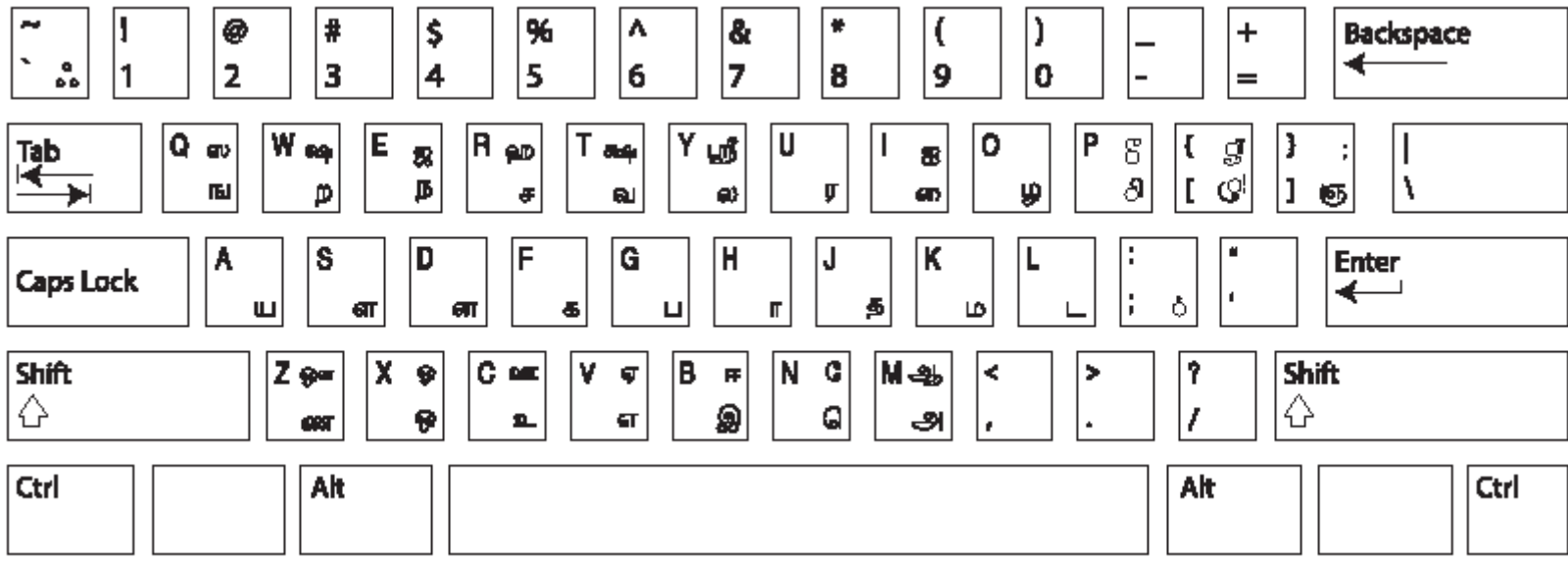


Design of Tamil Keyboard

- Studied over 10 variants
- Keys common to variants were retained
- 1st-row keys and common punctuation as in US-ASCII keyboard
- All Tamil consonants on unshifted keys, long vowels on shifted keys
- “Grantha” letters on shifted keys
- One key per vowel modifier – irrespective of shape



SL Tamil Keyboard 2007





Conclusion

- Many types of Indic keyboards exist
- User preference is overwhelmingly for typewriter or romanised types
 - they don't care about linguistics
- Attempt to introduce Tamil99 in Sri Lanka **failed**
- Need **robust** and **out-of-the-box** support for **typewriter-style keyboards**