Dasher in Korean

David J.C. MacKay
Cavendish Laboratory
J J Thomson Avenue
Cambridge
CB3 0HE
mackay@mrao.cam.ac.uk

September 12, 2006 — Draft 1.1

This technical report describes how Dasher works in Korean, with an emphasis on the default colour scheme.

1 Dasher

Dasher is an information-efficient text-entry interface, driven by natural continuous pointing gestures [2, 3, 4]. Dasher is a competitive text-entry system wherever a full-size keyboard cannot be used - for example,

- on a palmtop computer;
- on a wearable computer;
- when operating a computer one-handed, by joystick, touchscreen, trackball, or mouse;
- when operating a computer with zero hands (i.e., by head-mouse or by eyetracker).

The eyetracking version of Dasher allows an experienced user to write text in English as fast as normal handwriting – 29 words per minute; using a mouse, experienced users can write at 39 words per minute.

Dasher can be used to write efficiently in any language.

Dasher is **free software**. It's distributed under the same license as GNU/Linux, the GPL.

For further explanation of Dasher, please see the Dasher Manual or the built-in Help.

2 Korean

Written Korean (Hangul) is a beautifully logical phonemic language. [1] Each **hangul** character represents a single syllable, and is made up of two or three letters called **jamo**. There are **24** jamo letters: 14 consonants, and 10 vowels. The first letter in a hangul character is one of the 14 consonants. The second letter is one of the vowels. If there is a third letter, it is one of the consonants. Some five of the consonants exist in 'double' form.

In **Unicode**, any of the Hangul characters can be created from **67** 'combining' jamo characters. Why 67 – why not 24? Well, there are 5 double consonants, which get their own jamo character in unicode; and every consonant is available in an 'initial' form and a 'final' form. Plus there are about 9 extra final consonants that represent consonant-pairs like 'lm', 'nj', and 'lh'. Plus instead of having just 10 vowels, there are 21 vowels, including various diphthongs. Figure 1 shows the 67 combining jamo characters as rendered by Dasher.

There are two ways to use Dasher in Korean. The recommended way is to write using these 67 'combining' jamo characters. It's also possible to write directly in the Hangul alphabet, but this is not recommended as the alphabet consists of about 11 000 characters, and Dasher's performance with such a large alphabet is likely to be poor.

In Dasher's default 'European/Asian' colour

[1] See http://en.wikipedia.org/wiki/Hangul

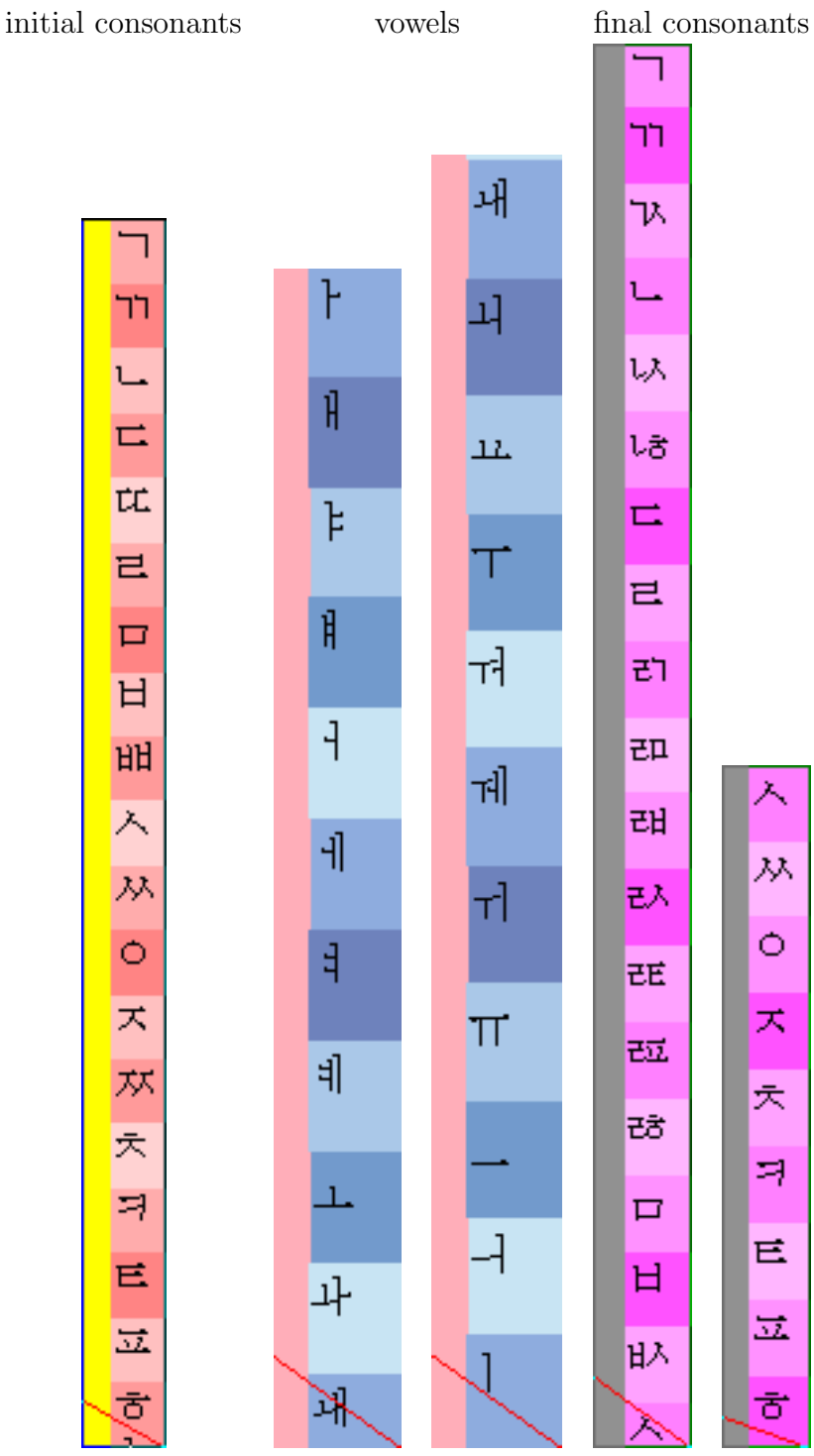


Figure 1: The 67 jamo characters in unicode

scheme, Korean writing proceeds as follows:

1. every new hangul character is initiated by entering a **yellow** box; this box contains all the **initial** consonants (*choseong*).

2. The initial consonants are coloured sand/pink. After the initial consonant, you must enter a vowel (*jungseong*); the vowels are in **blue** boxes.

3. After the vowel, you may either start a new hangul syllable by entering the yellow box, or add a final consonant (*jongseong*) – found in the **grey** box; final consonants are contained in **magenta**-coloured boxes.

4. Another option after the vowel is entered, or after a final consonant is entered, is to enter a non-jamo character such as a space character (shown by the box in a white square).

2.1 Practical details

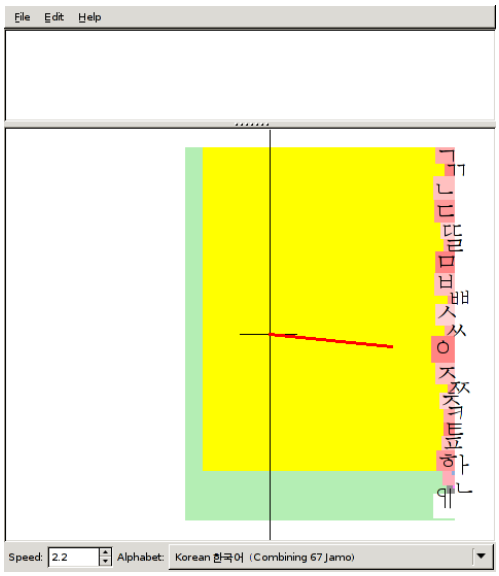
It's recommended that the Dasher option 'display box outlines' is switched **off** when using Dasher in Korean. (Because the group boundaries between initial consonants and vowels might otherwise be found distracting.)

You may also find that the user experience is improved by turning the Language→Smoothing parameter down from its default value of 5.0.

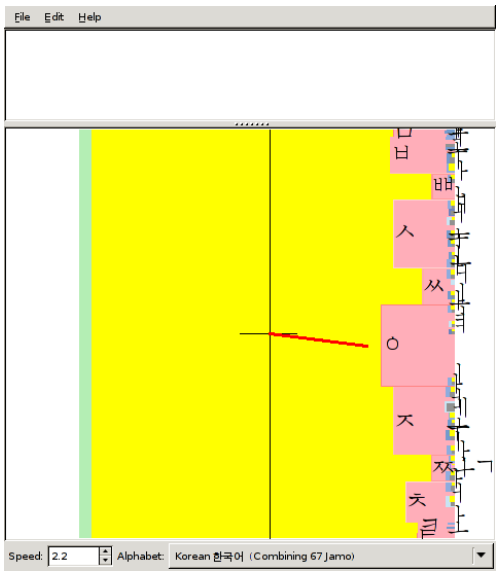
3 Screenshots

Korean works beautifully on linux systems, thanks to the excellent Pango rendering engine; the following screenshots are made on an Ubuntu linux machine. As of 2005, Microsoft Windows did not provide such good font support for Korean, so Phil Cowans had to add a special workaround to get Korean to look good on Windows.

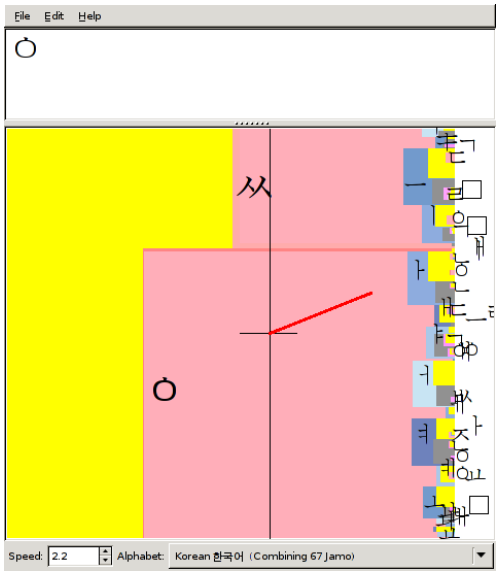
Here are screenshots showing writing 'hello' in Korean.



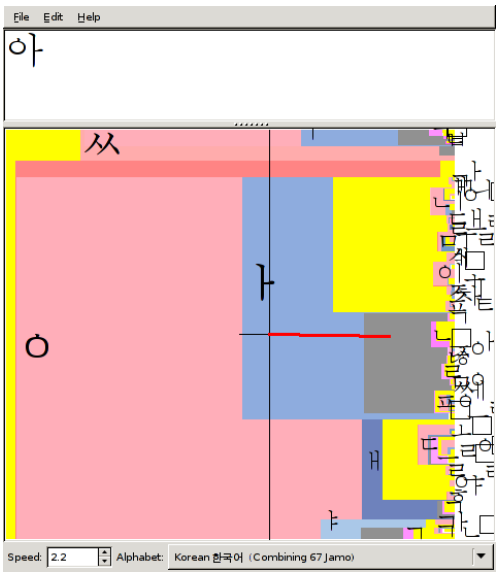
Initial Dasher view



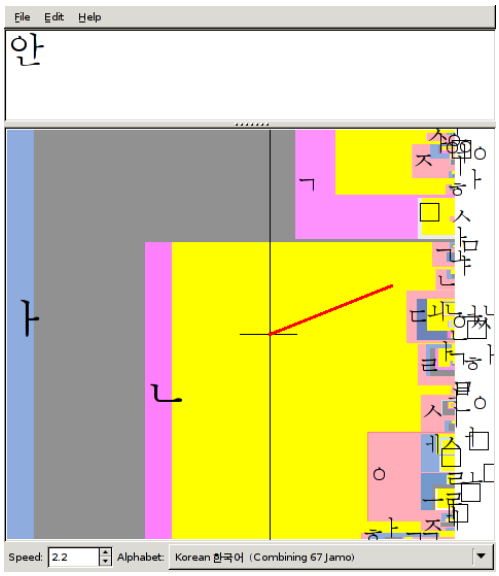
Entering the first consonant



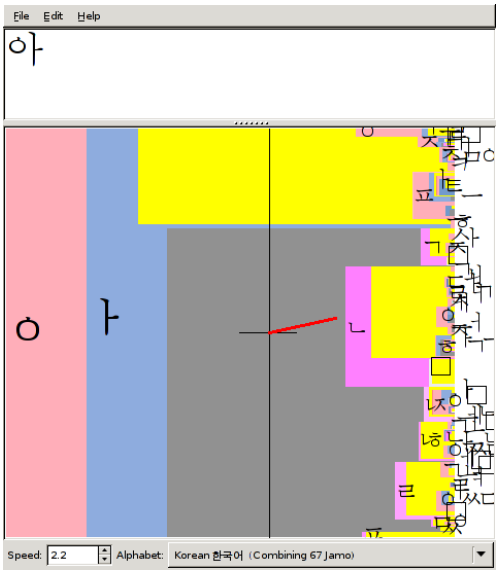
Choosing the vowel



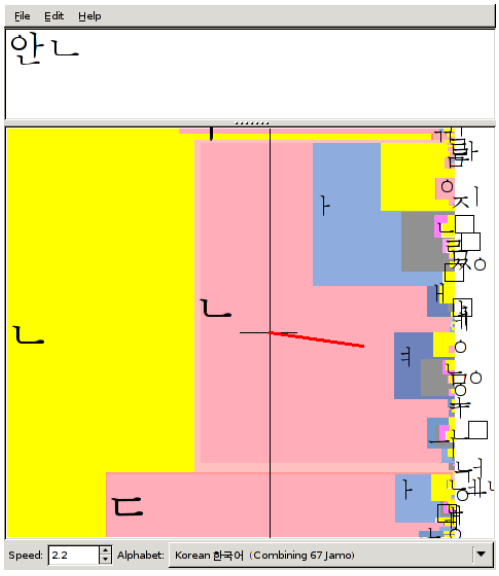
As we enter the vowel, we see two main options: yellow box – start a new syllable; grey box – add a final consonant to this syllable. Here, we choose the grey box.



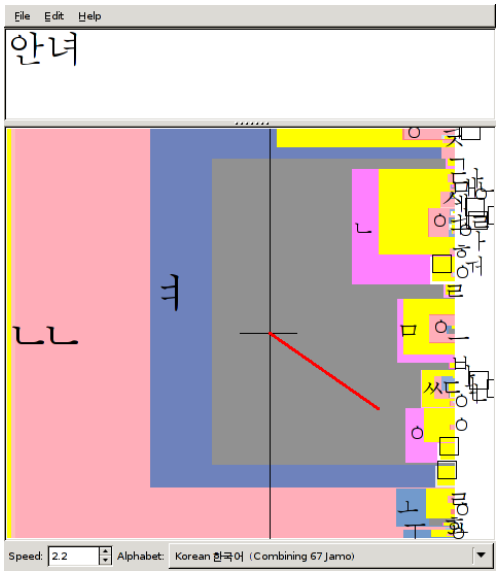
We start a new syllable with another 'n'.



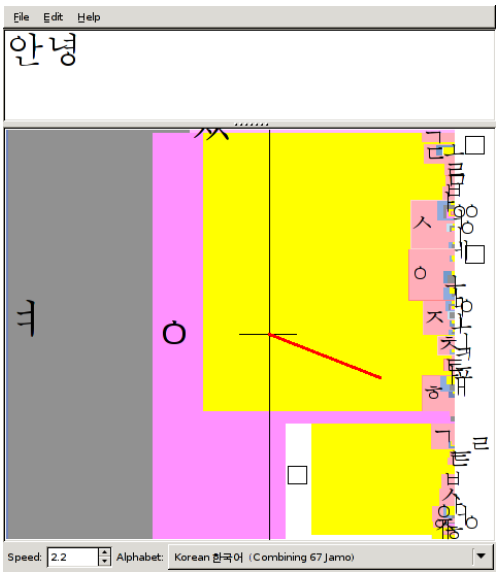
When we enter the final consonant 'n', the two probable options inside its magenta box are: yellow box (start a new syllable) and white box (enter a space character).



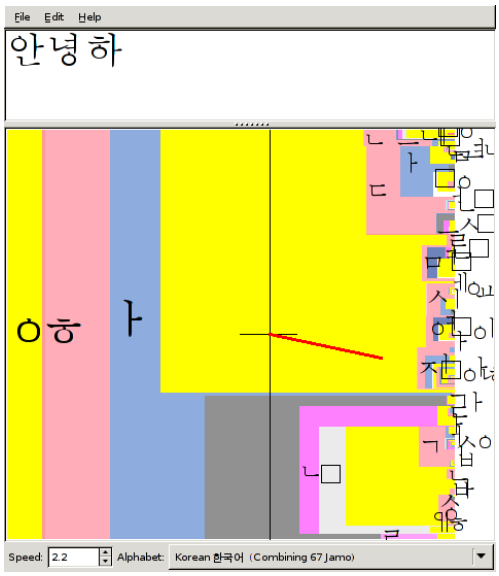
Choose the next vowel



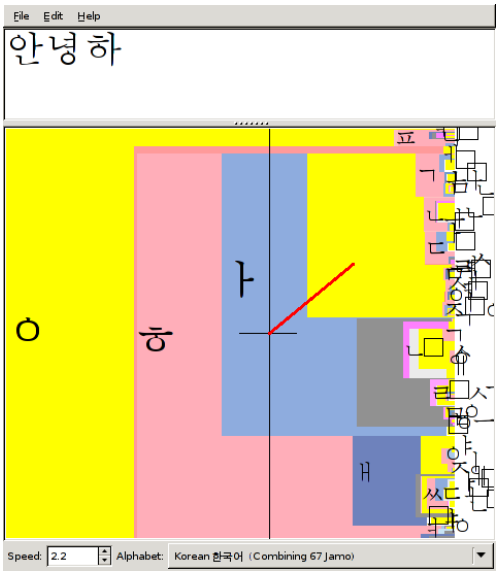
Choose the final consonant



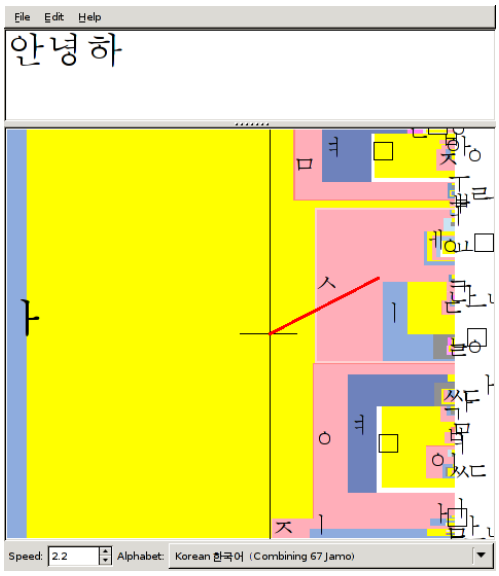
Enter the yellow box to start the next syllable



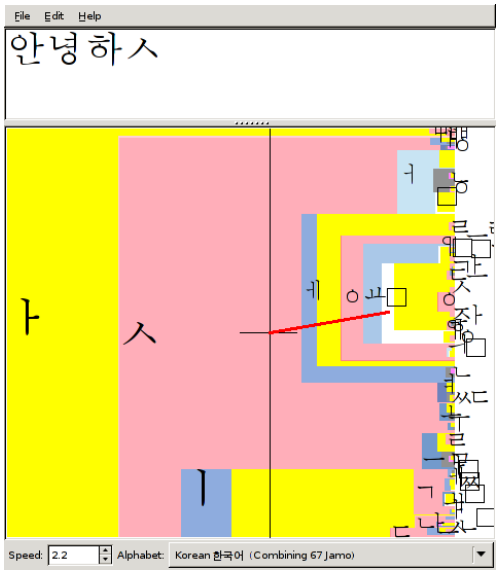
As we select the leading consonant 's'...



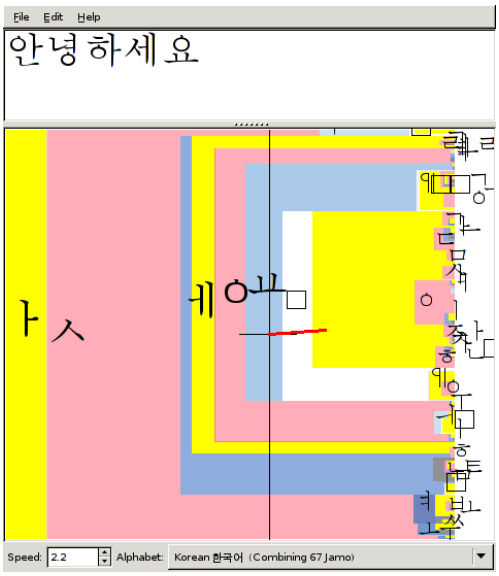
Grab the consonant and vowel; again we can either start the next syllable by entering the yellow box or add a final consonant by going in the grey box. This time we go in the yellow box, as the syllable is complete.



...and the next vowel...



the language model does a good job of predicting the entire next syllable.



As this syllable (which ends the word we are writing) is completed, the white-space characters (space and newline) are both very probable.

4 Training texts

We gratefully acknowledge the help of KAIST BOLA who provided a 258-kbyte corpus [1], which was used to make the screen-shots. This corpus is available for research purposes but may not be freely distributed.

At present we have very little public-domain training text. **We would very much like to receive corpuses of freely-distributable Korean text.** We are aware that there are several Korean writing styles, and ideally we would like to have a separate training text for each style. Please send corpora to David MacKay. UTF8 is the preferred plain text format.

References

[1] Key-Sun Choi. Kaist language resources, 2001. Ministry of Science & Technology Software Project results, 1995- 2000 (kibs.kaist.ac.kr).

[2] D. J. Ward, A. F. Blackwell, and D. J. C. MacKay. Dasher – A data entry interface using continuous gestures and language models. In *Proceedings of User Interface Software and Technology 2000*, pages 129–137, 2000.

[3] D. J. Ward, A. F. Blackwell, and D. J. C. MacKay. Dasher – A data entry interface using continuous gestures and language models. *Human-Computer Interaction*, 17(2-3):199–228, 2002.

[4] D. J. Ward and D. J. C. MacKay. Fast hands-free writing by gaze direction. *Nature*, 418(6900):838, 2002.