PAN
Localization

Survey of Language Computing in Asia
2005

Sarmad Hussain
Nadir Durrani
Sana Gul

Center for Research in Urdu Language Processing
National University of Computer and Emerging Sciences

IDRC ✳ CRDI
Canadä

www.nu.edu.pk   www.idrc.ca

Korean

Korean is the national language of Korea spoken by about 78 million people [1], mostly in North and South Korea. Scholars conflict on the development of Korean language. Few believe that Korean is related to Japanese while others disagree and argue its independent development.

Language Isolate
KOREAN

Figure 1: Language Family Tree for Korean [1]

Korean uses two writing systems known as Hangul and Hanja. Hangul is the syllabic writing system used in Korea. Hanja refers to the Chinese ideograph characters. Earlier, since about 5 AD, Korean has been written in Chinese script. The Korean alphabet started during the reign of King Sejong (1418-1450). The alphabet was originally called Hunmin jeongeum, or "The correct sounds for the instruction of the people", but has also been known as Eonmeun (vulgar script) and Gukmeun (national writing). The modern name, Hangeul, was given by a Korean linguist Ju Si-gyeong (1876-1914). The academic papers and official documents tend to be written in a mixture of Hangul and Hanja [2].

Character Set and Encoding

The primary character set encoding is KSC 5601 standard for Hangul and Hanja. This includes 17,100 characters [3]. This was augmented by KSC 5657. However, these standards have been superseded by KSC 5700, which contains the same set of characters as in Unicode. Other Korean standards include ANSI Z39.64 (also called as REAC), EACC, CCCII, CP949, EUC-KR, GB12052, ISO2022-KR, KS C 5636 and MOJIKYO [4]. Also see [6].

One of the encoding that emerged in 1992 is Johab standard. Johab is a 16 bit-code having first bit denoting the Johab encoding, the following 5 bits for initial consonant, next 5 for vowels and the last 5 for final consonant. Following figure shows some relative positions of the symbols. This standard has now been replaced by other standards discussed above [5].

Figure 2: Johab Encoding for Korean [5]

Fonts and Rendering

Many Korean fonts are available for most of the computing platforms, which are based on different encodings discussed in the previous section. Bating, Dotum, Gulim, Gungsuh and Arial Unicode are some fonts which work for Microsoft platform. Hanyang font is available for Linux with commercial license terms, Baekmuk font is supplied under BSD license, while Un-font is available under GPL license.

Keyboard and Input Method Engine

KSC 5715 is the standard keyboard layout [6]. However, there are other phonetic layouts also available (e.g. see [7]). Dubeolsik is also a common Hangul keyboard layout in use in South Korea [8]. In addition, various platforms also provide support for entering Korean, which require complex input methods in addition to the keyboard.

Microsoft Platform

Microsoft Windows supports the Korean standard and alternative keyboard layouts. The Korean IME enables users to input Korean (Hangul) and Chinese characters (Hanja). Hangul is entered by Jamos, which are 24 basic elements and combination of elements, on the standard 101 keyboard. By combining these Jamos, all 11,172 Hangul character combinations are produced [9]. Figure 3 shows MS keyboard layout for Korean language.


Figure 3 Korean Keyboard Layout [9]

The following figure explains how the Korean characters are entered into an application.

Action	Result
Type the letter "ㅇ" that corresponds to English key "d."	ㅇ
Now type the letter "ㅓ" that corresponds to the English key "j." The character "ㅇ" is replaced with the combined syllable "어."	어
Type the letter "ㄴ" that corresponds to English key "s." The character combination is now replaced with "언", which finishes the first of the two hanguls needed to represent language	언
To create the second hangul, type the letter "ㅇ" again.	언 ㅇ
Now type the letter "ㅓ" again followed by a space to finish the word.	언 어

Figure 4: Process of Input of Korean Characters through MS Korean Keyboard [9]

Once Hangul has been formed, the user can then press a Hanja key that will allow Hangul to be transformed into corresponding Chinese characters. Microsoft also provides IME Pad that allows the user to input Hangul via soft keyboard, shown in Figure 5. In addition, Microsoft also provides a handwriting recognition based interface to input Korean, also shown in Figure 5.
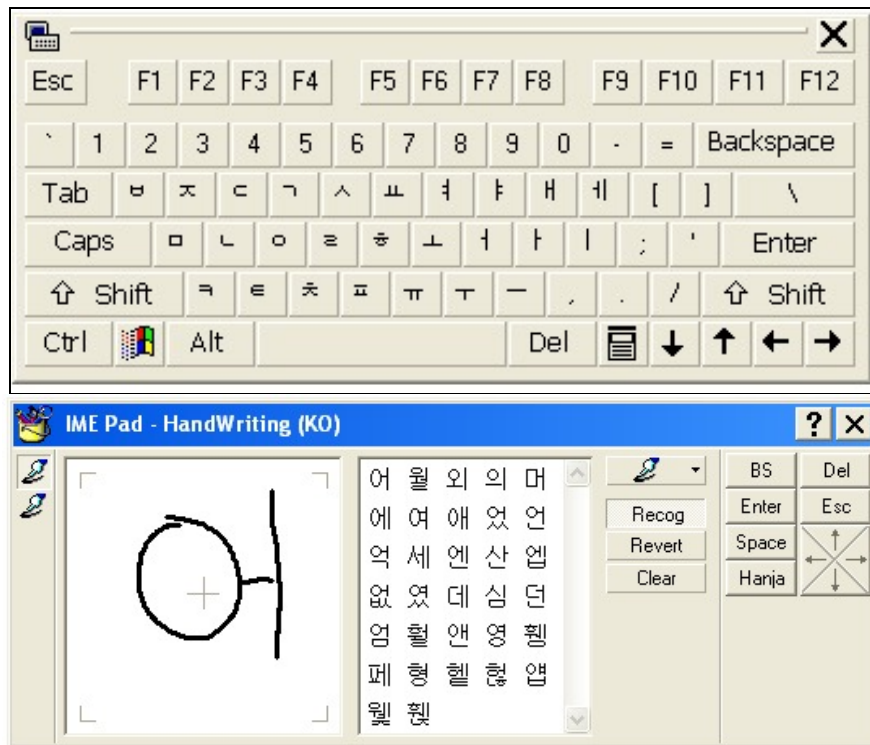
Figure 5: Soft Pad for Korean and Handwriting Recognition Based Input Method by Microsoft [7]

Linux Platform

Korean input methods in Linux support both the standard and alternative keyboards. The standard keyboard is widely used. "Ami" is an input method used for Korean [12].

Collation

Korean collation is supported in all MS applications, Linux platforms, and also defined in Glibc. There is some work on defining a standard collation of Korean. See [10] for Korean collation methodology. Collation rules are also defined in the locale definition of Korean (see below).

Locale

Standard Locale for Korean has been defined in IBM ICU. The code is ko_KR (for South Korea). This locale definition constitutes localized settings for date, time, days, months, currency symbols, number format and collation [10]. CLDR 1.3 also includes a locale for Korean.

Microsoft Platform

Microsoft Windows XP provides support for Korean locale. If system locale is switched to Korean, changes may be monitored in the date, time, and currency symbols within all application of Microsoft Windows. The following figure shows Korean locale on MS platform.
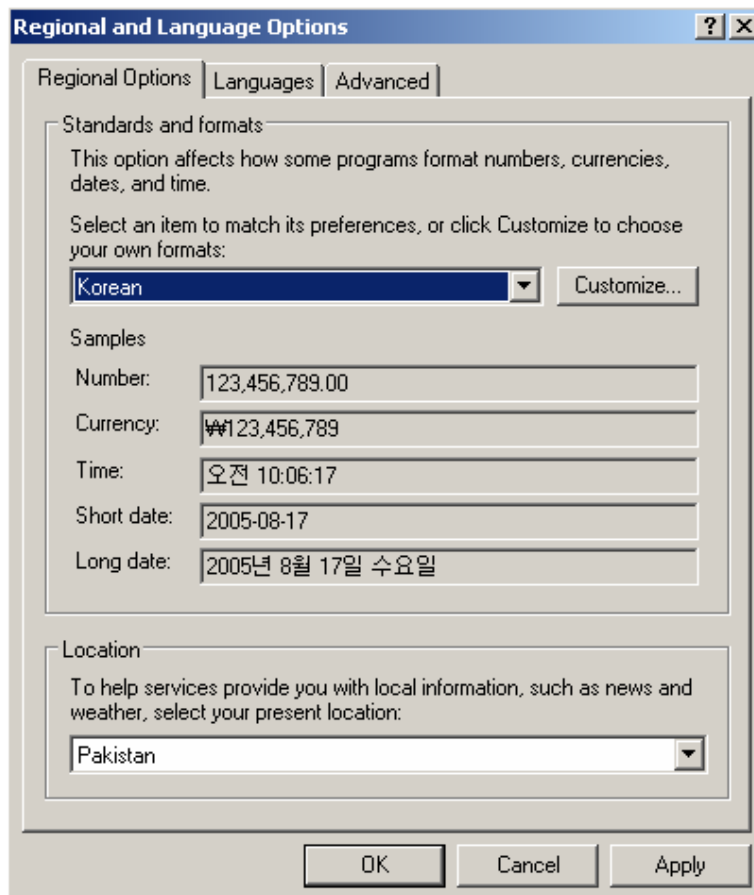
Figure 6: Korean Locale on Microsoft Platform

Linux Platform

Korean locale is also available on Linux platforms, e.g. see [12].

Interface Terminology Translation

Most of famous Linux distributions such as RedHat Linux, Fedora Core, Debian GNU/Linux, Mandrake and Novell Linux (or Suse Linux) support Korean. In addition, there are also Linux distributions by Korean vendors, which include Hansoft Linux, Hancom Linux, and Wow Linux.

Microsoft Platform

A complete version of MS Windows is shipped in Korean. The Korean Windows has translated interface, menus and dialogue boxes.

Linux Platform

Terminology translation for localization of GNOME is managed by GNOME Korean user community [13]. 94% for GNOME 2.12 has been translated for Korean language [14] while 14% of KDE has been done [15]. Mozilla terminology is maintained by volunteers under the Hangul Mozilla project [16]. On the Linux platform the Open Office interface and help files have also been localized [17]. There is a significant push by Korean government to develop open source platforms and many other initiatives are under progress [18].

Status of Advanced Applications

Significant language resources are available for Korean language, including lexicons, corpora (e.g. [19]), spell checkers and grammar checkers. Korean grammar checker has been embedded in Microsoft products (e.g. Word) [20] and Hangul (Korean-made word processing program). Word breaking utilities, stemmers and morphological parsers are also available, e.g. [21].

Korean text-to-speech systems [22, 23], and speech recognition systems [24, 25] are also commercially available. OCR software and handwriting recognition systems for Korean are widely available for a variety of platforms [26]. There are also Korean machine translation systems available for multiple languages, e.g. [27, 28].

References

[1] http://www.ethnologue.com/14/show_language.asp?code=KKN
[2] http://www.omniglot.com/writing/korean.htm
[3] http://www.ascendercorp.com/cjk.html
[4] http://www.jbrowse.com/text/charsets.html
[5] http://homepages.cwi.nl/~dik/english/codes/stand.html
[6] Hyun-Il, K. "Trends and Plans for the IT Standardization in Korea," in the *Proceedings of 5th AFSIT,* Tokyo, Japan, 1991. Available at http://www.cicc.or.jp/english/hyoujyunka/af05/5-07.html
[7] http://www.freepatentsonline.com/6462678.html
[8] http://www.absoluteastronomy.com/encyclopedia/k/ke/keyboard_layout.htm
[9] http://www.microsoft.com/globaldev/handson/user/IME_Paper.mspx
[10] http://www.open-std.org/jtc1/sc22/wg20/docs/n1037-Hangul%20Collation%20Requirements.htm
[11] http://www-950.ibm.com/software/globalization/icu/demo/locales/?d_=en&_=ko
[12] http://www.freshports.org/korean/
[13] http://www.GNOME.or.kr/
[14] http://l10n-status.GNOME.org/GNOME-2.12/ko/index.html
[15] http://i18n.KDE.org/stats/gui/stable/ko/
[16] http://www.mozilla.or.kr/
[17] http://projects.openoffice.org/native-lang.html
[18] http://www.ima.umn.edu/~klee/kr.linux.html
[19] www.ldc.upenn.edu
[20] http://www.translation.net/microsoft_proof_2003.html

[21] http://www.teragram.com/oem/asian_lang.htm#korean
[22] http://www.microsoft.com/MSAGENT/downloads/user.asp
[23] http://www.voicesignal.com/news/press/release_12_03_03.html
[24] http://www.mobileburn.com/news.jsp?Id=912&source=BROWSER
[25] http://www.asahi-kasei.co.jp/vorero/en/news/
[26] http://www.worldlanguage.com/Products/Korean/OCR/Page1.htm
[27] http://world.altavista.com/
[28] http://www.worldlingo.com/en/products_services/worldlingo_translator.html