A Proposed Character Set for the GE 636   May 25, 1965

by F. J. Corbató and R. Morris

The present memo is an outgrowth of extensive discussion at MIT
and Bell concerning a suitable character set for the GE 636 and super-
cedes MAC memo M223-1, CC-250-1.

It is recommended that the proposed revised ASCII character set
(Comm. ACM, April 1965, p. 207) be adopted as a standard. This code,
which is a 7-bit code, (mapped into a 9-bit field on the GE 636),
contains 128 characters. Of these 128 characters, 95 (including space)
have associated standard graphics, one is an escape character (with
an optional graphic), and the remaining 32 are control characters.
All devices which are not able to create these codes or the corresponding
graphics, will require conventions (hopefully uniform and standard)
which are suitable to represent the full character set. It is emphasized
that there are no character subsets meaningfully defined in revised
ASCII (c.f. Appendix B3, p.210).

For the reader's convenience, a brief comparison is given of revised
ASCII graphics compared with those of current ASCII:

<u>deletions</u>:   \ (reverse slash)

                   ↑ (left arrow)

                   ← (left arrow)

                   ' (apostrophe)

<u>additions</u>:   (escape character - no standard graphic)

                   | (vertical bar)

                   _ (underline)

                   ´ (accent acute)

                   ` (accent grave)

                   ∧ (circumflex)

                   ≈ (tilde)

<u>additions con't:</u>   ¬ (logical negation)

{ (left brace)

} (right brace)

a-z (lower case alphabet)

In review the complete graphics of the revised ASCII set are:

A-Z, a-z, 0-9,(space)

! " # $ % & ´ ( ) * + , - . / : ; < = > ? ` [~] ∧ _ @ { ¬ } |

Since no current devices contain all of the above graphics,

escape conventions will now be proposed for some of the devices likely

to be used.

<u>1050 Consoles</u>

A special ~~xixxtix~~ Selectric type ball has been ordered by Project

MAC for the IBM 1050. This 88-graphic ball will have most of the re-
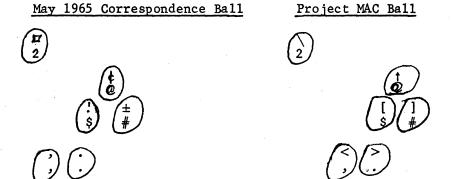
vised ASCII set.

Additional characters on the ball are:

\ (reverse slash) to be used as "escape".

↑ (vertical arrow) to be used as | (vertical bar) if it cannot
            be changed to vertical bar.

' (apostrophe) to be used as ´ (accent acute)

Missing characters are:

! (exclamation) for text typing, may be synthesized from backspace,
            period and apostrophe.

` (accent grave)

∧ (circumflex)

~ (tilde)

¬ (logical negation)

{ (left brace)

} (right brace)

Keyboard changes from the present 1050 are:

<u>May 1965 Correspondence Ball</u>　　　<u>Project MAC Ball</u>

Escape conventions for the 1050:

Missing characters are created by the use of the escape character. Whenever the escape character \ is encountered in a serial left-to-right scan of text, it always means "give the next character another meaning than the usual one". Similarly, \\ means a double escape, etc. In addition, whenever one is inputting or outputting on a console, a local carriage return (i.e. one which occurs physically at the console but is not included in the computer-stored character string) is designated by \ (carriage return).

In the case of text, where one wants the exclamation point in a graphically correct form, it is necessary to use three characters to create the symbol, namely, period, backspace, and apostrophe as has long been done on many typewriters.

A summary of the 1050 escape codes follows:

    \. for ! (exclamation)

    \C for ∧ (circumflex)

    \' for ` (accent grave)

    \T for ~ (tilde)

    \N for ¬ (logical negation)

    \(for { (left brace)

    \) for } (right brace)

$\backslash 0 d_1 d_2 d_3$ for the octal code $d_1 d_2 d_3$ modulo 128 where d_1, d_2 and d_3

are from 0 to 7.

$\backslash$ (return) for a local carriage return not in the computer-stored

string. xxxx

$\backslash V$ for vertical tab.

$\backslash F$ for form feed.

35KSR and 33KSR Teletypes

In the case of the model 35KSR and 33KSR Teletypes, it is pro-
posed that the TWX console with the M.I.T. modifications which do not
affect keyboard layout or graphics be adopted as a standard. This
console has exactly the current ASCII character set. Thus the characters
↑ ← and ' should be considered stylized versions of | _ and '. (There
is no backspace and the carriage return may not include a line feed
but the computer may by software automatically respond to every return
with a line feed.) In addition to those of the 1050, the following escape codes
are needed:

$\backslash$- designates the character "backspace"

$\backslash U$ designates that all subsequent letters are in upper case
    until a $\backslash L$ sequence is encountered

$\backslash L$ designates that all subsequent letters are in lower case
    until a $\backslash U$ sequence is encountered

$\backslash n$ where n=1,2,...9, designates that the next n characters are
    to be in the opposite case from the previous case. (This
    is valuable in capitalizing the initial character of a sentence
    in ordinary text.)

It is also proposed that the present locally modified 35 KSR
Teletypes at M.I.T. be phased out and replaced by the modified TWX
type. Since this will take time, it can occur in two stages. First,

the present non-standard graphics should be replaced by standard

ones. In particular,

    __ becomes \

    ⊃   "   %

    ∧   "   &

    ∨   "   @

    ~   "   #

Second, the keyboard layout should revert to the standard TWX form.

64-Character Keypunches

Although multiple-punching can be used to generate the full set

of 128 codes, the escape convention will usually be easier. A re-

commended choice of graphics is that of the current 64 character ASCII

(using \ for the escape) with the exception of ↑ ← and 1 being changed

(if possible) to | _ and ´. In this context, punched cards correspond

to typewriter lines. On reading a card, an "end of card" should generate

a "carriage return" code; thus \ in column 80 of an 80 column card

acts as \ (return). It is useful however, to introduce two other

escape characters for input convenience:

    * for "carriage return and skip reading the remainder of this card"

    \+ for "carriage return and keep reading this card"

    \H for horizontal tab

    \V for vertical tab

    \F form feed

60-Character Keypunches

The most commonly available character set on IBM keypunches will

probably be the IBM 60 character set of PI/I (e.g. NPL). This set

diffs from revised ASCII as follows:

additions: ' (apostrophe)

omissions: ´ (accent acute)

                  ` (accent grave)

                  ∧ (circumflex)

                  ~ (tilde)

                  { (left brace)

                  } (right brace)

                  [(left bracket)

                  ] (right bracket)

                  ! (exclamation)

               a-z (lower case letters)

                  \ (escape)

Beyond the escape conventions needed for the 64-character keypunch, there remains a need for conventions for the escape character and the escape codes for [] ´ ` and !. For a choice of escape character, the character ' (apostrophe) is recommended. Then the escape codes can be:

        '/ for ´ (accent acute)

        '. for !

        '(for [

        ') for]

        '/ for ` (accent grave)

48-Character Keypunch

Here the ' (apostrophe) is again suggested as the escape character. Additional escape codes over the above are:

''. for :

'', for ;

''D for &

''P for %

''V for |

''(for [

'') for]

''L for <

''G for >

''S for #

''Q for ?

''U for _

''A for @

''K for "

Telex Terminals

The Telex terminal is the most restricted case known. The only graphics always available on both domestic and international sets are:

A-Z 0-9 (space) -'()/:?,.

Using the apostrophe as the escape, additional needed conventions are:

''R for $

''M for *

''X for +

''E for =

Printers:

For printers which have 64 character sets of current ASCII the teletype conventions will work although computer output will in general be tortuous to read. It is strongly recommended that printers be used with the GE 636 which have the complete revised ASCII set. The IBM 1403 line printers have such capability and represent a possible mechanism to use.

Summary

The above escape conventions have been proposed in such a way as to be uniform over all the devices mentioned. These can be summarized where it is understood that the \\| _ and ´ are sometimes stylized.

	1050	35KSR	64C-KP	60C-KP	48C-KP	Telex
\\. for !						
\\C for ∧						
\\´ for `						
\\T for ~						
\\N for ¬						
\\(for {						
\\) for }						
\\0$d_1 d_2 d_3$ for octal code						
\\(return) for local return						
\\V for vertical tab						
\\F for form feed						
\\- backspace						
\\U upper case letters						

	35KSR	64C-KP	60C-KP	48C-KP	Telex
\L Lower case letters					
\n (n=1,...,9) for reverse letter case for n characters					
\H horizontal tab					
* return and skip rest of card					
\+ return and continue card					
\/ ´ (accent acute)					
\\/ ` (accent grave)					
\\([					
\\)]					
\\. for :					
\\, for ;					
\\A for ₤					
\\P for %					
\\V for						
\\(for [					
\\) for]					
\\L for <					
\\G for >					
\\S for #					
\\Q for ?					
\\U for _					
\\A for @					
\\K for "					
\\R for $					
\\M for *					
\\X for +					
\\E for =					

Of the above, the 1050 requires the fewest uses of escape; for practical convenience, the seven characters ! ∧` ∼ ¬{} should be avoided if possible in initial software.

Design Notebook

Appendix I, Supplement 1

July 9, 1965

Amendment to the Character Set for the GE 636

by F. J. Corbató and R. Morris

By informal communication from the American Committee considering the revised ASCII standards, it appears that the "escape" code given in the published proposal was essentially intended as a fifth device control code. Since all 33 of the control characters have been squandered for other purposes, it consequently becomes necessary to use one of the 95 normally printing text characters for the software escape character. It should be clear that the character chosen as escape will no longer be conveniently useable in GE 636 software, text, etc., (although it can still be inputted with the software escape mechanism), and it is for all practical purposes removed from the stand- ard set of characters. As the best candidate for the "most-unwanted- character-in-the-set" it is proposed that ∧ (circumflex) be selected as the software escape. On equipment which does not have a circumflex but does have a \ (left slash), the latter should be considered a stylized circumflex and used as the escape.

In any case, line printers now only need have 95 graphics (including space) instead of 96. For future flexibility, it is felt that it would be wise to have 96 graphics on any line printer, unless the inclusion of the extra graphic causes a serious degradation in per- formance.