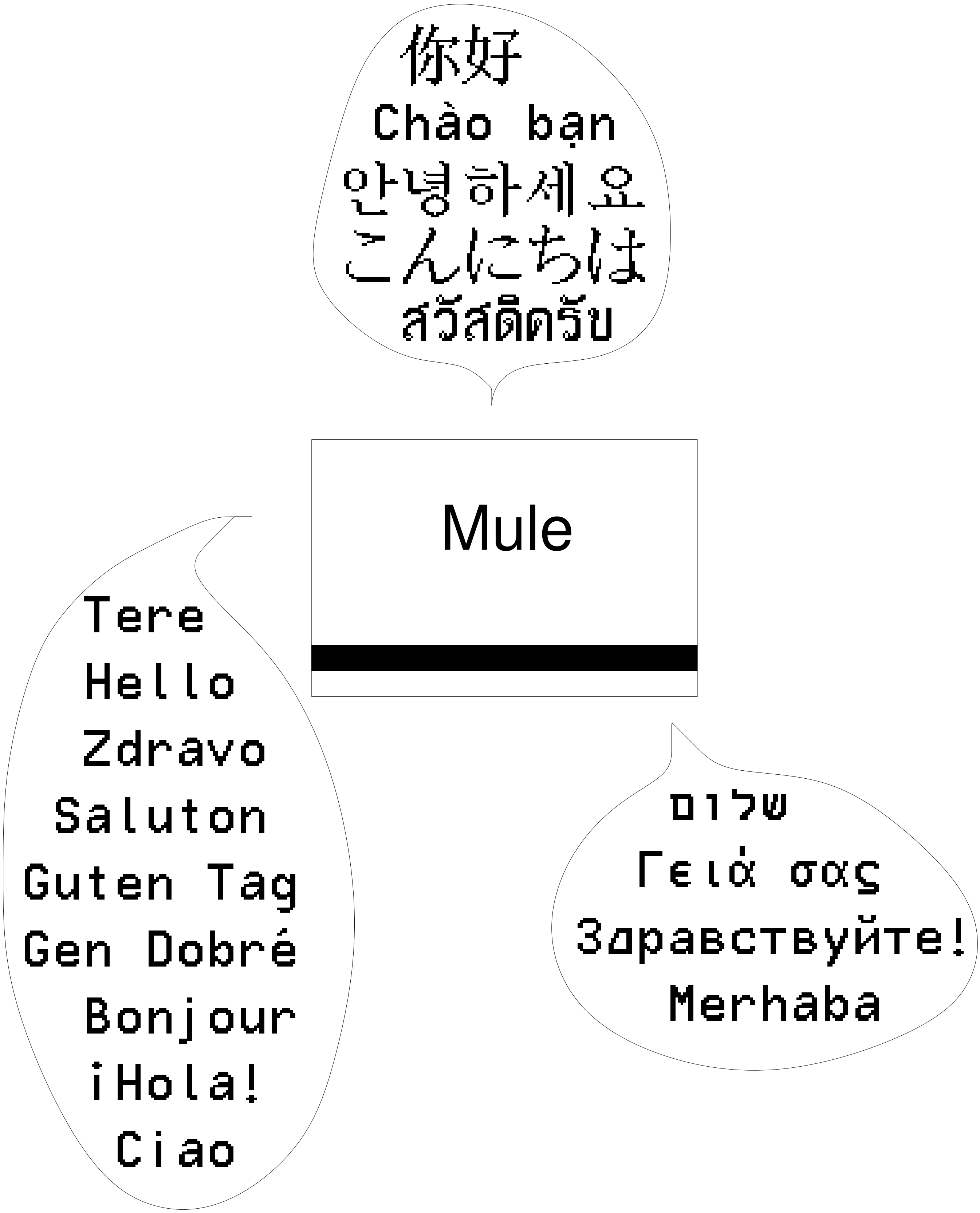
Mule: MULtilingual Enhancement to GNU Emacs

Nishikimi Mikiko
Handa Kenichi
Tomura Satoru

Electrotechnical Lab

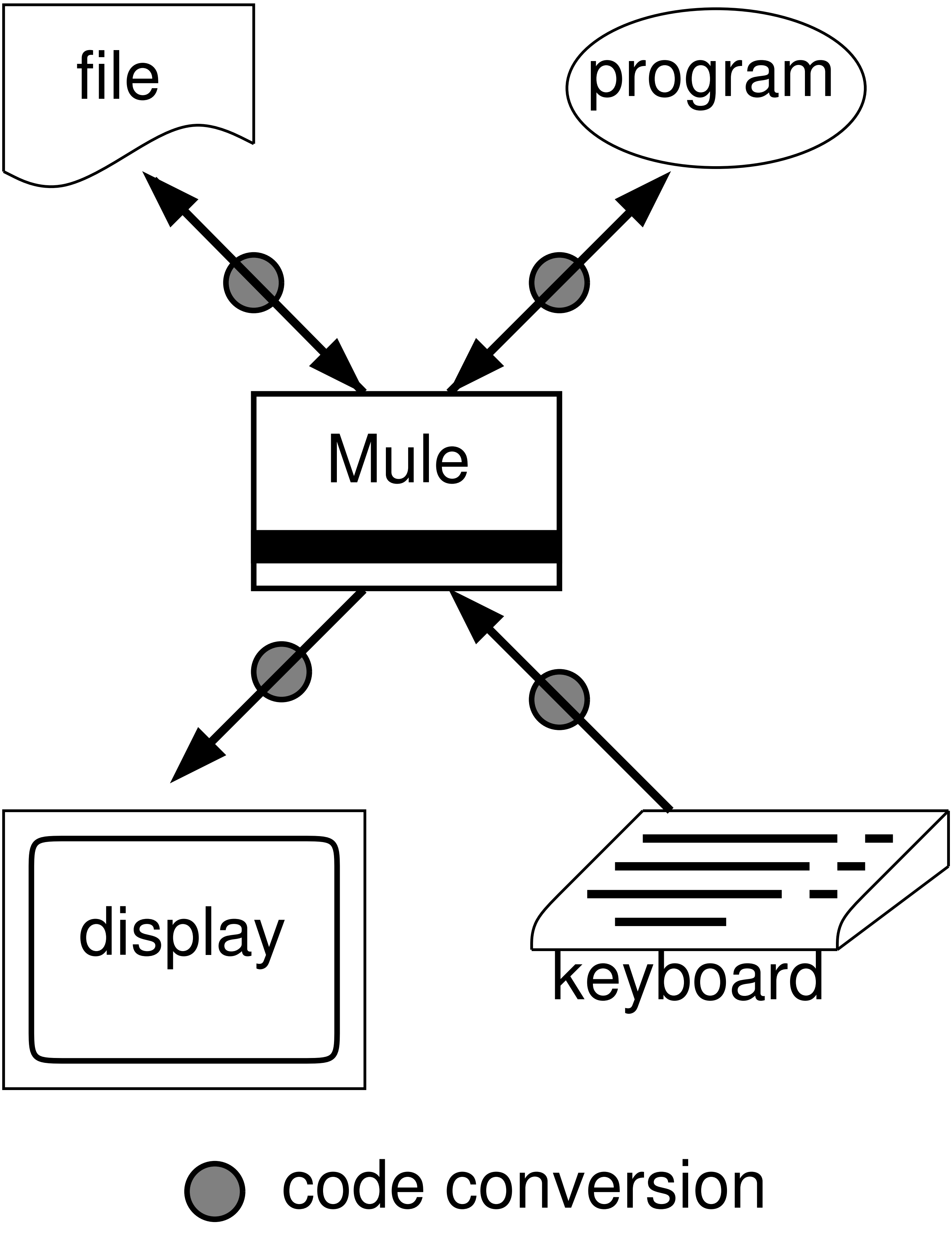

你好
Chào bạn
안녕하세요
こんにちは
สวัสดีครับ
Mule
Tere
Hello
Zdravo
Saluton
Guten Tag
Gen Dobré
Bonjour
¡Hola!
Ciao
שלום
Γειά σας
Здравствуйте!
Merhaba

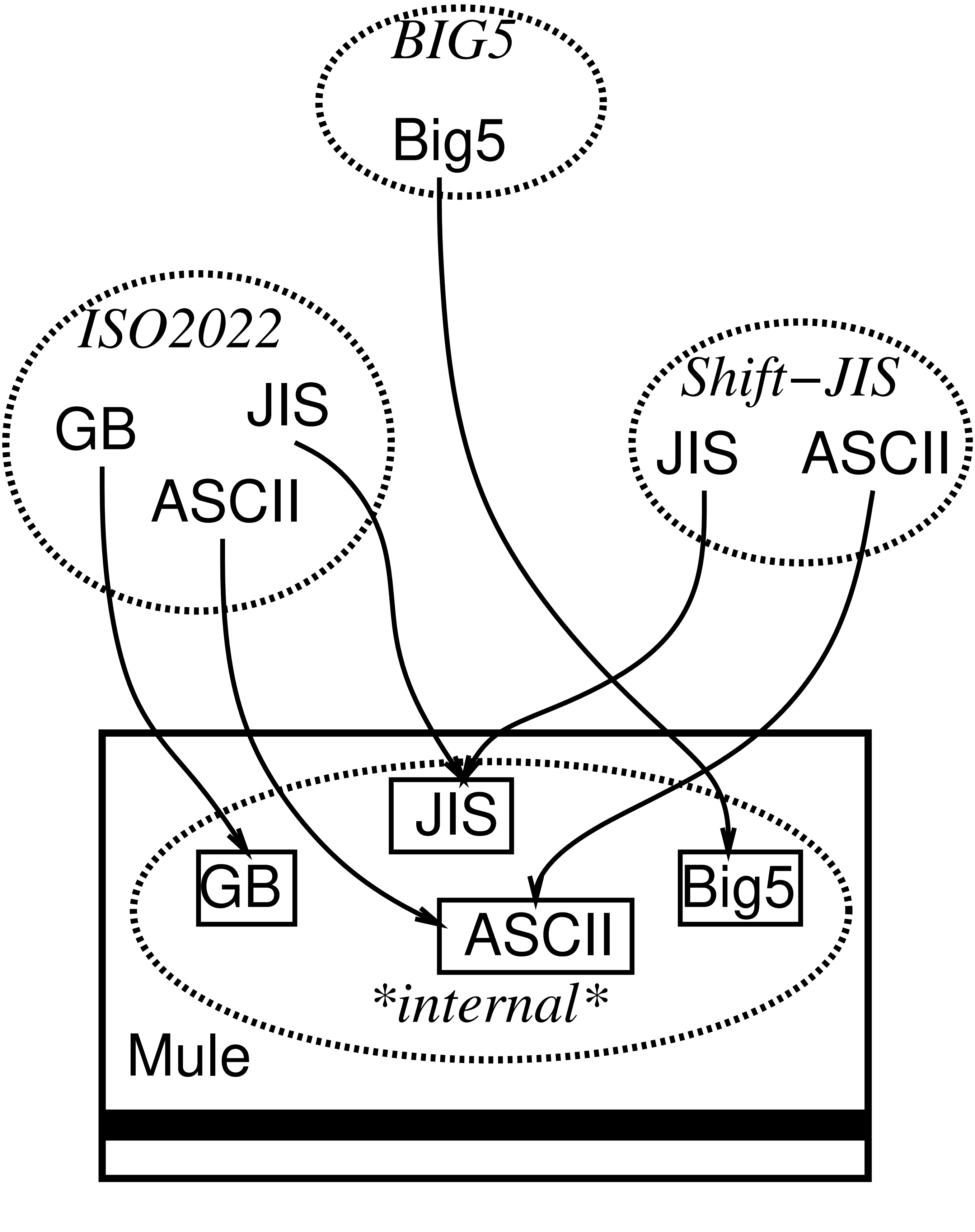
```
You need many fonts to read all.
Please correct this incomplete list and add more!

-----------------------------------------------------------
Danish (Dansk)            Hej, Goddag
English                   Hello
Esperanto                 Saluton
Estonian                  Tere, Tervist
FORTRAN                   PROGRAM
Finnish (Suomi)           Hei
French (Français)         Bonjour, Salut
German (Deutsch Nord)     Guten Tag
German (Deutsch Süd)      Grüß Gott
Greek (Ηλλένικα)          Γειά σας
Hebrew                    שלום
Italiano                  Ciao, Buonjorno
Maltese                   Ciao
Nederlands, Vlaams        Hallo, Dag
Norwegian (Norsk)         Hei, God dag
Polish                    Gen Dobré
Russian (Русский)         Здравствуйте!
Spanish (Español)         ¡Hola!
Swedish (Svenska)         Hej, Goddag
Thai (ภาษาไทย)            สวัสดีครับ, สวัสดีค่ะ
Turkish (Türkçe)          Merhaba
Vietnamese (Tiếng Việt)   Chào bạn
Yoguslavian               Zdravo

Japanese (日本語)                  こんにちは, ｺﾝﾆﾁﾊ, 騏驎
Chinese (中文,普通话,汉语)         你好
Cantonese (粵語,廣東話)            早晨, 你好
Hangul (한글)                      안녕하세요, 안녕하십니까

[--]J:-----Mule: demo            7:28pm  (Fundamental)--Top
Wrote /usr/share/src/mule-0.9/doc/demo
```


file
program
Mule
display
keyboard
code conversion


BIG5
Big5
ISO2022
GB
JIS
ASCII
Shift−JIS
JIS
ASCII
JIS
GB
Big5
ASCII
internal
Mule

◇◇ CharSets and CodingSystem ◇◇

- *Character Sets (non-ECMA*)*

ASCII – English
ISO8859 – European
GB2312, CNS11643, BIG5* – Chinese
JISX0201, 208, 212 – Japanese
KSC5601 – Korean
TIS620 – Thai
VISCII*, VSCII* – Vietnamese
...

- *Coding Systems*

ISO–2022 – generic
EUC – ASCII and GB,JIS,KSC
ISO-2022-JP – ACII and JIS
CTEXT – generic
...
Shift-JIS – ASCII and JIS
BIG5 – Big5
HZ, ZW – ASCII and GB
...

◇◇ Internal Representation ◇◇

- ISO's character set

$$\texttt{CODE1 [CODE2]}$$

$$\texttt{0x20} \leq \texttt{CODE1,CODE2} \leq \texttt{0x7F}$$

- Mule's internal representation
  - for ASCII

$$\cdots \underline{\texttt{CODE1}} \cdots$$

  - for the ISO's character sets

$$\cdots \underline{\texttt{LC C1 [C2]}} \cdots$$

$$\texttt{0x81} \leq \texttt{LC} \leq \texttt{0x99}$$

$$\texttt{0xA0} \leq \texttt{C1}(= \texttt{CODE1} + \texttt{0x80}) \leq \texttt{0xFF}$$

$$\texttt{0xA0} \leq \texttt{C2}(= \texttt{CODE2} + \texttt{0x80}) \leq \texttt{0xFF}$$

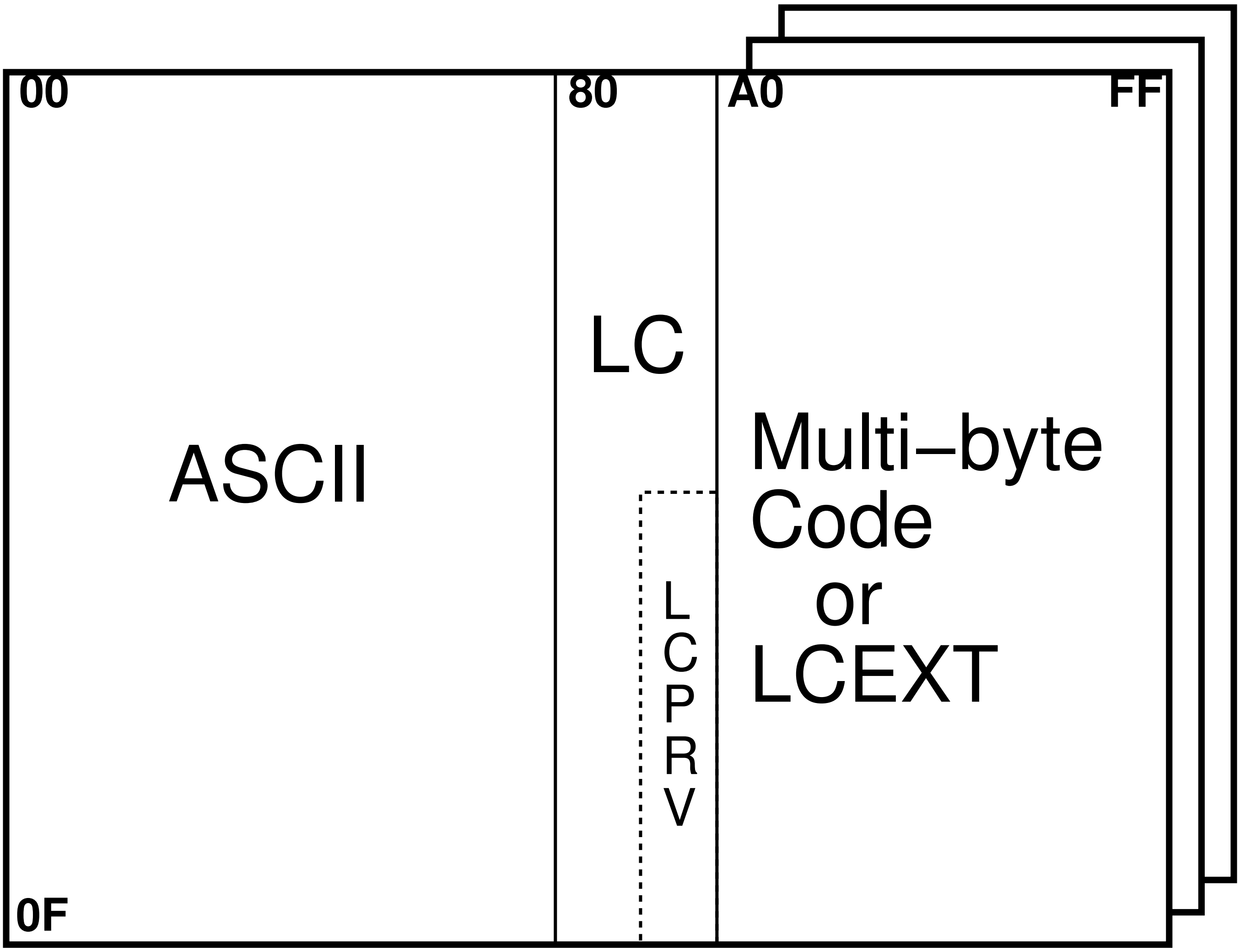
  - for private character sets

$$\cdots \underline{\texttt{LCPRV LCEXT C1 [C2]}} \cdots$$

$$\texttt{0x9A} \leq \texttt{LCPRV} \leq \texttt{0x9E}$$

$$\texttt{0xA0} \leq \texttt{LCEXT} \leq \texttt{0xFF}$$

- Code area map



- Properties of character sets

1. **BYTES**
   how many bytes it occupies in a memory?
   1, 2, 3, or 4
2. **WIDTH**
   how many columns it occupies on a screen?
   1 or 2
3. **DIRECTION**
   right-to-left or *left-to-right*
4. **DESCRIPTION**
   short document about the character set
5. **TYPE** – ISO specs
   how many characters are in it?
   94, 96, 94x94, or 96x96
6. **GRAPHIC** – ISO specs
   where to designate by default?
   G0, G1, G2, or G3
7. **FINAL-CHARACTER** – ISO specs
   final character of the designation sequence.

◇◇ Inputting Methods ◇◇

- QUAIL (European, Korean, Thai, Vietnamese)

  key-sequence

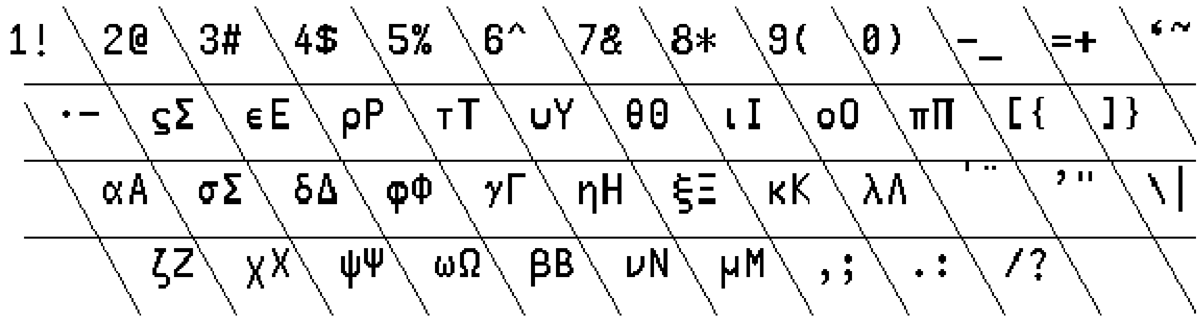
  $\Downarrow$

  characer code

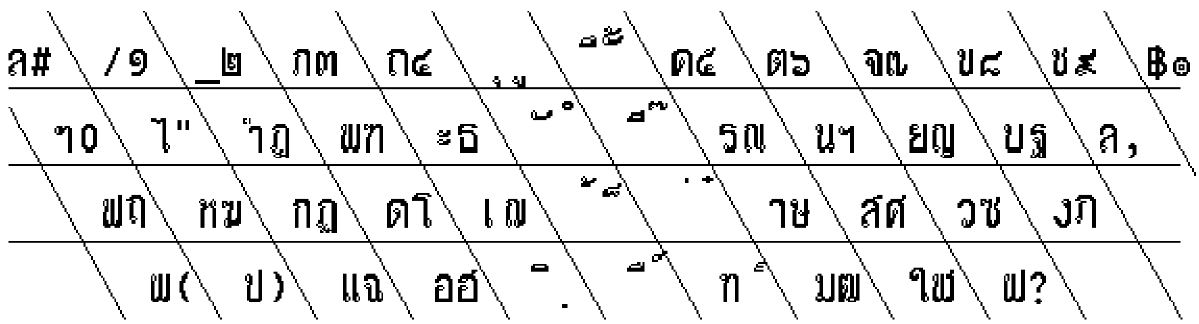
  - Use Roman mnemonics

    `o~ -> õ, ?/ -> ¿`

  - Use alternate keyboard



Greek keyboard



Thai keyboard

- EGG (Japanese, Chinese)

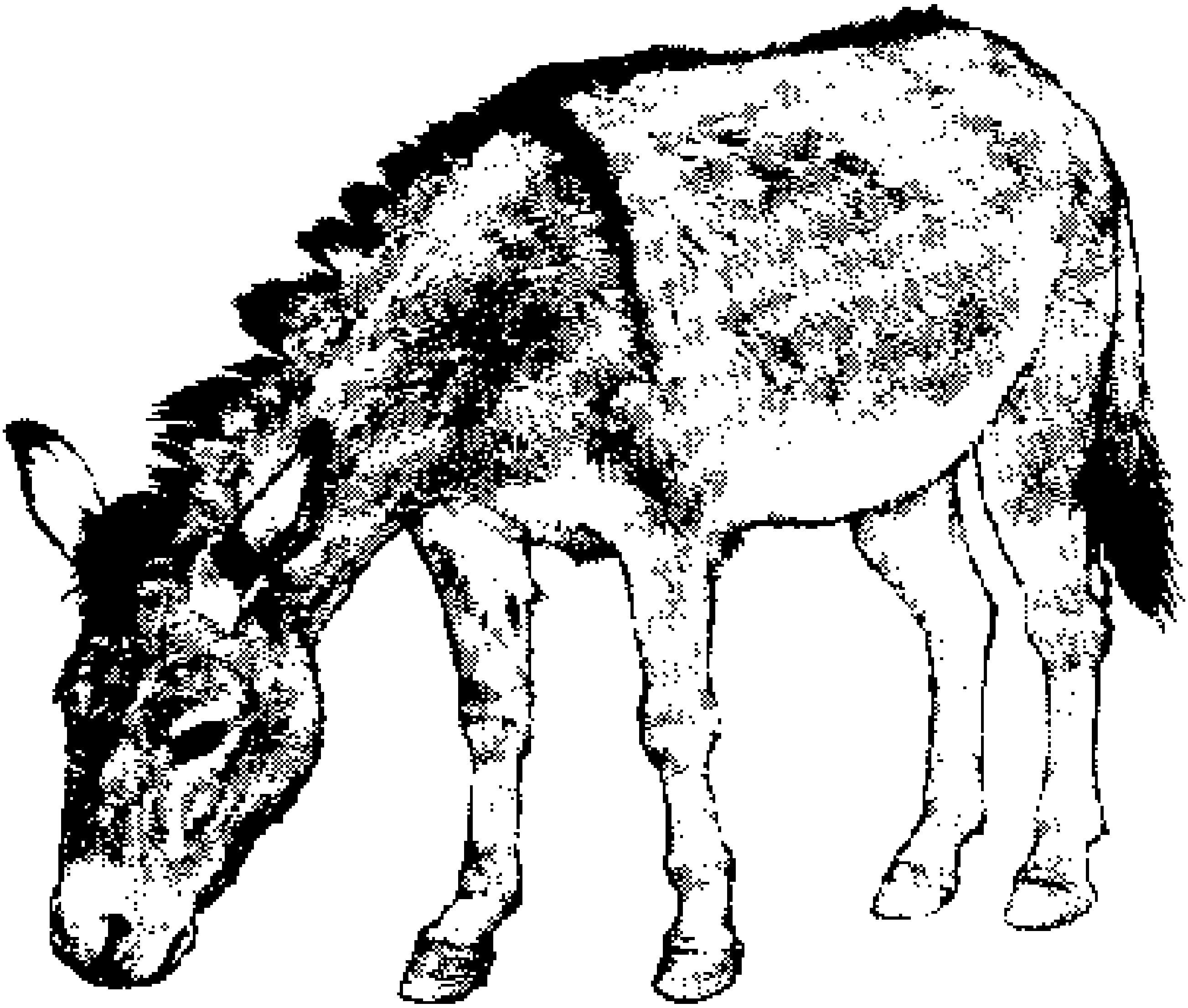
$$\underbrace{\begin{array}{ccc}\text{key-sequence} & \text{key-sequence} & \text{key-sequence}\\ \Downarrow & \Downarrow & \Downarrow \\ \text{characer code} & \text{characer code} & \text{characer code}\end{array}}$$

$$\Downarrow$$

$$\boxed{\text{translation server}}$$

$$\Downarrow$$

$$\overbrace{\text{characer code sequence}}$$



horse(♀) x donkey(♂) –> mule

horse(♂) x donkey(♀) –> hinny