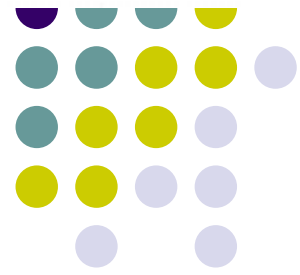


DB



- 
- 
- ◆ ISO 646 IRV & ISO 2022
- 
- ◆ ISO 10646
- ◆ Unicode
- &
- 
-

- ( )
- ,
- (character)
- 가
- ,
- '90 ,
  - ◆ BCD < eBCDi c < ASCII + NAPLPS < Unicode
  - ◆ ISO 646IRV+ISO 2022 < ISO 10646
- : < < +
- , 가



# 1

- BCD (binary coded decimal)
  - ◆ nibble – 4
- EBCDIC (Extended BCD Interchange Code)
  - ◆ one byte – 8
- ASCII (American Standard Code for Information Interchange)
  - ◆ 7 – ASCII
  - ◆ 8 – ASCII
- NAPLPS (North American Presentation Layer Protocol Syntax)
  - ◆ ASCII ( , )
  - ◆ 가 ( 2400bps)
  - ◆ : , ,

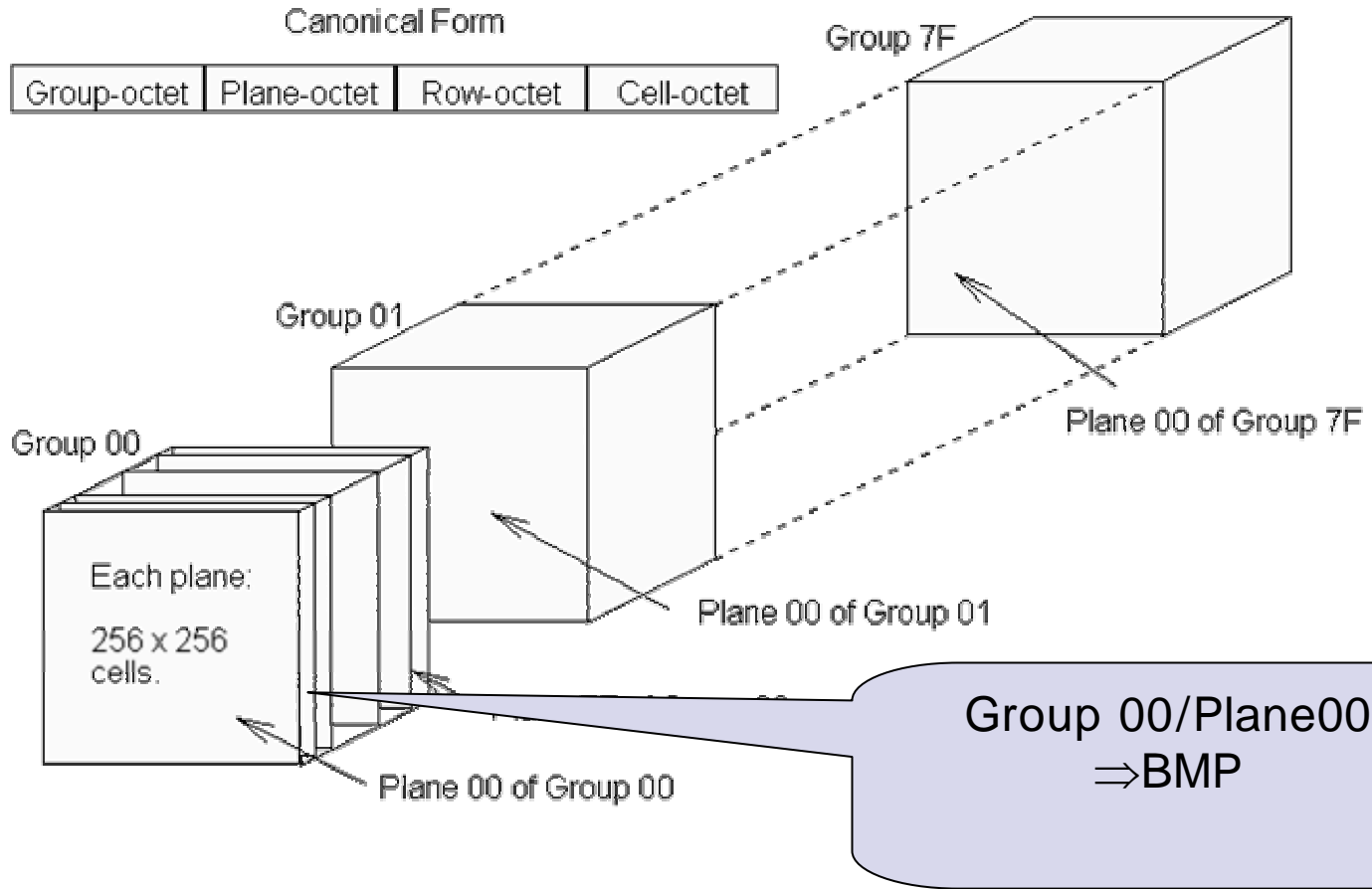


# ISO 10646

- ISO/IEC JTC1/SC2 (Coded Character Set)/WG2 (Multi-Octet Codes)
  - ◆ Unicode : Apple, Metaphor, RLG, Sun, Xerox 1989
- 1991 ISO 10646 BMP Unicode
  - ◆ Basic Multilingual Plane, Group 00/ Plane 00
  - ◆ Universal Multiple Octet Coded Character Set  
ISO 10646-1:1993 ↔ Unicode 1.1
    - ✓ Part 1: (BMP)
    - ✓
  - ◆ ISO 10646-1:2000 ↔ Unicode 3.0

# ISO 10646 ( )

- 4





- 1991 ([www.unicode.org](http://www.unicode.org))
  - ◆ “ 가 ” , ’ , ’
  - ◆ , ,
  - ◆ : Apple, HP, IBM, JustSystem, Microsoft, Oracle, SAP, Sun, Sybase, Unisys
  - ◆ SW: XML, Java, ECMAScript (JavaScript), LDAP, CORBA 3.0, WML, etc.
  - ◆ ISO/IEC 10646
- 1997 W3C
  - ◆ HTML 4.0, XML 1.0, CSS2 / [www.w3c.org](http://www.w3c.org)
- 38,000 16
  - ◆ 65536(2<sup>16</sup>) 18,000
- ISO 646IRV(ASCII), ISO 8859-1(Latin-1)



( )

● Unicode 5.0 ↔ ISO 10646-1:2005?

◆ 2006 4 . 2007 1

◆

✓

✓ GB18030, JIX X0213, HKSCS 가  
4.1

✓

가 1,2 , ISO/IEC 10646:2003

✓

( , , ),

, ,

SW



( )

- Unicode 5.0

Graphic	98,884
Format	140
Control	65
Private Use	137,468
Surrogate	2,048
Noncharacter	66
Preserved	875,441



# ISO 10646

- , , ,
- 가
- CJK
- FEFF
- UTF-16, UTF-8, UCS-2 (surrogate )
- 
-



# UTF-8

- - ✓ ISO 646IRV, ISO 2022
- UCS Transformation Format
  - ◆ UCS: Universal Multi-Octet Coded Character Set
    - ✓ UCS
    - ✓ 4 ,  $2^{32} = 40$  가 가
    - ✓ ,
- Unicode Transformation Format
  - ◆ UCS-4 BMP = UCS-2 = Unicode



# UTF-8 ( )

- UTF-1

- ◆ ISO-2022

CO, SPACE, DEL, C1  
. (ISO/IEC-10646:1993 Annex G)

- UTF-7

- ◆ 7 bit Unicode

- UTF-8

- ◆ ISO 646(ASCII) Unicode

- UTF-16 (surrogate )

- ◆ ISO 10646

- ◆ Group 00 16 plane



# UTF-8

- UTF-8
- U+0000~ U+007F UCS                      0x00 ~ 0x7f
- ASCII                      7                      ASCII  
                     ASCII                      UTF-8
- U+007F                      UCS  
                     가                      ,                      가                      ASCII                      (bit set)  
                     가                      .                      0x7f)                      (0x00-
- ASCII가                      0xC0                      0xFD                      ,                      가                      .  
                                          가                      0x80                      가                      0xBF                      가  
                     .                      resynchronization



# UTF-8

- 가 UCS .
- UTF-8 6 .  
가 , 16 BMP  
3 가 .
- Big endian UCS-4 .
- 0xFE 0xFF UTF-8 .



# UTF-8 ( )



▪



- U-00000000 - U-0000007F: 0xxxxxxx
- U-00000080 - U-000007FF: 110xxxxx 10xxxxxx
- U-00000800 - U-0000FFFF: 1110xxxx 10xxxxxx 10xxxxxx
- U-00010000 - U-001FFFFF: 11110xxx 10xxxxxx 10xxxxxx 10xxxxxx
- U-00200000 - U-03FFFFFF: 111110xx 10xxxxxx 10xxxxxx 10xxxxxx 10xxxxxx
- U-04000000 - U-7FFFFFFF: 1111110x 10xxxxxx 10xxxxxx 10xxxxxx 10xxxxxx 10xxxxxx



▪

가



▪



1

▪



# UTF-8

UTF-8 . A0..BF

가

◆ <C0, AF> . C0

◆ <E0, 9F, 80> . E0 , 9F

◆ <F4, 80, 83, 92> . 가 ( )

80..BF

Code Points	1st Byte	2nd Byte	3rd Byte	4th Byte
U+0000..U+007F	00..7F			
U+0080..U+07FF	C2..DF	80..BF		
U+0800..U+0FFF	E0	A0..BF	80..BF	
U+1000..U+FFFF	E1..EF	80..BF	80..BF	
U+10000..U+3FFFF	F0	90..BF	80..BF	80..BF
U+40000..U+FFFFFF	F1..F3	80..BF	80..BF	80..BF
U+100000..U+10FFFF	F4	80..8F	80..BF	80..BF





# Unibook

- ,
- 5.0.0 -Beta2 / 2006.4.
  - ◆ : [www.unicode.org/unibook](http://www.unicode.org/unibook)
  - ◆ unibook.exe
- 4.1.0
- 3.0



# Linux

- 2006.2. ETRI

SW Linux

- ◆ TTA

- ◆ ISO 10646, Unicode

- ◆ UTF-8

- ◆

- ◆

- ◆ (GNOME,

)

- ◆ , 가 ,



# DBMS Unicode

- Oracle
  - ◆ DB Unicode columns(UTF8,AL16UTF16,etc) code page
  - ◆ Oracle 9.x: NCHAR/NVARCHAR column
- SQL Server
  - ◆ Unicode Column: UCS2 little endian format
  - ◆ : U+FFFF, : U+FEFF
- Informix
  - ◆ Unicode column: UTF8
- IBM DB2
  - ◆ Unicode

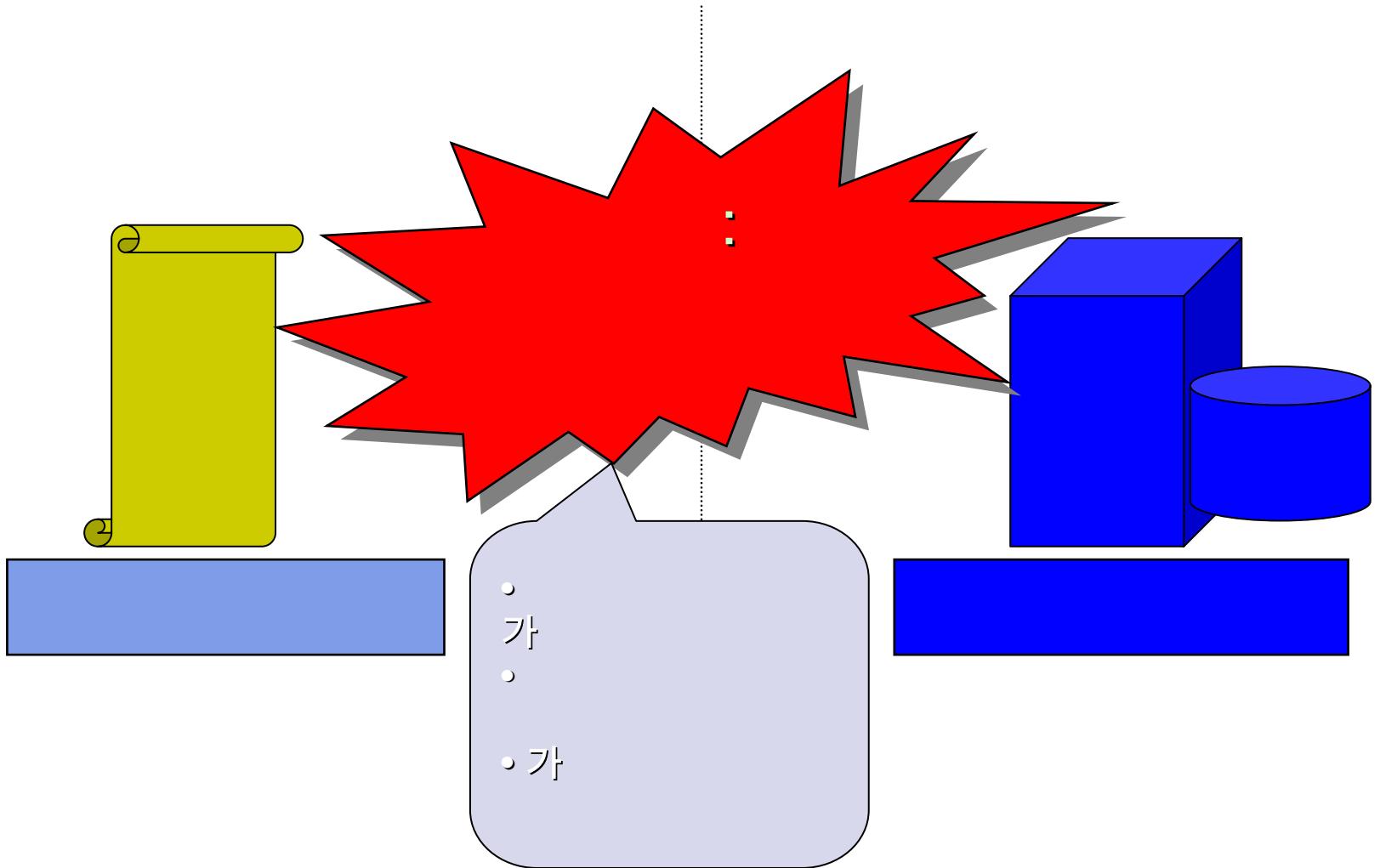


/



\_\_\_\_\_





( )



◆ 1987 ? ?



815

WP

◆ MS



hwp



MS

ISO 10646





- 
- 
- 
- 
-



# -1



:

(ㄱ ㄴ ㄹ ㅁ ㅂ ㅅ ㅇ)

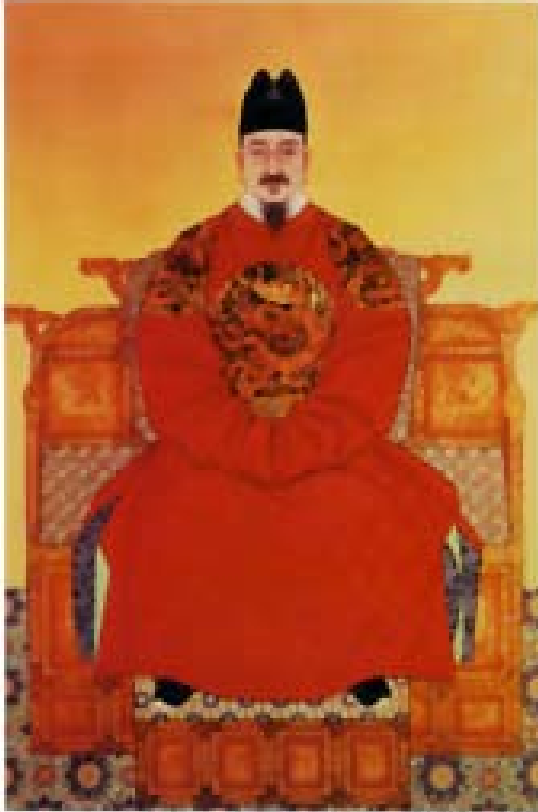


:

( • — | )



# -2



조선의 세종대왕 (1397-1450)

- (1397 - 1450)
- ◆ :
- ◆ 25 (1443)
- 
- ◆ 500
- ( ) 가
- ...

# - 3

- 
- 
- 

- ◆ 1443
- ◆ 1446
- ◆ 1447

,

,

世宗 御製 訓民正音  
 製정신은글지슬州 御製 製정신은글지슬산  
 그리란해후은글지슬州 御製 製정신은글지슬산  
 이오음음은수리니해후민인정정음음  
 은음외姓상그리차시논정정음스리라  
 國語 語音 이 國語인나라이관사음인  
 키지라語音그말씀이라  
 나랏말씀이  
 異字 中 國 語 異字 中 國 語  
 그어하은근거쳐쓰는字宗이라中 國 語  
 皇 帝 帝 王 帝 王 신나라히우리나랏





# - 1

- ◆ 爲螢. 日, 如 爲薪, . 爲蹄. 口, 如
- ◆ : 爲虎, : 심 爲泉. 人, 如: 爲海松, .
- ◆ 爲池. 己, 如 . 三 爲月, : 爲星之類



- ◆ 有天地自然之聲, 則必有天地
- ◆ 自然之文, 所以古人因聲制字
- ◆ 以通萬物之情,

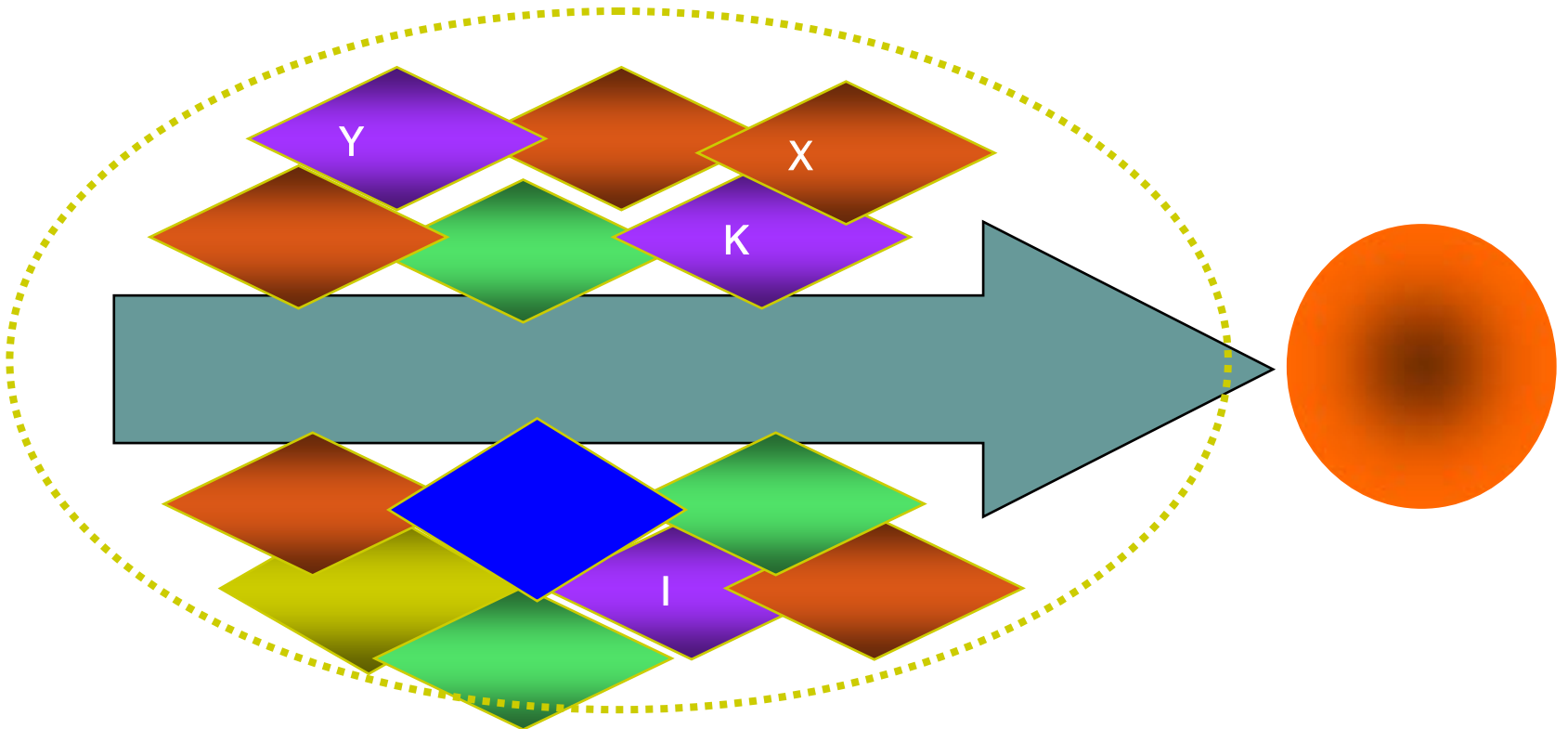


## - 2

- ◆ 括. 以二十八字而轉換無窮, 簡
- ◆ 而要, 精而通. 故智者不終朝而
- ◆ 會, 愚者可浹旬而學. 以是解書,
- ◆ 可以知其義. 以是聽訟, 可以得
- ◆ 其情. 字韻則清濁之能辨, 樂歌
- ◆ 則律呂之克諧. 無所用而不備,
- ◆ 無所往而不達. 雖風聲鶴戾, 鷄
- ◆ 鳴狗吠, 皆可得而書矣. 遂

-5

•  
◇ =>



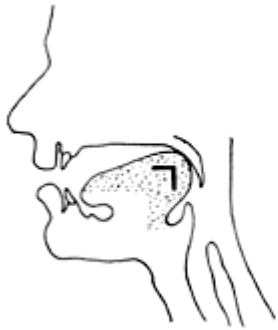


# -6 ( )

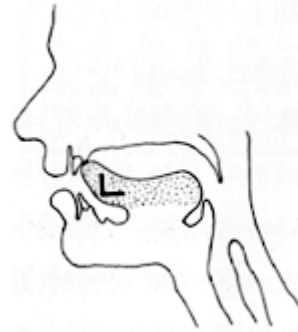
- :  
◆ 天地人: { · — | }, 牙舌脣齒喉: { ㄱ ㄴ ㄹ ㅁ ㅂ ㅅ }
- : 17 { ㄱ ㄴ ... ㅎ △ ○ ㅎ }
- : 11 { ㅏ ㅑ ... — | · }
- : 終聲不用初聲( )
- : 字素(grapheme)
- ◆ : 脣音(ㅁ ㅂ ㅃ ㅍ) => 脣輕音(ㅇ ㅁ ㅂ ㅃ ㅍ)
- ◆ : (2, 3字) 各字, 合用
- ◆ : (가 - , - )
- ◆ : 凡字必合以成音( )



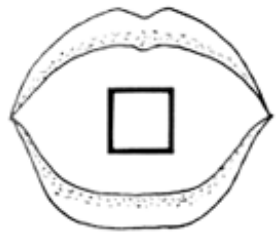
# (ㄱㄴㅇㅅㅇ)



⇒ ㄱ ⇒ ㅋ  
⇒ ㆁ



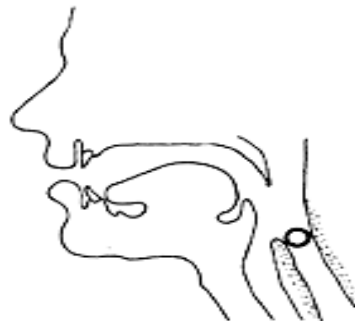
⇒ ㄴ ⇒ ㄷ ⇒ ㅌ  
⇒ ㄸ  
⇒ ㄹ ⇒ ㄺ



⇒ ㅁ ⇒ ㅂ ⇒ ㅃ ⇒ ㅍ ⇒ ㅈ  
⇒ ㅍㅍ ⇒ ㅍㅇ  
⇒ ㅍㅇ  
⇒ ㅍㅇ



→ ㅅ → ㅆ → ㅆ → ㅆ  
→ ㅆ  
→ ㅆㅆ → ㅆㅆ  
→ ㅆㅆ  
→ ㅆ  
→ ㅆ  
→ ㅆㅆ → ㅆㅆ  
→ ㅆㅆ  
→ ㅆ  
→ ㅆ



→ ㅇ → ㆁ → ㅇ → ㅇㅇ  
→ ㅇㅇ  
→ ㅇ

# -7



- ◆ 2,3 : ,
  - ✓ : 17 + 17Π2 + 17Π3 = 5219
  - ✓ : 11 + 11Π2 + 11Π3 = 1463
  - ✓ : (17+) + 17Π2 + 17Π3 = 5220
- ◆ 가
  - ✓ 5219 x 1463 x 5220
  - ✓ = 39,856,772,340 ( 399 )

- ( ) : |H| ( )

{ , }  
⋮

$$|H| = |C1| \times |V| \times |C2|$$

$$= 39,856,772,340$$

2,3

⋮

|C1|:5219, |V|:1463, |C2|:5220

⋮

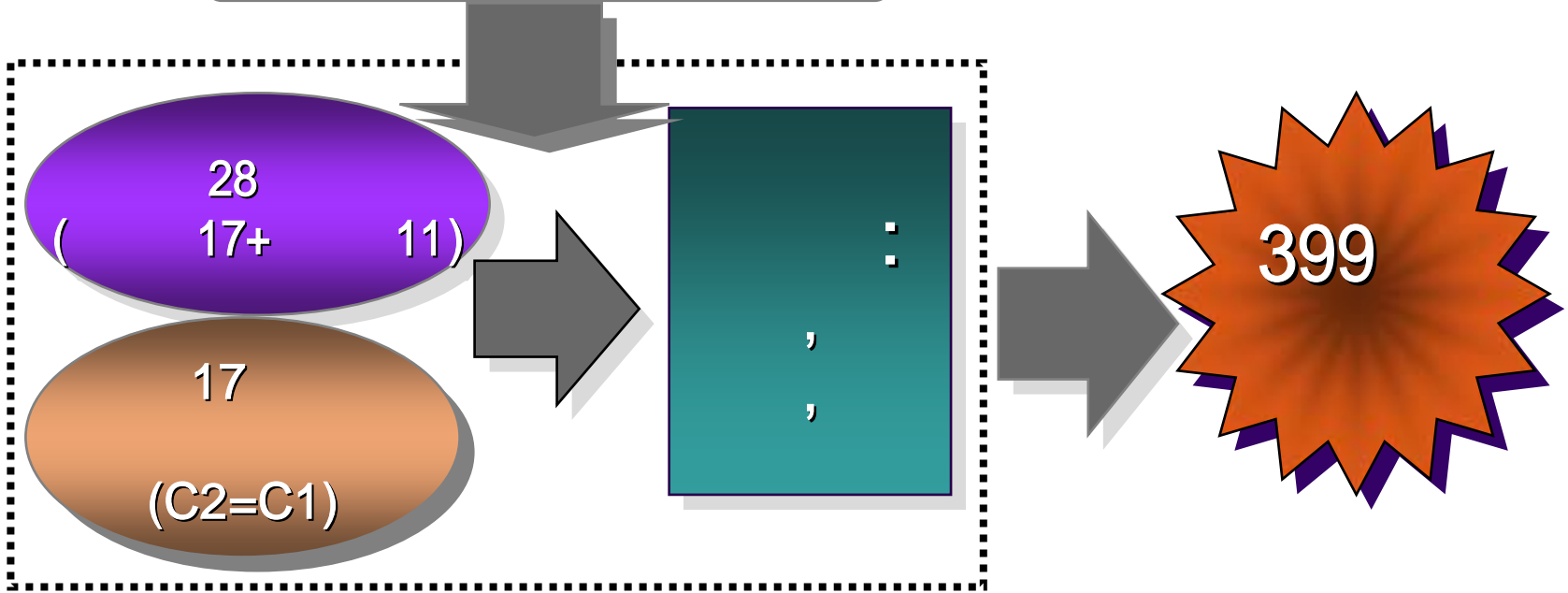
$$(17 \quad +11 \quad +17)$$

$$- 17 \quad ( \quad )$$

$$= 28$$

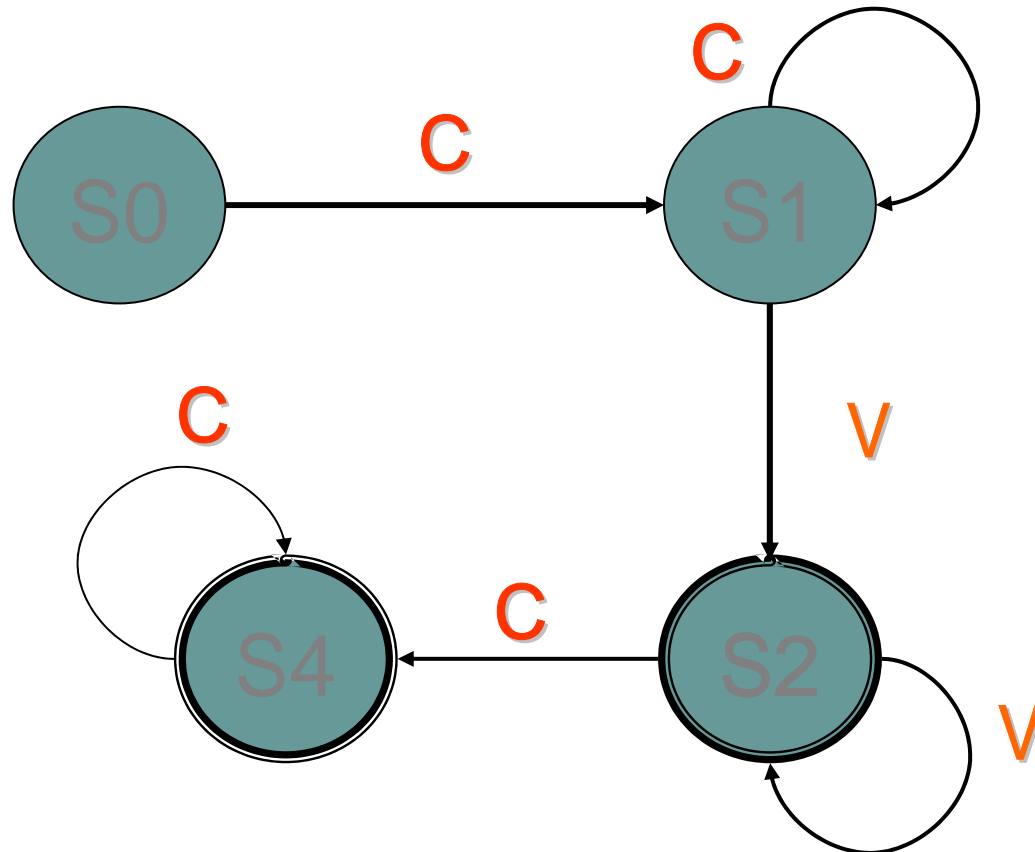
- 1

# 天地自然之聲



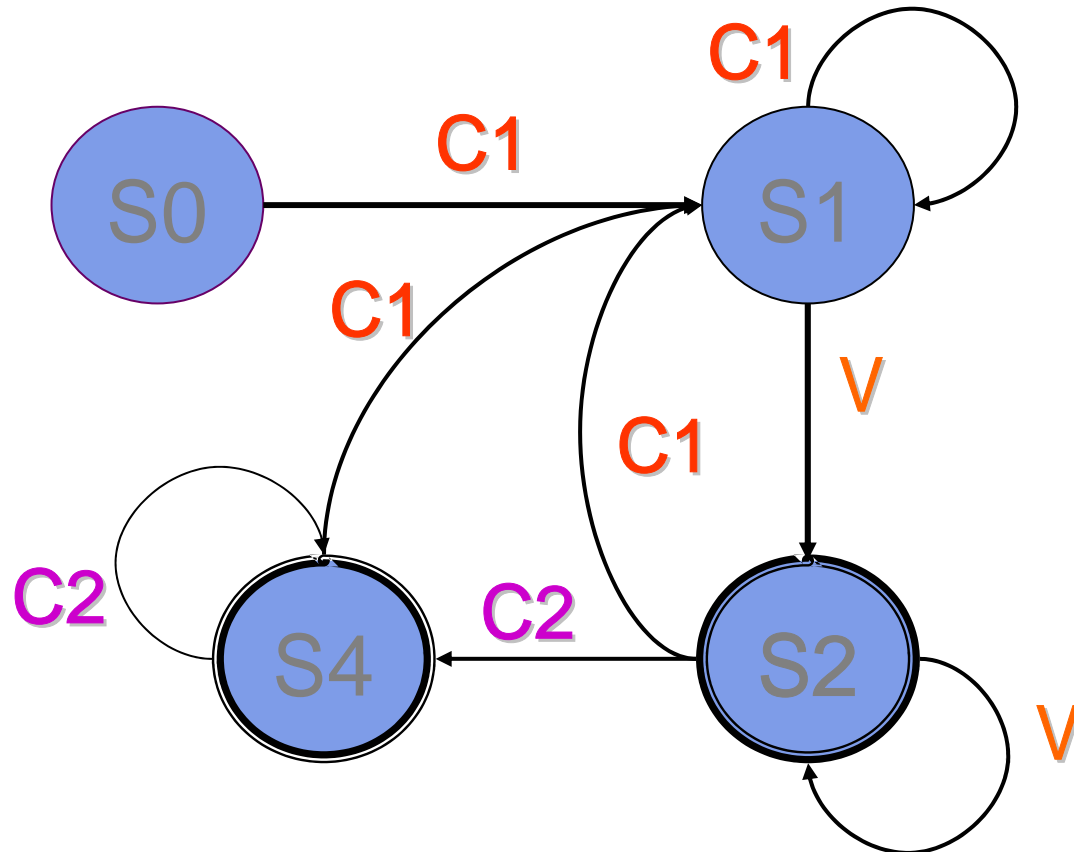
# (Automaton)

- JHS- (Acceptor: )



# (Automaton)

- BJY- (Accepter: )



- 2

有天地自然之聲 即必有天地自然之文

28 + 17

5219, 1463, 5220

399



-1

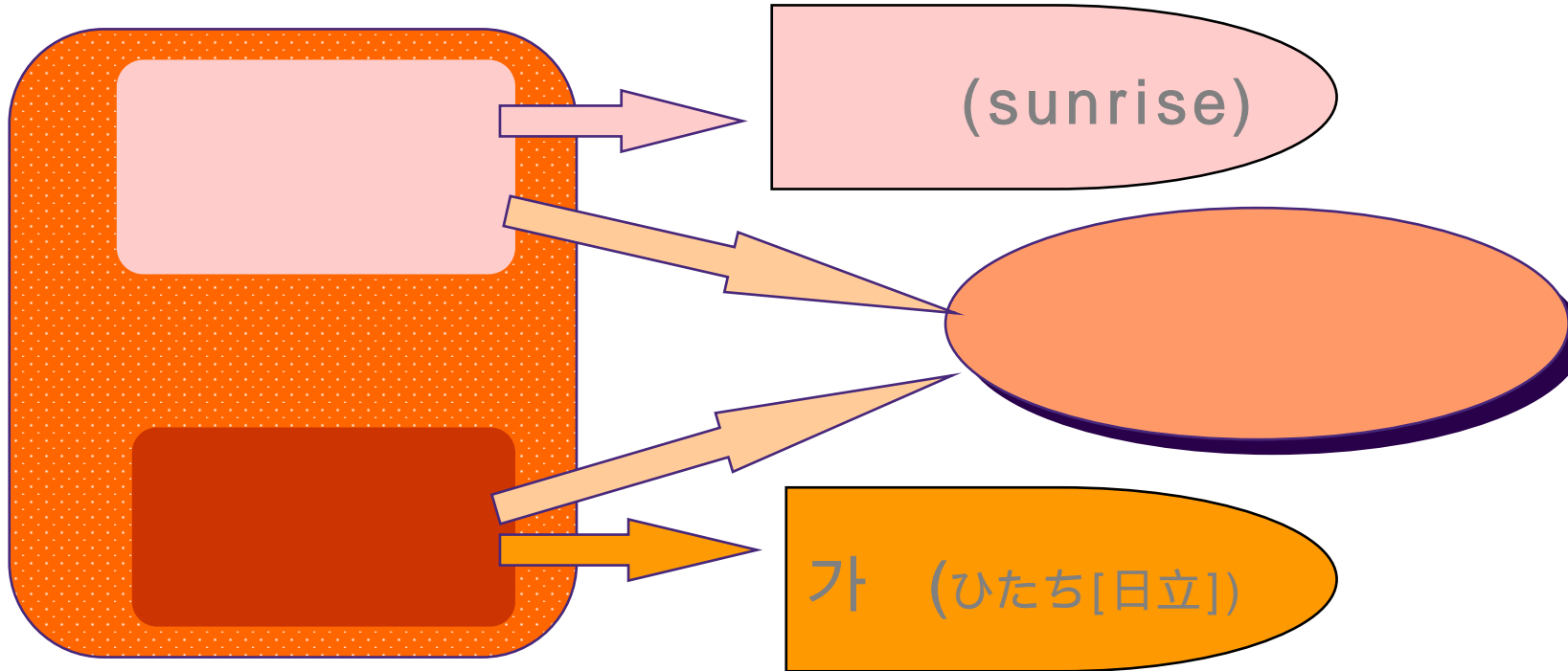


:

,

:

:가





# -2

- 

- ◆

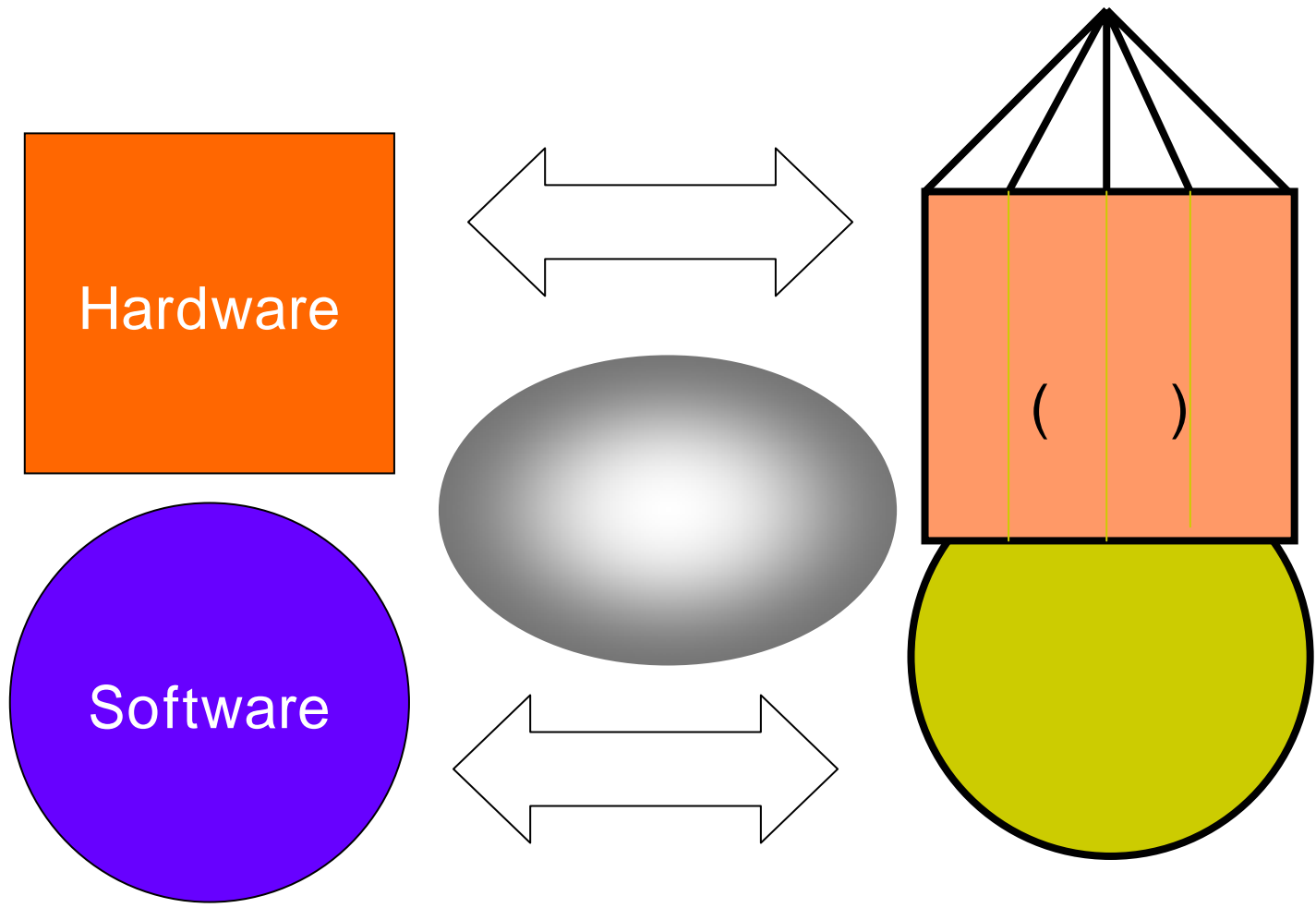
- ◆       : {                    }

- ◆       : {                    } {                    }

- 

- ◆       :                   ,                   ,

- ◆       : 가                   ,                   ,                   ,







◆ KS C 5601-1987=> KS X 1001



✓ 1933 “ ”  
✓ 19 x 21 x (27+1) = 11172



, , ,



4 가



=> 320

- 1



⇒



가



:

,



,



◆ KPS 9566-1997

# -2

- $\Rightarrow$  가
- ◆ ( )
- ◆ 가 ( )
- ✓ : ㄱ ㄴ ㄷ ㄹ ㅁ ㅂ ㅅ ㅈ ㅊ ㅋ ㅌ ㅍ ㅎ ㄱ ㄷ ㅂ ㅅ ㅈ ㅊ
- ✓ : ㅏ ㅑ ㅓ ㅕ ㅗ ㅛ ㅜ ㅠ ㅡ ㅣ ㅏ...
- ✓ : ㄱ ㄴ ... ㅅ ㅈ ... ㅎ ㄱ ㄷ ㅂ ㅅ ㅈ ㅊ ...

- 1996

- ◆ , 가 가

- 
- ◆ 가 ,



# Unicode / ISO 10646-BMP



◆ **KS C 5700-1995=> KS X 1005**

◆ Unicode

◆ 가 가? =>



✓ Hangul Jamo( ) 240

✓ Hangul compatability Jamo( ) 51 +x

✓ Hangul( ) 2350 + ... ( )

✓ Hangul( ) 11172



# Unicode

- KS X 1005-1 : 3

- ◆ , ,

- 11172

- ◆

- 

- ◆ 11172 240

- 240

- ◆





# -1996

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	
8																	
9																	
A																	
B																	
C																	
D	Fc	Г	L	С	ㄹ	□	ㅁ	人	△	○	◌	ㅈ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ
E	Fj	ㅌ	ㅍ	ㅊ	ㅋ	ㅌ	ㅍ	ㅈ	ㅊ	ㅈ	ㅊ	ㅈ	ㅊ	ㅈ	ㅊ	ㅈ	ㅊ
F	Fg	Г	L	С	ㄹ	□	ㅁ	人	△	○	◌	ㅈ	ㅊ	ㅈ	ㅊ	ㅈ	ㅊ



# 가 1



( )	( )	( )
( )	<b>399</b>	<b>O</b>
	<b>2,350</b>	<b>X</b>
<b>ISO</b>	<b>11,172</b>	<b>X</b>
<b>ISO</b>	<b>493,020</b>	<b>O(-)</b>
	<b>11,172</b>	<b>O(-)</b>
	<b>11,172</b>	<b>O(-)</b>

- - ◆ 가 ⇔
  - ✓
  - ◆ 가
  - ✓
  - ✓
  - ✓
  - QWERTY, Dvorak keyboard
  - ◆
  - ✓
  - ✓ , , 가

# -1



가



UNESCO



( ) :



...



...

.

.

-2

ISO 10646

3 가





# NEAOSS Forum

- WG3

  - ◆ [www.neaoss.org](http://www.neaoss.org)

- SWG1

  - ◆ SW IM Engine

    - ✓ IIIMF, SCIM, (GTK )

  - ◆ , ,

    - ✓ ,

    - ✓ ,

  - ◆ IM Engine /

- ISO/IEC JTC1/SC2



◆ SW



SW

=>



SW

