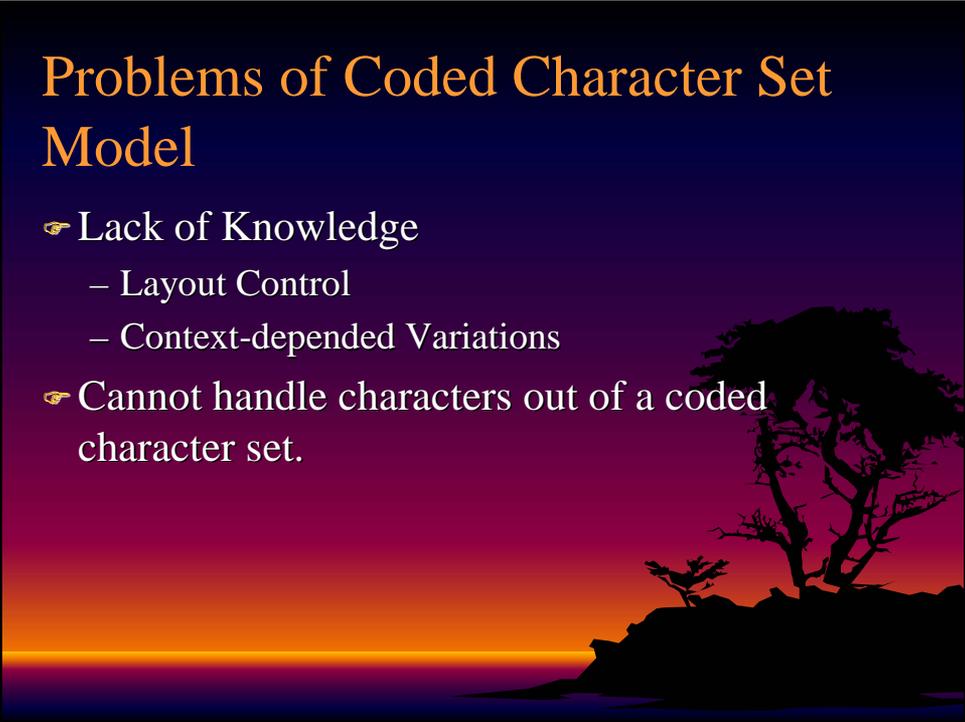# Surface or Essence:
Beyond Coded Character Set Model

Shigeki Moro, Hanazono Univ.
(s-moro@hanazono.ac.jp)

Problems of Coded Character Set Model

☞ Lack of Knowledge
  – Layout Control
  – Context-depended Variations
☞ Cannot handle characters out of a coded character set.

Essentialism of Unicode (1)

☞ 10 Unicode Design Principles (*Unicode 4.0*, p.14)
1. Universality
2. Efficiency
3. <u>Characters, Not Glyphs</u>
4. <u>Semantics</u>
5. Plain Text
6. Logical Order
7. <u>Unification</u>
8. Dynamic Composition
9. Equivalent Sequences
10. Convertibility

Essentialism of Unicode (2)

☞ Definition of *character* (1)

– "Characters are the abstract representations of the smallest components of written language that have semantic value. (…) Characters represented by code points… The Unicode Standard deals only with character codes." (ibid, p.15)

Essentialism of Unicode (3)

☞ Definition of *character* (2)

– "Characters have well defined semantics. Characters property tables are provided for use in parsing, sorting, and other algorithm requiring semantic knowledge about the code points." (ibid., p.17)

Essentialism of Unicode (4)

☞ Aristotle's Essentialism

– "A definition is an account (*logos*) that signifies an essence." (*Topics*, 102a3)

– "The formula (*logos*) of the essence of *x* is a formula in which *x* itself does not appear but which expresses (*legonti*) *x*." (*Metaphysics*, 1029b19-20)

Reason of Essentialism

☞ If a context changes, glyphs etc. will also change.
☞ Characters can cross contexts.

What crosses borders? (1)

☞ *character* / essence ?

to run

to go

What crosses borders? (2)

☞ écriture / surface?

Derrida's Theory of écriture

☞ *polysémie*
  – Polysemy which essence has

☞ *dissémination*
  – Footprints which it left when écriture passed contexts

☞ The simplicity of a character is always earlier than the polysemy of a character.

Metaphysics of Character

Example

Example

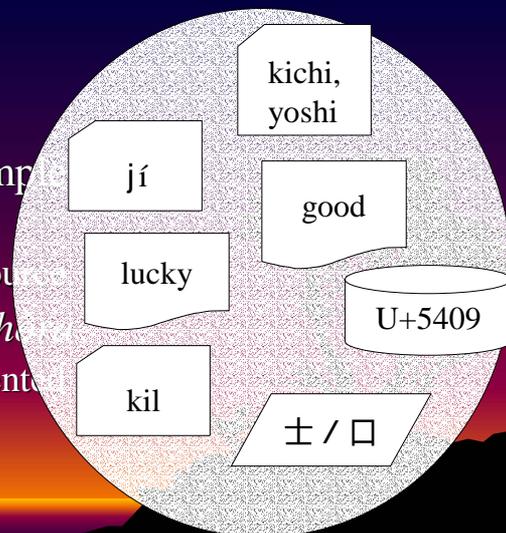Example   }   Dictionary

Example

Example

Example

Example

Example

Example

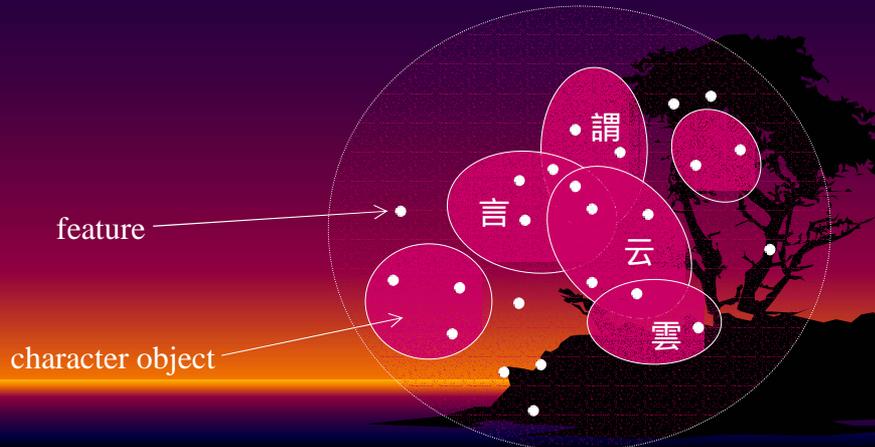# Chaon Model (1)

☞ Character model for CHISE Project

☞ Set of Features / Examp
– Is NOT of Attributes
– Each feature has its sou
☞ Character Object as *kha*
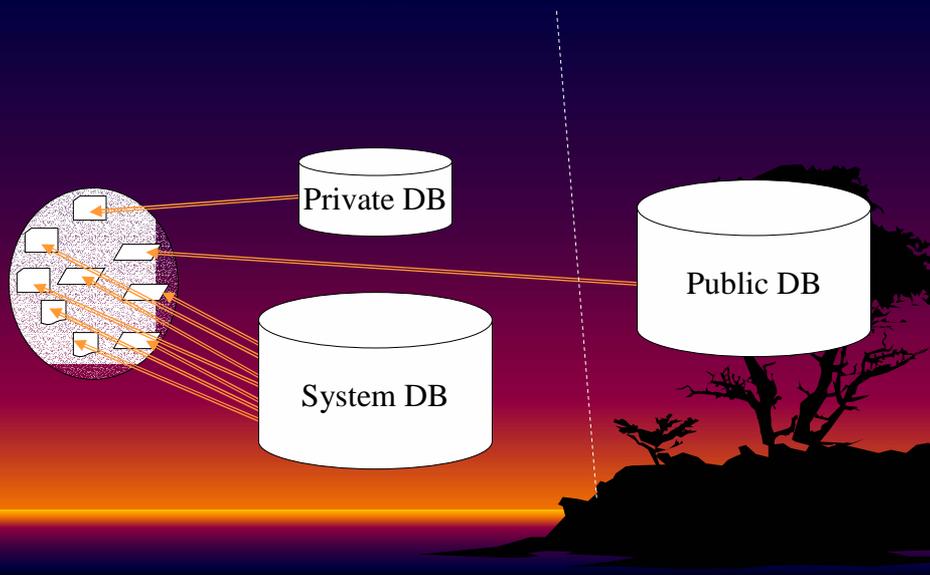– Is NOT of Object-orient Model
– cf. Moro 2003

kichi, yoshi

j í

good

lucky

U+5409

kil

Chaon Model (2)

☞ Comparison of characters = Set operation

feature

character object

Chaon Model (3)

Private DB

Public DB

System DB

Chaon Model (4)

☞ Set of Surfaces

Visual Surface

Pronunciation Surface

feature

Character-code Surface

Semantic Surface
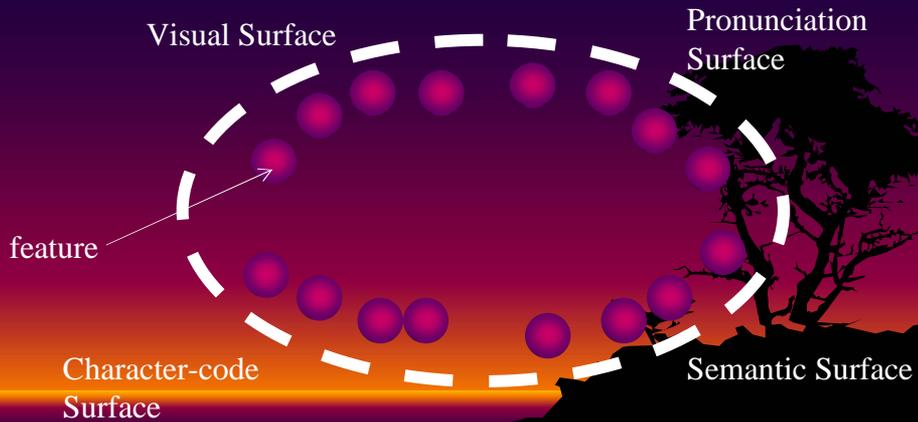
Join CHISE Project !

☞ CHISE Project (http://cvs.m17n.org/chise/)
  – Open source development
    • Character Databases
    • XEmacs CHISE
    • libchise
    • Ruby/CHISE
    • Perl/CHISE
    •   /CHISE
    • KAGE
  – Mailing lists (Japanese / English)

Works Cited

- Aristotle. *Topic.*
- Aristotle. *Metaphysics*
- Derrida, Jacques. "Signature Événement Contexte." *Marges*. Minuit. 1972.
- Moro, Shigeki. "Chaon   (Past, Present and Future of Chaon Model)." CHISE Symposium 2003. (http://www.kanji.zinbun.kyoto-u.ac.jp/projects/chise/symposium2003/)
- Unicode Consortium. *The Unicode Standard, Version 4.0.* Boston: Addison-Wesley. 2003.