

Chaonモデルの過去・現在・未来

師茂樹 (花園大学)

1

2003-03-15

CHISE Symposium 2003

Table of Contents

- 人文科学における文字の本質についての議論
- 集合としての文字と文字コード
- Chaonモデル

2

2003-03-15

CHISE Symposium 2003

Table of Contents

人文科学における文字の本質についての議論

- 集合としての文字と文字コード
- Chaonモデル

3

2003-03-15

CHISE Symposium 2003

“一般文字学”の必要性 (1)

- 言語学: 音声言語中心
 - 欧米では表音文字を使っているから?
 - 音声言語のほうが歴史的に古い?
- 文字学: 歴史的・文字研究・学術的研究
- 漢字学: 独自の長い伝統

4

2003-03-15

CHISE Symposium 2003

“一般文字学”の必要性 (2)

- 様々な提案
 - 後藤朝太郎「文字学の建設」
 - 西田龍雄: 記述文字学・歴史文字学・比較文字学
 - 犬飼隆: 字素論・字態論・統字論

5

2003-03-15

CHISE Symposium 2003

文字言語の独自性

- 空間と時間を超える間接的伝達
- 作者と読者のコンテキストの非共有
- 変遷の非歴史性
 - cf. 「声の文化」と「文字の文化」
 - cf. 「思考を抑圧する文字文化」
 - cf. 視覚言語(手話)

6

2003-03-15

CHISE Symposium 2003

文字の「同じものの性」と「同一性」

- コンテキストの中での「同一性」
- コンテキストを超える「同じものの性」

コンテキストを超えるということ

- 誤読
 - 創造的な読み
 - 劣化
 - 誤配
- } の可能性

要素の集合としての漢字 (1)

- 六書
 - 特に形声と会意
 - cf. 西田龍雄「世界のすべての文字に適用できる原理」



要素の集合としての漢字 (2)

- 形・音・義
 - 野村雅昭「字体素・音素・意義素」
- 字体・字形関連は多数
 - 「筆画」「部件」「整字」
 - 「字体」「字体単位体」「字体エレメント」
 - etc...

Table of Contents

- 人文科学における文字の本質についての議論
 - 集合としての文字と文字コード
- Chaonモデル

集合としての文字

- 内包的表現
 - {x | x = 同じ「骨」と判定される文字の集合}
- 外延的表現
 - {骨, 骨, 骨, 骨, 骨, ...}

JIS X 0208:1997の躰き

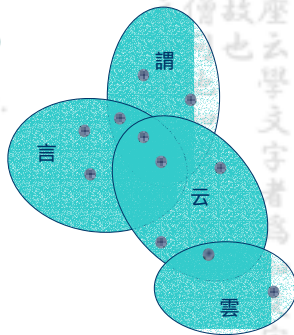
- 規定は放棄したものの実体視されている
 - そもそも規格書である
 - 唯一のコード値
 - 「一般に用いられている漢字」との対応が電子化されていない
 - 字体の変更という歴史的経緯

Table of Contents

- 人文科学における文字の本質についての議論
- 集合としての文字と文字コード
Chaonモデル

Chaonモデル (1)

- 文字素性の集合による文字表現
 - 「一般に用いられている漢字」との対応をモデル化
 - 器としての文字オブジェクト



Chaonモデル (2)

- コンテキストによる変化
 - 素性の見え方の変化
 - 新たな素性の追加

素性の集合演算 Chaonの文字処理

- 比較
 - 排他 / 部分集合 / 一致
- 和集合
 - 文字の同一視
- 積集合

Perl/CHISEによる文字処理

- 文字を「素性の束」として扱う
 - 文字オブジェクト
 - HASH配列へのリファレンス
 - 呼び出し
 - 生成
 - 比較
- 文字データベースへのアクセス

Chaonモデルへの批判 (1)

- クリプキの固有名論
 - フレーゲ・ラッセル・サルらの固有名論
 - 固有名 確定記述(の束)
 - シャカムニ {仏教を聞いた人, 父は浄飯王, ...}
 - クリプキの批判
 - 「シャカムニ」は仏教を開かなかったかもしれない
 - 「仏教を聞いた人」は仏教を開かなかったかもしれない?

Chaonモデルへの批判 (2)

- Chaonモデルの場合も...
 - 「草」の総画数が10画の場合もある
 - 「画数が9画の文字」が10画の場合もある?

多重所属としてのChaonモデル

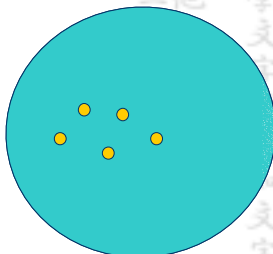
- 素性の束 確定記述の束
 - 複数のコンテキストへの多重所属
- 素性の交換可能性
 - 可能世界としての文字データベース

Chaonモデルにおけるコンテキスト

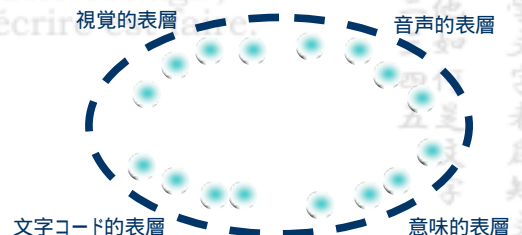
- 既存のコンテキストにおける振る舞い
 - 日本語、中国語...
 - 形・音・義のどれが重視されるか
- まったく新しいコンテキストにおける振る舞い
 - 文字オブジェクトに対する素性の追加
 - 弱いコンテキストと強いコンテキスト

既存のコンテキストにおける振る舞い

- コンテキストにしたがって、素性が背後に隠れ、見えにくくなる
 - 見えなくなるわけではない
 - 何が隠れるかについてはTopicMaps



表層と深層



視覚的表層の突出 (1)



31 2003-03-15

CHISE Symposium 2003

視覚的表層の突出 (2)

KPS 9566 マンセ-
V



32 2003-03-15

CHISE Symposium 2003

音声的表層の突出

- 音声入力 音声合成
- 視覚表現を伴わない“文字”

33 2003-03-15

CHISE Symposium 2003

文字コード的表層の突出？

- ¥/バックスラッシュの同一視
- リダイレクトなど
% perl hoge.pl | sort +1 > sorted.txt
- 言語と結びつかない文字は「字」に過ぎない？

34 2003-03-15

CHISE Symposium 2003

まったく新しいコンテキストにおける振る舞い

- 弱いコンテキスト
 - ある掲示板で
友
という表現が使われた
- 強いコンテキスト
 - その掲示板が巨大化した
文字素性の追加
TopicMapsへの登録

35 2003-03-15

CHISE Symposium 2003

「文字」「文語」の再定義(案) (1)

- 文字言語は音声言語に従属しない
 - 歴史的にも連続しない？
- 文字 視覚的な図形
 - コンテキストによる視覚的素性の背景化
 - 無文字社会にも「文語」がある

36 2003-03-15

CHISE Symposium 2003

「文字」「文語」の再定義(案) (2)

- 文語
コンテキストの越境を前提とした言語活動
- Chaon文字オブジェクト
= コンテキストを超える素性の器
 - = 文字の同じもの性

Chaonモデルの今後の課題

- 文字列をどう扱うか
 - cat / rat
 - 注釈・註釈 / 注水
- データベースの充実

Perl/CHISEの今後の課題

- もっときっちり作ろう
- 素性を一時的に隠すメソッドの追加
- データベースへの書き込み

個人的な今後の課題

- 「文字」「文語」の再定義についてもっと考える
 - 点字・手話などとの関係
- コンテキストと文字(列)との関係についてもっと考える
- 統計的テキスト処理への応用
 - Nグラムなど