

言語研究センター第1回シンポジウム

コーパスに基づく言語研究の現在

2004年1月24日
麗澤大学1号棟 1603教室
主催: 麗澤大学言語研究センター
URL: <http://www.fl.reitaku-u.ac.jp/LINC/>
e-mail: LINC@reitaku-u.ac.jp

ディスカッション

「コーパスに基づく言語研究の推進のために」

- アドヴァイザー: 後藤 斉氏, 滝沢 直弘氏
- 司会: 千葉 庄寿 (麗澤大学外国語学部・言語研究センター研究員)

2

ディスカッションの目的

- コーパスを利用している, 利用したいと考える言語研究者のための共通認識づくり (ブレインストーミング)
 - 検索・分析のコツ
 - 個人に必要なスキル
 - コーパス・ツール
 - 組織としてのコーパス研究体制づくり
- ディスカッションを総括し, LinCホームページ上で報告

3

進行

- 研究の進め方について
 - コーパスを上手に使うには
- ツール・プログラミングについて
- コーパスそのものについて
 - 日本語コーパスの言語とこれから
- 研究体制作りについて

4

はじめに: コーパスと言語研究の関係

- コーパスに基づく言語研究
 - ✂ コーパス言語学
 - ?
- コーパスを研究にどう位置づけるか?

5

1: 研究の進め方

- 研究者個人に必要なとされるコーパス利用のノウハウとコンピュータのスキル
 - ここではコーパスの種類や形式, 利用するツールなどは考えない
- コーパスの利用
 - A) 用例検索のポイント
 - B) 必要なコンピュータの知識・スキル
 - C) 工学系の人たちに協力してもらう際に必要な知識

6

A) 用例検索のポイント

- 検索したいことを形式的に述べることができるか (「形」を問題にできるか)
 - コンピュータの特性: あいまいでない命令が必要
- 必要な知識
 - コンピュータの基礎知識
 - 言語学的知識
 - 法律倫理的知識

7

B) 必要なスキル

- 利用するツール・サービスによる

ツール	データ
Webブラウザ	テキストのみ
特定のソフトウェア	形態解析入りデータ
Unix コマンド	構造化されたデータ

- 共通の知識としてあったほうがよいもの
 - 正規表現
 - テキストファイルの基礎知識と扱い方
 - Unicode の知識

8

C) 工学系の技術者との協力

- 「会話」が成り立つ程度の技術的な最低限の知識
- むしろ言語の分析能力が必要

9

2: ツール, ソフト・プログラミングについて

- ツールは必要?
- プログラミングの必要性
 - 誰が? どうやって?
- ツールの多言語対応の必要性

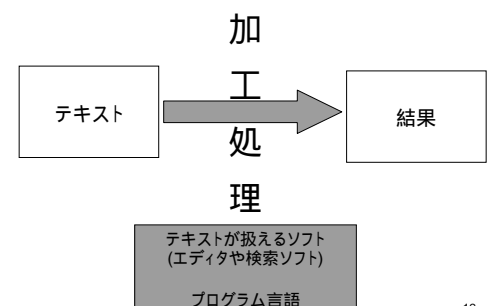
10

プログラム言語の利用

- 必要なプログラミングのスキルのレベルはいろいろ
- 個人として
 - テキストデータの検索・加工
- 共通で使うツールとして

11

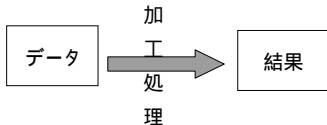
テキストデータの検索・加工



12

共通のツール, サービス

- 多くの人が簡単に利用できるようなツールやサービス
 - GUIツール
 - Webベースインターフェース (CGIなど)
 - データベース
- 高度なプログラミングが必要
 - セキュリティ
 - バグの除去
 - 高速化
 - マニュアルの作成



処理方法は同じ, しかしプログラミングの要求は高い。

13

3: コーパスそのもの

- 日本語コーパスの現状とこれから
- インターネット上のデータの利用
- コーパスの作成のコツ
 - ある記号を1つの文字で入力する: 文字とコードの一対一対応

14

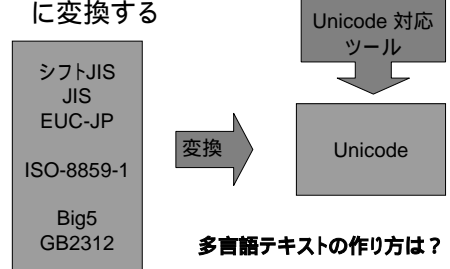
Unicode ベースのコーパスの時代にむけて

- Unicode の利点
 - いろいろな言語のテキストを同じように扱える: Unicodeに対応しているツールであれば, すべての言語のテキストを処理できる。
- Unicode の注意点
 - Unicode に対応したシステム(OS・フォント・ツール)が必要
 - 同じような記号・文字が複数ある
 - 同じ文字の表し方が複数ある

15

Unicode データの作成と利用

- 複数の文字バリエーションの混入を防ぐため, 単言語のデータはその言語・地域のエンコードで作成し, Unicodeに変換する



多言語テキストの作り方は? 16

インターネット上のデータの利用

- コーパスとしての利用の可能性と限界

17

組織としての研究体制作り

- 何を備えるか
 - コーパス
 - 利用環境, ツール
- 何を教えるか
 - シラバス
- どんな人的資源を配置するか
- 組織間でどんな研究ネットワークを作るか

18