

Corpus-Based Language Studies

UNIT A1 CORPUS LINGUISTICS: THE BASICS

英語專攻4年
浄泉 和博

まとめ (p.11)

本章ではコーパス言語学における基本的な事項を学びました。

たとえば...

- ①コーパス言語学の歴史のあらまし
- ②現代の言語学における「コーパス」の定義
- ③言語の研究でコンピュータを活用する利点
- ④intuition-based approachとcorpus-based approachの比較
- ⑤なぜコーパスは方法論として扱われるのか
- ⑥corpus-based linguisticとcorpus-driven approachの違いは何か

A1.1 Corpus Linguistics

Past and Present

Pre-Chomskyan period...

紙媒体のコーパスを活用

Not representative

実験的で観察データに基づく



“Corpus-based”

In the late 1950s...

コーパスの役割

音声の特徴の研究

文法の研究にコーパスを活用する学者も

Jespersen

Fries

In the early 1960s...

コンピュータの発達



膨大なデータ処理が可能に

アメリカ英語のmodern corpusの登場

“the Brown University Standard Corpus of Present-day American English”

A1.3 What is a Corpus?

Corpus

= a body of naturally occurring language

Sinclair, J

A corpus is a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language.

- × テキストのランダムな寄せ集め
- 言語の情報として様々なannotationがつけられた産出データのサンプル集

Corpusの4要素

- ① machine-readable
- ② authentic texts, which is
- ③ sampled to be,
- ④ representative of a particular language or language variety

Specialized Corporaはコーパスと言えるのか
ex) the Lancaster Corpus of Abuse

A1.4 Why Use Computers

to Study Language

- ① 速い
- ② 簡単
- ③ 正確
- ④ 信頼性
- ⑤ 量

A1.5 the Corpus-Based Approach vs. the Intuition-Based Approach

the Intuition-Based Approach



- ・分析に向けた例文
- ・言語外の要因の影響を受けない



- ・方言の影響を受けている可能性がある
- ・直感的に正しいと判断されても典型的な表現と言えるか
- ・分析結果の立証の難しさ

the Corpus-Based Approach



- authenticなテキスト
- 直感では見落としうる違いを発見
- 信頼性のある量的データ

corpusとintuitionのバランスが大事！

A1.6 Corpus Linguistics

A Methodology or a Theory?

言語学

統語論

音韻論

意味論

語用論

コーパス

コーパス

言語学

統
語
論

音
韻
論

意
味
論

語
用
論

A1.7 Corpus-Based vs.

Corpus-Driven Approaches

Corpus-Based Approach

既存の理論 → データで検証

Corpus-Driven Approach

データを分析 → 理論を構築

Types of Corpora Used

Corpus-Based Approach

Attitudes toward Existing Theories

データで検証

Focuses on Research

Corpus-Driven

データを分析

理論を構築

Paradigmatic Claims

まとめ (p.11)

本章ではコーパス言語学における基本的な事項を学びました。

たとえば...

- ①コーパス言語学の歴史のあらまし
- ②現代の言語学における「コーパス」の定義
- ③言語の研究でコンピュータを活用する利点
- ④intuition-based approachとcorpus-based approachの比較
- ⑤なぜコーパスは方法論として扱われるのか
- ⑥corpus-based linguisticとcorpus-driven approachの違いは何か