中日事象表現データベースの構築について

張勤[†], 星英仁[‡], 沈力[‡], 金明哲[‡]
[†]中京大学, [‡]同志社大学
qin@letc.chukyo-u.ac.jp(張勤)

本研究は中日事象表現データベース作製のための基礎準備をするものである。基本構想は 2007 年夏より練られ、以後、星と沈より日本語と中国語について項目の案が示され、金よりデータベース構築技術についてのアドバイスが出され、張においてデータを抽出する資料の選定が行われた。2008 年春より、「中国語と日本語の対照に基づく事象表現の総合的な研究（研究課題番号 19320064）」の分担金を受け、張において、上記の項目案や資料候補に対して修正を加えながら、データ抽出・構築の作業を開始し、現在データ抽出の作業がほぼ完了した。

1　中日事象表現データベースの構築目的

このデータベースの構築は以下の目的に基づいている。

1）中国語教育に実用のデータベースを供する

2）中国語学・言語学研究に日本語と対照することができるデータベースを供する

2　データベースの項目

初期の段階では、項目が中日両語の事象表現に限定されていた。2008 年春以降の具体作業において、主に中国語教育分野においてより実用性が高まるものとなるように、介词・助詞などをできるだけ広く収めることになった。その結果、中国語は 107 項目、日本語は 80 項目となった（ハンドアウトを参照）。

3　データ抽出用の資料

資料は、現代小説から以下の基準で選定を行った。

中国語

1）北方方言地域の出身の作家であること

2）女性作家もいること

3）現実の生活が題材にされ、リアリズムな手法で書かれていること

4）1945 年ごろから近年までの間に出版されたものを均等に選出する

5）日本語訳があること

日本語

1）できるだけ大きな方言地域をカバーすること。

2）女性作家もいること

3）現実の生活が題材にされ、リアリズムな手法で書かれていること

4）昭和から平成までの間に出版されたものを均等に選出すること

5）中国語訳があること

（資料一覧はハンドアウトを参照。）

4　データベースの構造

一つの項目につき、55 例を無作為に抽出する。

そのうち、5 例について、タグ・クロス・訳文をつける。残りの 50 例には訳文のみつける。

（項目例はハンドアウトを参照。）

5　特記事項

1）データ抽出の合理性

2）項目定義の依拠

3）項目の範囲（データベースの守備範囲）

6　今後の展開

1）現段階のデータを完備する

2）実用に向けて検索のフレームを確定させる

3）公開の方式を模索する

7　参考文献（データベース構築関係部分）

2008 石川慎一郎『英語コーパスと言語教育』大修館書店.

2003 ダグラス・バイバー他著／齊藤俊雄他訳『コーパス言語学』南雲堂.

1995 岡田毅『実践「コンピュータ英語学」』鶴見書店.

キーワード：中国語・日本語・事象表現・言語教育・データベース・検索