

# 日中対照漢字語データベースの開発と応用

松下達彦(東京大学)・陳夢夏・王雪竹・陳林柯(一橋大学大学院生)

## 1. 背景と目的

意味・用法の異同の記述的先行研究

意味対応のパターンを量的に詳しくとらえたものではなく、分類は現在の使用事実と対応しないところが見受けられる

数や意味対応を数量的に扱っている先行研究

コーパスの規模が小さく、対象の異なり語数も1万語に満たない

目的

- 日本語の学習・教育・研究により直接的に役立つデータベースの開発
- ⇒意味対応の判定結果のほか、字体の違いや意味のズレなども検索、授業の予復習、テスト作成などに活用
- ⇒漢語(字音語)の日中両語の意味対応パターンの量的分布を調査

中国語との違いなどを逐一検索することが困難

従来の対照研究の結果が十分に教育現場で応用されていない

## 2. 内容および作成方法

**対象**  
『日本語を読むための語彙データベース』(松下2011)の留学生用語彙ランクの上位2万語に含まれる漢語10054語

**項目**

- ・ID, 留学生用語彙ランク
- ・見出し語, 文字数
- ・標準的(新聞)表記
- ・標準的読み方
- ・品詞, 語種, 中国語表記
- ・意味対応
- ・共通義と独自義
- ・意味の類推可能性

**意味対応の分類**

- ・同形同義(日=中)
- ・同形異義(日≠中)
- ・同形類義(日>中)
- ・同形類義(日<中)
- ・同形類義(日><中)
- ・非同形( $\varphi$ )

**意味対応の分類の手順**

1. 日本で日本語学や日本語教育を専攻する中国語母語の博士課程大学院生3名A, B, Cのうち, 2名が独立して第1次判定
2. 一致したもの → 確定  
一致しないもの → 独立した第3判定者が第2次判定(第1次判定の一方と一致した場合はそれで確定)
3. 3者が異なる判定 → 3名のうちの2名以上の合議で第3次判定

## 3. 分析結果と考察

作業過程での一致率

表1 第1判定者ペア別一致語数・一致度

	A-B	A-C	B-C	合計
判定語数	3352	3351	3351	10054
一致語数	2567	2580	2490	7637
一致率(%)	76.6	77.0	74.3	76.0
Kappa係数	0.55	0.58	0.52	

判定はある程度一致が、易しくはない

特定の判定者の影響はない

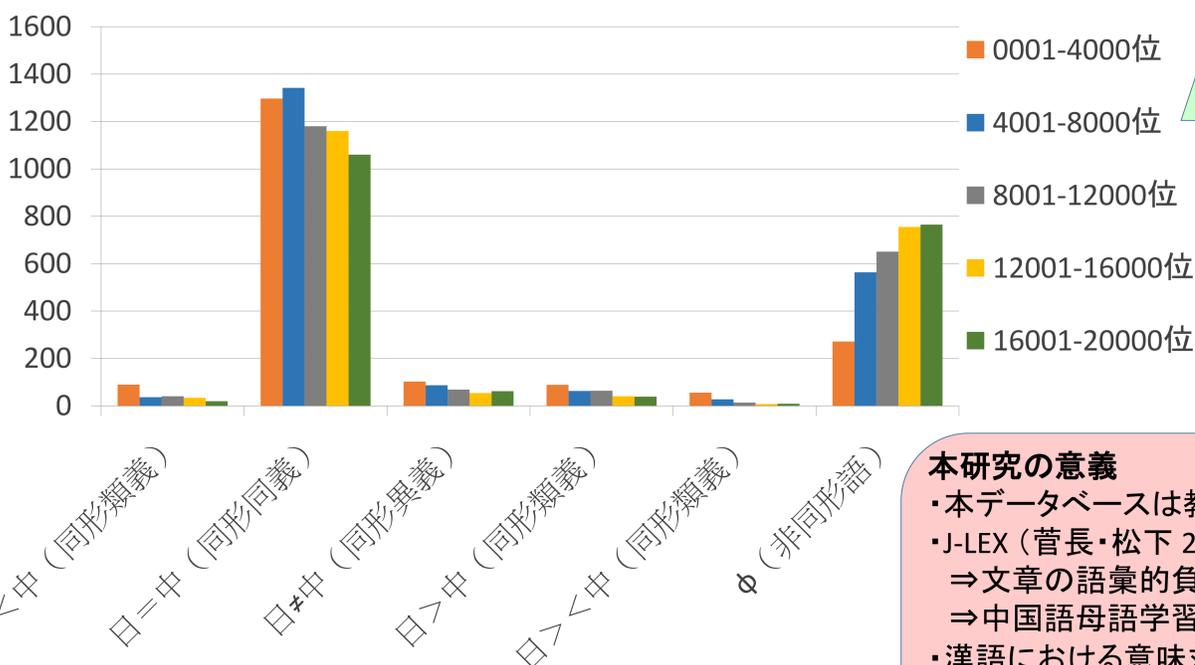
- ・同形語は86%が、非同形語は75.0%が第1次判定で確定
- ・同形同義は判定が一致しやすい
- ・不一致のうち, 56.4%が、判定者のどちらかが非同形語と判断した語 **そもそも中国語に存在するかどうかで判断が分かれやすい**
- ・どちらかの意味が広い語は、第2次判定で確定するものが多い
- ・日><中のパターンは第3次判定までもつれ込んだものが多い **かなり上級レベルの判定者でも両語に存在する意味の認定は容易でない**

表2 第1判定で不一致だった語の不一致パターン別/判定者ペア別の語数・割合

\* 予稿集の表2に誤りがありました。本表が正しい表です。

意味対応の分類	</=	</≠	</>	</><	</ $\varphi$	=/≠	=/>	=/><	=/ $\varphi$	≠/>	≠/><	≠/ $\varphi$	>/><	>/ $\varphi$	></ $\varphi$	計
A・B不一致	86	10	8	5	17	61	103	23	365	7	7	44	6	34	9	785
A・C不一致	95	12	10	5	12	65	108	12	341	19	8	50	6	27	1	771
B・C不一致	70	9	9	4	15	82	157	22	357	23	9	52	12	32	8	861
不一致合計	251	31	27	14	44	208	368	57	1063	49	24	146	24	93	18	2417
不一致語数に占める割合(%)	10.4	1.3	1.1	0.6	1.8	8.6	15.2	2.4	44.0	2.0	1.0	6.0	1.0	3.8	0.7	100.0

漢語, 同形語の割合, および同形語に占める意味対応の分布



- ・各レベルの漢語の割合は5割前後で安定
- ・漢語の70%(全体の35%)が同形語
- ・漢語の60%(全体の30%)が同形同義
- ・漢語の10%(全体の5%、同形語の14%≒1/7)が **同形類義&同形異義 = 要注意語**
- ・漢語の30%(全体の15%)が非同形語
- ・同形同義は中級前半で一番多い
- ・レベルが上がるにつれて同形同義、同形類義、同形異義は減り、非同形語が増える
- ・副詞に非同形が多い(45%)
- ・接辞に非同形は少ない(接頭辞2%、接尾辞10%)

**本研究の意義**

- ・本データベースは教師, 学習者, 研究者が語の検索など直接利用できる
- ・J-LEX(菅長・松下 2014)のような語彙頻度プロファイラーに搭載できる
- ⇒文章の語彙的負荷を中国語母語と非母語に分けて表示
- ⇒中国語母語学習者にとっての要注意点を画面表示
- ・漢語における意味対応パターンの量的側面を2万語レベルの規模で初めて明らかにした

**課題**

- ・データベースの判定や記述には、まだ改良の余地あり。随時、アップデート。

データベースのダウンロードURL(「松下」「言語」で検索):

<http://www17408ui.sakura.ne.jp/tatsum/database.html#cvd>

本研究は日本学術振興会科学研究費, 基盤研究A(課題番号 25244022, 研究代表者: 庵功雄), および基盤研究B(課題番号 17H02350, 研究代表者: 庵功雄)の助成を受けたものです。