

日中対照漢字二字熟語データベースの構築と語彙特性の分析に関する研究

小森和子(明治大学)・早川杏子(関西学院大学)・李在鎬(早稲田大学)・玉岡賀津雄(名古屋大学)
共同研究者:三國純子(文化学園大学)

1. 本研究の意義・目的

- 既存の資料**:
 - ・文化庁(1978)『中国語と対応する漢語』
 - ・約2,000の日本語の漢語の日中の意味的対応関係
(①同形同義(S), ②同形類義(O), ③同形異義(D), ④日本語にしかない漢語(N))
- 問題の所在**:
 - ・日本語, 中国語における**意味的変容**に伴う**対応関係の再検討**の必要性
 - ・研究や教育の用途としての**語彙特性の拡充**の必要性(意味だけでなく, 二言語間の形態的・音韻的異同, それぞれの言語における語彙使用頻度, 新たな指標に基づく語彙難易度etc.)
- 本研究の目的**:日本語と中国語の対照研究に基づく日本語の習得**研究**や日本語**教育**のための**基礎資料**(データベース)の構築
- 本発表**:構築中のデータベースの**一部紹介**, データベースの全体像を把握するための計量的**分析結果**の報告

2. データベースの構成

●本研究のデータベースの構成:日本語(N=2,078)と中国語との対照

一次資料:朴・熊・玉岡(2014)(旧JLPT4~2級の漢字二字熟語(N=2,058), 日本語・中国語・韓国語)

★日本語の語彙特性

- ①**見出し(=読み)**(例「あした(明日)」,「あす(明日)」,「みょうにち(明日)」), ②**漢字表記**
- ③**音訓情報**(O:音読み, K:訓読み, ZF:常用漢字付表, SR:熟字訓等), ④**難易度**(日本語教育語彙表(ver1.0))他
- ⑤**使用頻度**(大規模コーパス:Sketch Engine他, 新聞コーパス(朝日, 毎日)), ⑥**逸脱率**(荒牧他(2012))
- ⑦**語義, 品詞, 用例**(『新明解国語辞典(第7版)机上版』による, 語義毎に品詞と用例)

☆中国語の語彙特性

- ⑧**表記**(簡体字), ⑨**読み**(ピンインと四声) ⑩**使用頻度**(Sketch EngineとBCC(北京語言大学の中国語コーパス))
- ⑪**語義, 品詞, 用例**(『現代汉语词典(第6版)』, 『現代汉语规范词典(第3版)』)

★★日中の対応関係

- ⑫**日中書字異形度**(茅本(1996)の指標:「0」(完全一致)~「4」(完全不一致), 但し一部改変あり)
- ⑬**日中音韻類似度**(早川・于・初・玉岡(2017)の指標:Rのsdists関数により単語間の重み付け編集距離を計算)
- ⑭**日中意味一致度**(文化庁(1978)の分類, S(同形同義)O(同形類義)D(同形異義)N(日本語のみ))

👉**今後, 語義情報に基づいて, 日中意味対応関係を判定し, タグ付けを行う予定**

3. 計量的分析の結果

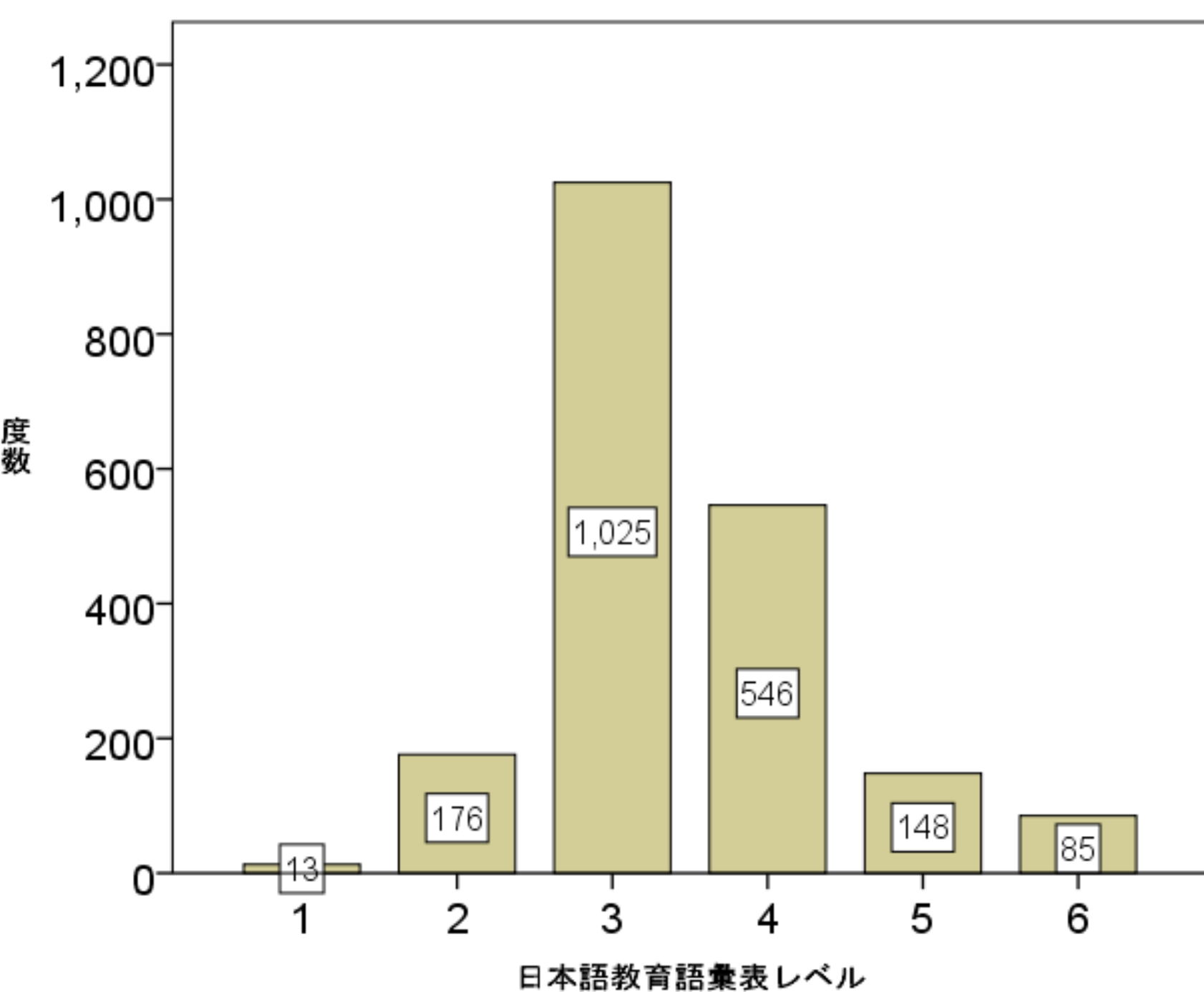


図1 度数(難易度別)

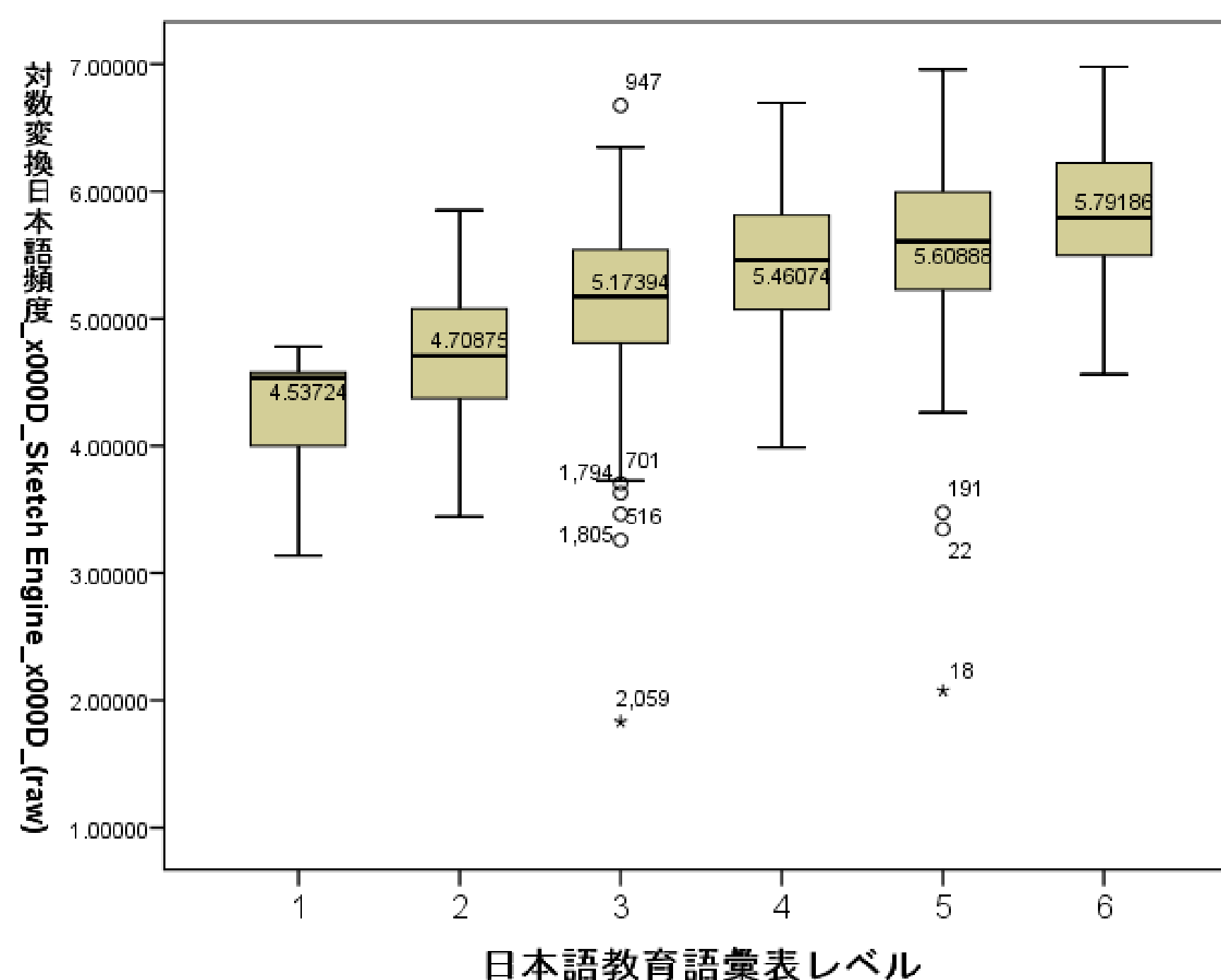


図2 使用頻度(難易度別)

表1 語彙難易度を予測する変数(重回帰分析)

変数名	標準偏差	β	t値	有意確率
予測変数				
語彙難易度(日本語教育語彙表)	0.907			
説明変数				
日本語使用頻度(Tono et al., 2013)	0.326	10.395	***	
中国語使用頻度(Sketch Engine)	-0.354	-5.117	***	
日本語使用頻度(Sketch Engine)	0.133	4.153	***	
逸脱率	-0.076	-3.507	***	
日中書字異形度	-0.072	-3.409	**	
中国語使用頻度(BCC)	0.179	2.582	**	

注1:n=1789. *** p<.001, ** p<.01

注2:決定係数(R²)は,.218であった

●語彙特性間の関係(全2,078語)

- ・難易度とSketch Engineなどの**使用頻度**に相関:難易度が低い語は**使用頻度が高い**(Sketch Engine [r=0.230, df=1993, p<.001]. Tono et al. (2013) [r=0.432, df=1993, p<.001])
- ・難易度を予測する変数:日本語使用頻度, 逸脱率, 書字異形度, 中国語使用頻度

●日中同形語の諸相(1,522語, 全体の73.24%)

- ・難易度と**新聞の使用頻度**に相関:難易度が低い語は**新聞での使用頻度が高い**(朝日新聞 [r=0.244, df=1475, p<.001], 毎日新聞 [r=0.227, df=1467, p<.001])
- ・難易度と**中国語の使用頻度**に相関:難易度が低い語は**中国語の使用頻度も高め**(中国語BCC [r=0.111, df=1473, p<.001])

●高頻度語の特徴

- 日中同形語が多い:
 - ・朝日新聞の上位20語中18語(14語はS語)
 - ・Tono et al(2013)の上位20語中14語(7語はS語)
- 👉**新聞は中国語NSに理解しやすい**

参考文献:予稿集参照. 付記:本研究はJSPS科研費15K02656, 15K16781, 16K02794の助成を受けた。