

# 不同水平日语学习者口语词汇复杂性特征研究

曲滢莹 袁丽杰 帅岩兰

大连理工大学 辽宁大连 116024

**摘要:** 本研究聚焦学习者日语口语能力发展, 分析了不同水平日语学习者词汇复杂性特征。结果显示, 从词汇多样性的角度来看, 随着日语水平的提升, 口语中使用的词汇数量增多, 但为注重表达的准确性及流畅度, 相应会降低词汇复杂性。口语复杂性、准确性和流利性三者相互联结、不断变化。本研究将有助于为日语口语教学提供启示, 帮助教师更有针对性地解决口语教学中所存在的问题, 从而精准有效地提高学习者的日语口语能力。

**关键词:** 第二语言习得; 日语口语; 能力动态发展

近年来, 以培养交际能力为目的的外语教学得到普及, 在课堂教学中语言输出任务的真实性、有效性也得到了关注。伴随二语口语能力受到了重视的同时, 二语习得研究领域中, 二语口语能力也成为了热点议题。随着二语习得的深入, 语言熟练度不断提高, 口语的复杂性、流畅性、正确性随之会发生如何变化, 二语语言知识与二语口语水平之间有何种关系, 口语任务的特征是否会影响二语口语能力的评价结果等课题得以开展。但此领域多以英语二语学习者研究对象, 日语相关研究还处于起步状态, 且多以学习者书面语分析为主。如郭圣琳, 朱一平(2022)探讨了日语学习者书面语的句法动态。关于日语口语能力发展方面的实证研究的数量还是比较匮乏, 研究对象和方法也不够丰富。除此, 大部分研究都聚焦同一水平的学习者, 有必要对不同水平的日语学习者的口语能力进行探讨。

本研究聚焦中国日语学习者日语口语能力发展, 基于大型日语语料库 I-JAS(International Corpus of Japanese as a Second Language), 通过大规模数据, 结合学习者的详细背景信息和日语能力水平, 以词汇复杂性为切入点, 探讨不同水平日语学习者的口语能力特征。以求帮助教师在实际教学中精准了解学习者口语发展的变化趋势, 帮助学习者更有针对性地提高日语口语能力。

## 一、研究方法

本研究使用 I-JAS 语料库 2019 年 5 月公开的日语母语者与 50 名中国日语学习者的自然对话数据, 研究过程中去掉了日语母语者数据, 只保留学习者数据。并以 50 名学习者的 C-CAT 成绩作为评定日语水平的标准, 按照“平均值

$\pm 2SD$ ”计算极端值, 去掉学习者 1 名, 然后去掉得分为中央数值的学习者 5 名。将剩余的 44 名学习者分成高水平组和低水平组 ( $t=11.14, p=.00$ )。

研究视角选取了词汇多样性、词汇密度、词汇难度三种。词汇多样性采用公式“ $Uber = (\log \text{Tokens})^2 / (\log(\text{Tokens} - \text{Types}))$ ”, 词汇密度援引佐野等(2009)提出的公式“词汇密度 = 实词数量 / 级阶小句的总数”, 其中实词和级阶小句依照佐野等(2009)的认定标准, 通过人工认定和手工标注后计算出词汇密度。词汇难度参照 Laufer & Nation(1995)提出的公式“ $LS = \text{Number of advanced tokens} / \text{Total number of lexical tokens}$ ”计算。在此基础上, 结合《日语教育词汇表》(<http://jhlee.sakura.ne.jp/JEV/>)的词汇等级分类来判定复杂词。

## 二、结果和分析

### 1. 不同日语水平学习者词汇复杂性指标差异

表 1 显示了高水平日语学习者的词汇多样性、词汇难度、词汇密度都要低于低水平日语学习者。且通过 t 检验发现, 高水平组和低水平组之间的各项指标的差异具有统计学意义。即随着日语水平的提升, 日语学习者使用的词汇复杂程度反而降低。

表 1 不同日语水平学习者词汇复杂性指标差异

日语水平	词汇多样性		词汇难度		词汇密度	
	平均值	标准偏差	平均值	标准偏差	平均值	标准偏差
高水平	5.90	0.72	0.09	0.01	0.32	0.03
低水平	6.34	0.63	0.10	0.01	0.35	0.03

口语能力包括复杂性, 流畅性、正确性三方面。从流

畅性角度来看,一般认为简单词汇的使用可以提高语言流畅度。当学习者使用常见、熟悉的词汇时,他们更容易构建自然的句子,不需要花费过多的时间思考词汇的选择。而复杂词汇的使用将导致语言表达变得繁琐,进而影响流畅性,学习者需要更多时间来思考如何正确使用这些词汇,这可能会使交流变得不那么流畅。另外从准确性角度来看,使用简单词汇可以降低语法和词汇错误的风险,复杂词汇的使用可能增加语言错误的风险,学习者也许不确定如何正确使用高度复杂的词汇,因此容易出现语法错误或用词不当的问题。相反因为学习者对这些简单的词汇更加熟悉,所以选择使用常见词汇会使他们的表达更准确。综上所述,随着语言水平的提高,学习者有可能更关注流利度和准确性,因此更倾向于使用熟悉和简单的词汇。

### 2. 不同水平学习者的词汇复杂性各指标的相关性

在此,算出学习者整体以及高水平日语学习者和低水平日语学习者的词汇多样性、词汇难度、词汇密度三个指标间的相关性。具体结果如表 2 所示。

表 2 学习者在词汇复杂性各指标上的皮尔逊相关性

指标间的相关性	整体	高水平组	低水平组
词汇多样性与词汇难度	.274+	-.200	.523*
词汇多样性与词汇密度	.048	-.148	-.106
词汇难度与词汇密度	.508**	.171	.509*
(**p < 0.01,*p < 0.05,+0.1 > p > 0.05)			

如表 2 所示,学习者整体上词汇难度与词汇密度呈中相关( $r=.508, p<.01$ ),词汇多样性与词汇难度呈相关倾向( $r=.207, 0.1 > p > 0.05$ );低水平组的词汇多样性与词汇难度( $r=.523, p<.05$ ),词汇难度与词汇密度( $r=.509, p<.05$ )呈弱相关;而高水平组三个指标未呈现任何相关性。可见学习者词汇复杂性指标之间并未观察到由不相关或仅个别指标弱相关向各个指标均相关的方向发展,而相反,低水平阶段观察到了词汇难度与词汇密度和多样性之间的相关,到了高水平阶段相关性消失了。

### 三、结语

本研究从词汇多样性、词汇密度、词汇难度的角度分析了词汇复杂性和学习者水平之间的关系。结果显示,高水平日语学习者的词汇多样性、词汇难度、词汇密度都要低于低水平日语学习者。随着语言水平的提高,学习者有可能更关注流利度和准确性,因此更倾向于使用熟悉和简单的词汇。

但这一现象并不意味着学习者完全放弃复杂词汇,他们可能会在适当的时候使用这些词汇,以丰富他们的表达方式。在不同的语境中,平衡流利度和准确性是学习者需要不断努力的目标。除此,研究结果显示随着学习阶段的发展,学习者词汇复杂性的内部子系统并未出现协同发展的趋势。基于上结果,建议教师在知道高水平学习者口语时,适当引导学习者在保证口语流畅性和准确性的基础上,关注词汇的复杂性。通过提升词汇的复杂性来促进口语能力的发展。

### 参考文献

- [1] 鲍贵.二语学习者作文词汇丰富性发展多维度研究[J].外语电化教学,2008(5):39-45.
- [2] 鲍贵.不同课程水平英语学习者词汇复杂性研究[J].解放军外国语学院学报,2011(4):57-62+129-130.
- [3] 王海华,周祥.非英语专业大学生写作中词汇丰富性变化的历时研究[J].外语与外语教学,2012(2):43-47.
- [4] 于涵静.英语学习者口语复杂性、准确性的动态发展研究[J].外语与外语教学,2019(2):104-114+154.
- [5] 于涵静,s 彭红英,周世瑶.英语学习者口语复杂性动态发展趋势研究[J].现代外语,2022(1):94-105.
- [6] 郑咏滢.基于追踪个案的二语书面语法复杂性发展研究[J].中国外语教育 2017(4):34-41+89.
- [7] 郑咏滢.高水平学习者语言复杂度的多维发展研究[J].外语教学与研究 2018(2):60-71+162.
- [8] 金銀姫.タスクの認知的複雑さが中国人日本語学習者の独話に与える影響—流畅さ,複雑さ,正確さ,語彙の豊富さを用いた分析—[J].第二言語習得研究会,2020,68-73.
- [9] 田島ますみ,深田淳,佐藤尚子,玉岡賀津雄.語彙指標数値と文章主観評価の関係:日本人大学生による2種類の書き言葉コーパスを使った実証研究[J].中央学院大学人間・自然論叢,2009(29):57-77.
- [10] 田島ますみ.KY コーパスを用いた語彙的複雑性の測定に関する研究—語彙の多様性及び密度と言語運用能力との関連[J]The tenth Princeton Japanese Pedagogy Workshop proceedings,2002,94-104.
- [11] 田島ますみ・深田淳・佐藤尚子.語彙多様性を表す指標の妥当性に関する研究—日本人大学生の書き言葉コーパスの場合[J]中央学院大学社会システム研究所

紀,2008,9(1):51-62.

[12] 建石由佳,小野芳彦,山田尚勇. 日本文の読みやすさの評価式 [J]. 情報処理学会研究報告ヒューマンコンピュータインタラクション (HCT)25,1988:1-8.

[13] 堀場裕紀江,金銀姫,松本順子. L2 発話に言語習熟度と語彙知識はどのように関わるか—流暢さ・複雑さ・正確さ・語彙の多様性の観点から— [J]. 言語科学研究: 神田外語大学大学院紀要,2021(27):105-123.

[14] 水本篤. 自由英作文における語彙の統計指標と評定者の総合的評価の関係 [J]. 統計数理研究所共同研究リポート 215 「学習者コーパスの解析に基づく客観的作文評価指標の検討」,2008,15-28.

[15] Carroll J. B. On sampling from a lognormal model of word-frequency distribution [C]. In H. Kucera & W. N. Francis

(Eds.), *Computational analysis of present-day American English*. Providence, RI: Brown University, 1967, 406-424.

[16] Read J. *Assessing Vocabulary* [M]. Cambridge: Cambridge University Press, 2000.

[17] Skehan, P., & Foster, P. The influence of task structure and processing conditions on narrative retellings [J]. *Language Learning*, 1999, 93-120.

[18] Verhallen, M. and Schoonen, R. Lexical knowledge in L1 and L2 of third and fifth graders [J]. *Applied Linguistics*, 1998, 19(4):452-470.

#### 作者简介:

曲滄莹 (2004 年 10 月 16 日), 女, 汉族, 辽宁省瓦房店市, 学生, 本科, 二语习得, 大连理工大学, 辽宁省大连市甘井子区凌工路 2 号