

## 汉语二语者同素单双音词语混淆分布特征及其生成机制

付冬冬\* · 朴炳仙\*\*

### 【目 录】

1. 引言
2. 同素单双音易混淆词的确定及其整体混用情况
3. 二语者同素单双音词语的混淆特征分析
4. 同素单双音词语混淆的生成机制
5. 结语

### 【摘要】

本文基于大规模汉语中介语语料库,综合考量词语混用的绝对频次和相对频度,筛选出混淆度排名靠前的27组同素单双音易混淆词,通过数据对比分析发现,同素单双音词语的混淆分布特征为:共通性强,同素单双音易混淆词基本上全为同词类混淆词语,其中动词多发生语体搭配误用,名词多发生韵律搭配误用和语体搭配误用,形容词多发生语义混用;双向误用的易混淆词占据绝对优势;同素单双音易混淆词的语义关系均较近。我们认为,一语标义词位的调节作用,以及汉语二语者心理词库中缺少词语的韵律信息和语体信息,是造成同素单双音词语混淆的主要原因。比较而言,词语的韵律混淆和语体混淆最难克服。

【关键词】词语混用;同素单双音易混淆词;汉语二语者;韵律混淆;语体混淆最。

\* 付冬冬,第一作者,上海财经大学国际文化交流学院讲师 (296478669@qq.com)。

\*\* 朴炳仙,通讯作者,韩国国立群山大学中文系教授 (parkbs@kunsan.ac.kr)。

## 1. 引言

在现代汉语中,同素同/近义单双音词是较为特殊的词语类聚,相关研究成果丰富(程娟、许晓华2004;刘智伟2005;季瑾2005;吴颖2009;骆健飞2017等)整体上看,已有研究多从汉语本体角度进行,以二语者词语偏误为研究基点的文章较少。

刘春梅(2007)从中介语视角归纳了同素同义单双音名词的偏误类型及其成因;于洋(2015)则探讨了二语者同素单双音名词的混淆特征及其成因。从词语偏误到聚焦词语混淆现象,基于中介语的词汇研究进一步细化和深入。

然而,以往研究还有不足:1)针对二语者的单双音名词的考察较丰富,而缺少对同素单双音形容词、动词习得情况的观察,也就难以发现二语者不同类型的词语混淆特征和规律。2)缺少对不同学习群体同素单双音词语使用真实情况的横向对比,因此,哪些混淆现象具有共性特征,哪些混淆现象是学习者的个性特征就无从判定。3)以往对二语者词语混淆成因分析多从语言层面开展,而二语者心理层面的深层因素却少有探讨。

不同群体的二语者同素单双音词语混淆特征表现有无差异?二语者词语混淆现象心理层面的生成机制是什么?不同类型的同素单双音易混淆词有无难度等级?这些问题,只有在对不同母语背景的二语学习者的词语混淆分布状况进行全面调查和对比分析的基础上才能弄清。

鉴于此,本文将基于北京语言大学“HSK动态作文语料库”、“汉语中介语语料库系统”以及“不同母语背景的汉语学习者词语混淆分布特征及其成因研究”项目组<sup>1)</sup>(以下简称“项目组”)采集的汉语中介语语料,对比分析日语、韩语、英语背景汉语学习者同素“单音词—并列式双音复合词”(以下简称“同素单双音词语”)混淆的分布特征,尝试探析二语者词语混淆的深层心理机制,并探明不同类型词语混淆的克服难度。

## 2. 同素单双音易混淆词的确定及其整体混用情况

同素单双音词语的选取源于“不同母语背景的汉语学习者词语混淆分布特征及其成因研究”项目组的准易混淆词词表<sup>2)</sup>。

1) “不同母语背景的汉语学习者词语混淆分布特征及其成因研究”项目由北京语言大学张博教授主持,30多位教师、中外博士生和硕士生参与研究。

2) 该词表是由项目组中有丰富对外汉语教学经验的一线教师、精通汉语的外国学者结合教学经验并根据调研和测试编制而成的准易混淆词词表。

首先,从词表的1264组词条中筛选出共有相同语素的“单音词—并列式双音复合词”<sup>3)</sup>114组。然后,在中介语语料库中对这些词目进行逐一检索,提取全部语例,删除无关用例并进行人工判定<sup>4)</sup>。之后,采用张博(2013)提出的兼顾混用绝对频次和相对频度的综合方法<sup>5)</sup>,对各组词语的混淆程度进行排序。

“绝对频次”指在同一语料库中相关词语因彼此不当替代而误用的次数,该数据用于跟其他词语的混用次数进行横向比较;相对频度指在同一语料库中词语混用的次数占该词及该组词总频次的比重,该数据用于跟词语的正确用法进行比较。(张博,2013)通过观察所有准易混淆词语的统计数据可发现,词语混用的绝对频次高,其相对频度不一定高,反之亦然。例如,“国⇌国家”的绝对频次为112,相对频度为1.58%,而“皮⇌皮肤”的绝对频次虽然仅为4,但是相对频度却达到2.86%。

我们认为,纳入研究范围的易混淆词语应该具有典型性和代表性。结合词语的混用实际,我们将绝对频次大于15、或相对频度高于1%的混淆词对均作为最终的研究对象。

符合此标准的词对共有27组,它们的混用数据如<表1>所示:

<表1> 同素单双音易混淆词混用统计

| 序号 | 易混淆词 | 绝对频次 |    |     | 相对频度(%)        |               |                 |
|----|------|------|----|-----|----------------|---------------|-----------------|
|    |      | 前词   | 后词 | 总计  | 前词             | 后词            | 总计              |
| 1  | 美⇌美丽 | 4    | 63 | 67  | 1.64(4/244)    | 13.82(63/456) | 9.57(67/700)    |
| 2  | 错⇌错误 | 32   | 20 | 52  | 9.14(32/350)   | 7.04(20/284)  | 8.20(52/634)    |
| 3  | 变⇌变化 | 78   | 8  | 86  | 8.26(78/944)   | 0.78(8/1025)  | 4.37(86/1969)   |
| 4  | 帮⇌帮助 | 70   | 14 | 84  | 20.71(70/338)  | 0.74(14/1892) | 3.77(84/2230)   |
| 5  | 穷⇌贫穷 | 18   | 20 | 38  | 8.18(18/220)   | 12.99(20/154) | 10.16(38/374)   |
| 6  | 生⇌出生 | 23   | 20 | 43  | 8.58(23/268)   | 3.70(20/540)  | 5.32(43/808)    |
| 7  | 建⇌建立 | 0    | 40 | 40  | 0(0/162)       | 13.51(40/296) | 8.73(40/458)    |
| 8  | 变⇌改变 | 36   | 22 | 58  | 3.81(36/944)   | 2.52(22/873)  | 3.19(58/1817)   |
| 9  | 成⇌成为 | 14   | 34 | 48  | 3.15(14/445)   | 3.55(34/957)  | 3.42(48/1402)   |
| 10 | 声⇌声音 | 12   | 32 | 44  | 7.59(12/158)   | 3.44(32/930)  | 4.04(44/1088)   |
| 11 | 国⇌国家 | 21   | 91 | 112 | 0.97(21/2159)  | 1.85(91/4931) | 1.58(112/7090)  |
| 12 | 时⇌时候 | 134  | 20 | 154 | 2.72(134/4931) | 0.25(20/8148) | 1.18(254/13079) |

3) 并列式的判定依据邵敬敏主编《HSK汉语水平考试词典》(2000)对词语结构类型的标注。

4) 人工判定和筛选误例是在义项层面进行的。如,“好—美好”是一对准易混淆词。在《现代汉语词典》(第7版, p. 519)(以下简称《现汉》)中“好”有15个义项,但二语者将“好—美好”混淆基本只发生在“好”的第一个义项(“优点多的;使人满意的”)上。因此,“好”的其他义项与混用无关,不纳入考察范围。

5) 分别按照绝对频次和相对频度从大到小给所有词对排序,将词对的两个序号相加赋值,就得到了词语混淆度序列。该计算方法兼顾了词语混用的绝对频次和相对频度,能相对客观反映词对的混淆程度。

|    |      |    |    |    |               |               |                |
|----|------|----|----|----|---------------|---------------|----------------|
| 13 | 事⇌事情 | 16 | 76 | 92 | 0.38(16/4162) | 3.19(76/2379) | 1.41(92/6541)  |
| 14 | 等⇌等待 | 10 | 13 | 23 | 2.42(10/414)  | 13.68(13/95)  | 4.52(23/509)   |
| 15 | 市⇌城市 | 9  | 36 | 45 | 0.95(9/950)   | 2.60(36/1384) | 1.93(45/2334)  |
| 16 | 树⇌树木 | 0  | 21 | 21 | 0(0/377)      | 35.59(21/59)  | 4.82(21/436)   |
| 17 | 过⇌度过 | 27 | 13 | 40 | 1.64(27/1645) | 10.92(13/119) | 2.27(40/1764)  |
| 18 | 活⇌生活 | 19 | 29 | 48 | 2.84(19/668)  | 0.69(29/4195) | 0.99(48/4863)  |
| 19 | 路⇌道路 | 8  | 13 | 21 | 0.88(8/906)   | 6.05(13/215)  | 1.87(21/1121)  |
| 20 | 钱⇌金钱 | 0  | 29 | 29 | 0(0/1725)     | 23.39(29/124) | 1.57(29/1849)  |
| 21 | 皮⇌皮肤 | 2  | 2  | 4  | 2.02(2/99)    | 4.88(2/41)    | 2.86(4/140)    |
| 22 | 美⇌美好 | 11 | 2  | 13 | 4.51(11/244)  | 0.69(2/289)   | 2.44(13/533)   |
| 23 | 有⇌具有 | 0  | 41 | 41 | 0(0/24373)    | 11.75(41/349) | 0.17(41/24722) |
| 24 | 贫⇌贫穷 | 4  | 0  | 4  | 6.15(4/65)    | 0(0/154)      | 1.83(4/219)    |
| 25 | 改⇌改变 | 15 | 2  | 17 | 5.15(15/291)  | 0.23(2/873)   | 1.46(17/1164)  |
| 26 | 好⇌美好 | 5  | 23 | 28 | 0.06(5/8347)  | 7.96(23/289)  | 0.32(28/8636)  |
| 27 | 难⇌困难 | 10 | 9  | 19 | 0.81(10/1242) | 0.65(9/1385)  | 0.72(19/2627)  |

在27组易混淆词中，混用绝对频次在40以上的共有14组，占比51.85%；其中超过100的有两组，分别是“时⇌时候”154次，“国⇌国家”112次。混用绝对频次在10以下的只有2组，占比7.41%。在27组易混淆词中，混用相对频度在4%以上的共有9组，占比33.33%。1%以下的共有4组，占比14.81%。大多数词语的混用相对频度均不低。数据详见<表2>。

<表2> 单双音易混淆词混用绝对频次和相对频度分布

|           |            |             |             |             |        |     |
|-----------|------------|-------------|-------------|-------------|--------|-----|
| 绝对频次N     | 0 < N ≤ 10 | 10 < N ≤ 20 | 20 < N ≤ 30 | 30 < N ≤ 40 | N > 40 | 总计  |
| 词对数量      | 2          | 3           | 5           | 3           | 14     | 27  |
| 比重 (%)    | 7.41       | 11.11       | 18.52       | 11.11       | 51.85  | 100 |
| 相对频度F (%) | 0 < F ≤ 1  | 1 < F ≤ 2   | 2 < F ≤ 3   | 3 < F ≤ 4   | N > 4  | 总计  |
| 词对数量      | 4          | 8           | 3           | 3           | 9      | 27  |
| 比重 (%)    | 14.81      | 29.63       | 11.11       | 11.11       | 33.33  | 100 |

3. 同素单双音易混淆词的确定及其整体混用情况

同素单双音词语的混淆特征可以从学习者母语背景分布、混淆词语的词类分布、误用方向、误用词与当用词的语义关系几个方面进行观察。

## 1) 学习者母语背景分布表现出强共通性

根据学习者的母语背景分布,易混淆词可以分为共通性易混淆词和特异性易混淆词。共通性易混淆词指不同母语背景学习者都会混用的词语;比如“原因-理由”是英、日、韩、蒙、印尼等不同母语背景的学习者都会混淆的词语,被称为共通性易混淆词语;特异性易混淆词指单一母语背景学习者混用的词语,或者虽然被多种母语背景学习者混用,但某一母语背景学习者的混用表现具有区别性特征的词语。(申旻京,2011)比如,“恩惠-恩情”只有韩国学习者会出现混淆,因此是特异性易混淆词;再如“爱-爱情”是多种母语背景学习者都会混用的词语,混淆表现出双向性,但是只有韩国学习者混淆表现出单向性,即,当用“爱”时,误用成了“爱情”。因此,这组词语也是韩国学习者特异性易混淆词语。

学习者的汉语易混淆词语可以表示为:

|   |                  |   |
|---|------------------|---|
| { | 单一母语背景学习者特异性易混淆词 | { |
|   | 多母语背景学习者共通性易混淆词  |   |
|   | 混淆表现相异的易混淆词      |   |

混淆表现相同的易混淆词

从理论上推测,不同学习群体的共通性词语混淆主要与目的语特征有关,而特定学习群体的特异性词语混淆则可能来自母语词汇知识的干扰(张博,2016, p. 3)。

本文对日、韩、英三种母语背景学习者语料进行横向对比后发现,在27组易混淆词中,三个学习群体全都混用的词语共有25组,占比92.59%。只被日、韩学习者混用的词语1组,为“改 $\rightleftharpoons$ 改变”。单一母语背景学习者混用的词语只有“贫 $\rightarrow$ 贫穷”,它只出现在日语背景学习者语料中。

这表明,同素单双音词语混用在不同学习群体中具有超高的广泛性,也就是说,单双音词语混淆情况非常严重。

据此推断,该类词语混淆应与汉语单双音词语的独特属性具有强相关性。

## 2) 同素单双音易混淆词基本上全都是同词类词语

按照实际发生混用时误用词和当用词的词性,易混淆词可以分为同词类混淆词语和跨词类混淆词语。在27组同素单双音易混淆词中,25组属于同词类混淆词语,2组属于跨词类混淆词语。如:

(1) 不仅如此,【变】对个人发展起着举足轻重的作用。[韩]

(2) 跟成家的人相比,未成家的学生们做完作业之后就没有别的【难】了。[日]

误例(1)中的误用词“变”只有动词词性，而这里的当用词“变化”则是名词，该误例属于“动一名”跨类混用。误例(2)中，学习者想要表达的应当是“生活中遇到的不易解决的问题或障碍”，当用名词“困难”。此误例中的“难—困难”属于“形一名”跨类混用。

在25组同词类混淆词语中，动词数量最多，共计10组，它们是：帮—帮助、变—改变、成—成为、等—等待、改—改变、过—度过、活—生活、建—建立、生—出生、有一—具有。名词共有9组，分别是：国—国家、路—道路、皮—皮肤、钱—金钱、声—声音、时—时候、市—城市、事—事情、树—树木。形容词共有5组：好—美好、美—美好、美—美丽、贫—贫穷、穷—贫穷、形名兼类词1组：错—错误。

#### A. 同素单双音动词多发生语体搭配混用

(3) 政府平等的处理，支持和鼓励，【帮】他们成为国家的一分子。[英]

(4) 我可以【帮助】你带你的孩子。[日]

以上语例中的混用词对在语义和句法上似乎不存在差异，但词语所在的句子或其搭配词却具有语体差异。如误例(3)所述内容似乎属于新闻性质，其中的“政府”“平等”“鼓励”“关注”等均是非常正式的词语，而误例(4)中“带孩子”明显是口语的非正式表达。在成对的同素单双音动词中，单音词常常具有口语色彩，而双音词则相对较为正式，以上误例均属于语体搭配混用。

#### B. 同素单双音名词多发生韵律搭配混用和语体搭配混用

(5) 坐火车的【时】，我跟中国人一边谈话一边玩纸牌了。[英]

(6) 街上的场面很热闹，各个【市】都有游行。[英]

(7) 好的是帮弱小【国】从敌人的侵略中保护人民。[韩]

(8) 我看中国无疑能走经济发展的【路】，能走社会主义市场经济的【路】。[英]

(9) 我也知道病不治好，加上家里人的【金钱】不够，精神上的负担更多。[日]

以上误例(5)–(7)属于韵律搭配混用。王洪君(2001)将定语为指代或数量成分、或带“的”的定中结构归为“自由短语”，韵律黏着的单音词往往不能充当其中的中心语。误例(5)中的“时”和(6)中的“市”均属于此类错误。另外，[2+1]式不是形名的常规模式，(7)中的“国”受到双音形容词“弱小”修饰，不符合韵律搭配规则。

误例(8)(9)属于语体搭配混用。单音名词常常具有口语色彩，对应的双音名词则具有书面语色彩，以上语例中所用词语与整个句子在语体上不一致。

### C. 同素单双音形容词多发生语义混用

- (10) 其中荷花更【美好】。[韩]
- (11) 我喜欢运动, 可是没留下【美好】的成绩。[日]
- (12) 音乐是一种【美丽】的艺术。[日]
- (13) 这篇小说的内容太【美丽】, 结果是大家都幸福。[日]
- (14) 在美国, 学生没有那么【困难】的考试。[英]

“美好”多用于生活、钱途、愿望等抽象事物, 它们较少与具体事物搭配, 而“美”和“好”则没有这方面的限制。误例(10)(11)中的“荷花”和“成绩”均是具体可感的, 不能受“美好”形容。

“美丽”形容的一般是从视觉上让人感到愉快的人或事物, 而“美”能够形容的既可以是视觉, 也可以是听觉、味觉、感知等方面让人感到愉快的人或事物。(12)“音乐”是听觉方面, (13)“小说内容”指感知方面, 因而形容词当用“美”。

“困难”侧重于强调做成某件事的阻碍很多或者条件不足, “难”侧重于形容客观对象本身难度较大, 如考试、题目、语言等。(14)在形容“考试”时, 当用“难”。

以上是名动形三类同素单双音易混淆词的主要混用倾向, 实际上少数同素单双音名词和动词也存在语义混用, 如:

- (15) 其次, 吸烟者很容易感到【皮】老。[韩]
- (16) 【树木】下面有一口井。[韩]
- (17) 古时候生活条件没有现代好, 那么何必在山上【建立】个庙呢? [日]
- (18) 鲁迅倘若还在【生活】着则觉得中国已经改革了传统社会的虚伪性和落后性, 或者中国人还在互相吃。[英]

“皮”和“皮肤”都可用于指人身体表面的组织, 但“皮”侧重于与“肉”相区别, “皮肤”则侧重于呈现其色泽、状态(误例15)。“树木”指树的总称, 属于集合概念, 而(16)中则是指称具体的一棵树, 因而当用“树”。“建立”一般指“友谊”“制度”“机构”“政权”等开始形成或成立, (17)中对象是具体建筑物时, 当用“建”。“活”侧重于表示有生命, 与“死”相对, “生活”侧重于表示人进行各种活动。(18)想要表达的是“如果鲁迅还有生命”, 因而当用“活”。

个别同素单双音形容词也存在语体搭配混用, 如:

- (19) 利用外表去评价一个人是绝对【错】的。[英]

(20) 如果我们选择绿色食品的话,很多【穷】地区的孩子就濒临死亡。[韩]

(21) 他家非常【贫穷】,他小时死了父亲,在母亲的教育下长大了。[日]

(19) 整个句子都具有正式语体色彩,而形容词“错”常常用在非正式语体中,这里当用双音词“错误”。(20)中“地区”“濒临死亡”都具有书面语色彩,因而也应选择正式语体的双音词“贫穷”与之搭配。(21)中的“他家”及整个句子都具有口语色彩,因而当选用非正式语体的单音词“穷”。

### 3) 同素单双音易混淆词双向误用占绝对优势

词语混淆按照误用方向可以分为单向误用和双向误用。单向误用指在成对的易混淆词中,CSL学习者只混用了其中一词,另一词却没有混用误例。如“建←建立”在语料库中只出现了当用“建”而误用“建立”的语例,而没有相反的情况。双向误用则指成对易混淆词中的二词均存在混用误例,如“成⇌成为”。

同素单双音易混淆词以双向误用占绝对优势。在27组混淆词语中,双向误用共22组,占比81.48%;单向误用5组,占比18.52%。双向误用直接反映了CSL学习者对易混淆词中的两个词语均掌握不好,因而无论使用哪个词语都会出现混用。

同素单双音易混淆词具有如此高的双向误用比重,说明CSL学习者对该类词语混淆严重。

### 4) 同素单双音易混淆词的语义关系均较近

同素单双音易混淆词主要根据实际发生混用的义项判断其语义关系。我们依据词语在『现汉』中的释义,即词语理性意义的相同和相近,将同素单双音易混淆词分为同义关系词语和近义关系词语。

同素单双音易混淆词在『现汉』中的释义包括以下五种类型:

a. 以双音词释单音词。如“事:①名事情。”这类易混淆词共有10组,它们是:帮—帮助、变—变化、等—等待、改—改变、国—国家、路—道路、生—出生、声—声音、市—城市、事—事情。

b. 以单音词释双音词。如“金钱:名货币;钱。”这类易混淆词共有2组,它们是:过—度过、钱—金钱。

c. 单双音词的释语相同。如“错:⑤形不正确。错误:①形不正确;与客观实际不符合。”这类易混淆词共有4组,它们是:错—错误、贫—贫穷、穷—贫穷、时—时候。



d. 单双音词释语的中心成分一致,但其中一词的释语还具有其他限定成分。如“美:③形令人满意的;好。美好:形好(多用于生活、前途、愿望等抽象事物)。”这类易混淆词共有8组,它们是:变—改变、成—成为、好—美好、活—生活、美—美好、皮—皮肤、树—树木、有—具有。

e. 实际发生混用的单双音词的义项不同,释语不同,但所表示的语义仍相近。如“难:①形做起来费事的;不容易。困难:①形事情复杂,阻碍多。”这类易混淆词共有3组,它们是:建—建立、美—美丽、难—困难。

通过对『现汉』释义的查检和比对,以上五种类型的27组同素单双音易混淆词的语义关系都较近。其中,前三类的易混淆词属于理性义相同词语,后两类的易混淆词属于理性义相近词语。具体分布如下:

<表3> 同素单双音易混淆词的语义关系分布

| 语义关系 | 数量 | 易混淆词   |
|------|----|--|
| 同义   | 16 | 帮—帮助、变—变化、错—错误、等—等待、改—改变、国—国家、过—度过、路—道路、穷—贫穷、生—出生、声—声音、时—时候、市—城市、事—事情、贫—贫穷、钱—金钱。 |
| 近义   | 11 | 变—改变、成—成为、好—美好、活—生活、建—建立、美—美好、美—美丽、难—困难、皮—皮肤、树—树木、有—具有。                          |

## 4. 同素单双音词语混淆的生成机制

### 1) 汉语复合词表征结构

心理语言学认为,人类习得的语言信息被存贮在记忆里,这种记忆被称为表征(representation)。(李荣宝等,2000)Levelt(1989, pp. 181-232)提出,母语者的词语表征应该包括语义、句法、语音/拼写和形态信息,前两者合称为标义词位,后两者合称为形式词位。Levelt(1989, pp. 181-232)基于印欧语的研究提出了词语表征结构,然而其形态信息中的内部屈折、异根等并不适用于汉语词。因此,我们应结合汉语词的特点,构建复合词的表征结构。

#### (一) 汉语复合词表征内应包括韵律信息

韵律信息指某词与他词的韵律搭配规则,它应属于汉语词的基本属性。在现代汉语词汇中,双音节词占比高达七成左右。据『现代汉语频率词典』(1986)统计,使用度排名前

5000的词语在各种文体语料中的覆盖面达到91.67%。在这5000词中,单音节词1795个,占比35.9%,双音节词3103个,占比62.06%,二者合计占比97.96%。现代汉语词汇中的常用词以双音节为主,单音节的比重也很大,这就带来了单音词和双音词的区分,以及由此造成的“单双组配”的区分,主要是[2+1]和[1+2]的韵律组配区分。这种区分不是单纯语音上的区分,它同时也是语法、语义、语用上的区分,它是一种形态手段<sup>6)</sup>。冯胜利(2007)也曾提出,汉语可以通过韵律这种超音段的手段,发挥其类似于音段形态手段的作用。汉语母语者的词汇能力中应该存在韵律信息,单音词使用受限的知识及单双音节的组配规则是下意识自动提取的,母语者不会产出“\*市的措施”、“\*十八岁的时”、“\*互相帮”、“\*绝对错”等韵律搭配错误。而不同母语背景的二语学习者均会产出表现一致的韵律搭配混用,如名词单独作定语或单独作含“的”定中结构的中心语,单音动词或形容词受双音副词修饰等。而且,学习者一般不会对这些错误进行自我修正。

因此,对于汉语来说,韵律信息应属于词汇能力的一部分,被存储在词语表征中。母语者能够下意识地提取,二语者在韵律信息缺失或不足的情况下,便会产出韵律搭配错误。

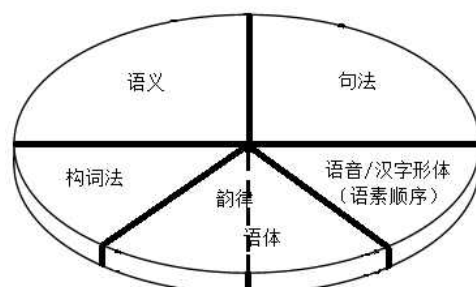
## (二) 汉语复合词表征内应包括语体信息

语体信息也应属于汉语词的基本属性。汉语词的语体差异同样最常表现在单双音节词语中。张国宪(1989)、李临定(1990)、董秀芳(2002)都曾指出,单音节动词常用于口语,双音节动词多具书面语色彩。冯胜利(2010)则提出现代汉语存在“书面体-口语体”和“文雅体-白话体”两对语体范畴,并认为相互对立的语体各具自身的词汇。例如,“金钱”具有书面语语体色彩,“钱”具有口语语体色彩,这种语体信息应是自然储存在母语者心理词库中的,在激活和使用“金钱”和“钱”时,语体信息会自动提取。然而,日语、韩语、英语背景学习者均存在“钱—金钱”的语体搭配混用。说明二语者相关词语的语体信息仍未正确整合入词语表征内。

综上所述,汉语复合词的表征应包括形式、语义、句法、形态、韵律和语体信息。除韵律和语体之外,汉语复合词的形式信息和形态信息也与英语等印欧语不同。印欧语系诸语言的书写系统属于拼音文字,因而其形式信息包括语音和拼写两部分,其中的拼写指字母的书写规则。而汉字是一种意音文字,最小的音义结合体是语素。汉语复合词的形式信息应当包括语音、汉字形体和构词语素顺序。另外,不同于英语词的内部屈折等形态信息,汉语复合词的形态信息应当指构词法信息。综合以上分析和考虑,我们对Levelt(1989, 181-232)基于印欧语提出的词语表征结构进行了补充修订,构建了汉语复合词表征结构,如<图1>所示:

6) 沈家煊2016年3月18日在上海外国语大学报告语。

&lt;图1&gt; 汉语复合词表征结构



与Levelt (1989, pp. 181-232) 的相比, 汉语复合词表征结构具有以下几点不同:

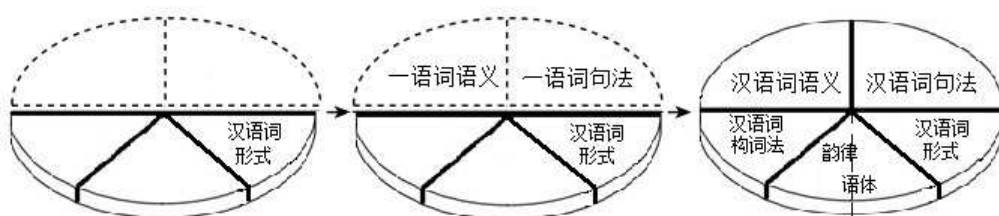
- 1) 形态信息不包括内部屈折、异根等, 而是表现为构词法信息;
- 2) 形式信息表现为语音、汉字形体和语素顺序;
- 3) 汉语复合词表征结构内包括韵律信息和语体信息, 二者存在部分交叉。冯胜利 (2010) 曾指出, 双音构词法是语体正式化的手段之一, “‘各’是典雅体嵌偶单音词”, “各”有[各+N]必双的使用要求。韵律和语体有时密切相关, 因此, 表征结构中的这两部分信息之间用虚线隔开。

## 2) 汉语二语词汇发展模型

Jiang (2000) 将二语词汇习得的抽象过程具体化为词汇能力四类信息整合进词语表征结构里的过程, 并根据词语表征的不同形式, 将二语词汇习得过程划分为三个阶段: 形式阶段、一语标义词位调节阶段、二语词整合阶段。在词汇发展的不同阶段, 学习者心理词库中具有不同的词语表征信息, 各阶段的词语输入和输出机制也有所区别。

本文推测, 汉语作为第二语言的词汇发展也大致符合Jiang (2000) 的三个阶段: 1) 形式阶段; 2) 一语标义词位调节阶段; 3) 汉语词条整合阶段。汉语二语复合词发展三阶段图示如下:

&lt;图2&gt; 汉语作为第二语言的复合词发展模型



### 3) 同素并列式单双音词语混淆的生成机制

从构词语素之间的语义关系看,本文的同素单双音易混淆词中的双音复合词基本上均为同义并列。它们大多是由同素单音词经过双音化而来,因而成对词语在语义上相同或高度相近。在本文探讨的27组同素单双音易混淆词中,同义关系词语16组,近义关系词语11组。同义关系词语主要发生了韵律混用和语体混用,近义关系词语则主要发生了语义混用。

具有源流关系且语义相同相近的同素单双音词语是汉语特有的词汇类聚,它们在日语、韩语、英语中大多对应于同一个词。CSL学习者的同素单双音词语使用初期会受到一语标义词位的调节作用,从而产生语义误推,造成词语的语义混用。随着词汇习得的发展,学习者从输入词语中提取并整合到词语表征内的汉语词标义词位信息越来越多,他们也将尝试利用汉语词信息直接产出词语。此时,整合不完善或不准确的语义或句法信息将导致词语混用。

理性义基本相同的同素单双音词语常常发生韵律混用和语体混用。汉语二语学习者很难在词汇习得中关注并提取到词语表征中存在的韵律信息和语体信息。我们推测,其原因有三点:

1) 从语言对比的角度看,CSL学习者母语中不存在同素同义单双音词语类聚,成对单双音词语在其母语中往往对应于同一个词。如“错”和“错误”的日语对应词都是“誤り”,韩语对应词为“잘못”,英语则均对应于wrong或mistake。单双音词语在韵律信息和语体信息上存在差异,如“错”常同单音词搭配,它适用于口语和俗白文字,具有口语色彩;而“错误”多同双音词搭配,它通用于书面语和口语(刘叔新主编,1987/2011:94)。这种差异在学习者母语中是不存在的,因而他们基本上很少能自觉关注到单双音词语在韵律和语体上的差异,他们也缺乏从输入词语中提取单双音词韵律信息和语体信息的相关经验。

2) 从词汇习得的主体看,二语者在词汇习得过程中,更多关注的是词语的意义和用法。正确理解和传达语义是语言交际的基本需求,因而学习者将大部分注意力都集中于此。而且,即使单双音词语使用发生了韵律搭配错误或语体搭配错误,有时也不会影响交际的正常进行。CSL学习者长期忽视同素单双音词语在韵律信息和语体信息上的差异,因此也就无法有效提取相关知识信息。

3) 从词汇习得的客体看,单双音词语的韵律信息和语体信息规律性较弱,特异性较强,这在客观上给CSL学习者关注和提取它们带来了难度。刘叔新、周荐(1992, pp. 81-84)认为单双音节的韵律搭配限制为“单对单、双对双”。但实际情况却较为复杂,如

“大国”“小城”可以说，而同为“单对单”的“\*大城”“\*大市”“\*小市”却不能说；“双对双”的“\*日本国家”“\*北京城市”“\*劳动节”则都不成立（以上错误均在学习者语料中大量存在）。词语的语体信息也缺乏一致的规则性。虽然不少研究均认为，单音动词常用于口语，双音动词常用于书面语，但程娟、许晓华（2004）对《等级大纲》中181组同素单双音动词的语体色彩统计结果显示：19.3%的单音词比双音词的书面语色彩更浓，另有33.7%的单双音词语语体色彩差异不明显。可见，在同素单双音词语类聚中，语体色彩的倾向性并不强。CSL学习者多为认知能力发达的成年人，他们在二语词汇习得过程中，会对已经获得的知识进行总结归纳，并将其类推到同类词语使用上。而面对同素同义单双音词语纷繁复杂的韵律制约规则和语体色彩分布，学习者即使有时注意到并提取了某个词语惯常的韵律搭配和语体搭配，也很难将其类推到其他词语上。也就是说，即使二语者正确提取了某对单双音词语的韵律信息和语体信息，若对其进行了不当类推，则会导致其他单双音词语的相关信息整合错误，从而造成词语混用。

4) 从二语词汇教学来看，印欧语词汇的形态信息是显性知识（Bialystok, 1982, p. 181-206），词语的形态知识是二语课堂教学中必然教授的内容，教材中也有相关知识的讲解。二语者可以通过有意识地学习，获得词语形态信息，并且能够用语言将其表达出来。如apple的复数形式要加-s，go的过去式是went等。而汉语词的韵律信息和语体信息则不然，对外汉语教师较少将单双音词语的韵律搭配规则和语体搭配规则作为授课内容，教材中也大多不存在相关内容的说明和针对性练习。因而，CSL学习者可能不知道“\*该国”不成立，以及为什么不成立；也不清楚为何“好朋友”“大城市”能说，但“\*穷地区”却不能说。这其中涉及“国”和“国家”、“穷”和“贫穷”在语体信息上的差别。

综上所述，汉语二语学习者较难关注到单双音词语的韵律信息和语体信息，因而也就难以从词语输入中有效提取这些信息，这导致其心理词库中该部分语言能力的缺失，因而在词语输出时，单双音节的使用更可能是一种随机的选择，又或者存在学习者的个体倾向性<sup>7)</sup>。即使学习者通过词语输入提取到某对词语的韵律信息和语体信息，也很难将其类推到其他词语的使用上。这都将导致同素同义单双音词语的混淆现象。

7) 比如在“国”和“国家”的使用上，某一学习者可能都选择“国”，而另一学习者可能都选择“国家”。

## 5. 结语

本文基于大规模汉语中介语语料库,综合考量词语混用的绝对频次和相对频度,筛选出混淆度排名靠前的27组同素单双音易混淆词,通过数据对比分析发现,同素单双音词语的混淆分布特征为:共通性强,同素单双音易混淆词基本上全为同词类混淆词语,其中动词多发生语体搭配误用,名词多发生韵律搭配误用和语体搭配误用,形容词多发生语义混用;双向误用的易混淆词占据绝对优势;同素单双音易混淆词的语义关系均较近。我们认为,一语标义词位的调节作用,以及汉语二语者心理词库中缺少词语的韵律信息和语体信息,是造成同素单双音词语混淆的主要原因。比较而言,词语的韵律混淆和语体混淆最难克服。

本研究可以为并列式复合词教学提供一些有针对性的教学建议:

1) 并列式同素词语混淆发生在词汇发展的一语标义词位调节阶段。学习者利用一语标义词位调节二语词产出,这是词汇习得的必经过程,因而这种调节作用是不可避免的。但我们可以采用一些针对性的方法,帮助学习者加速词汇发展,向词语整合阶段靠近。具体而言,一方面,本文发现了与并列式相关的高频混用词语以及导致混用的具体因素,这使得我们在词语教学中能够有的放矢,在学习者已掌握成对相关词语之后,针对二词的易混淆点进行辨析,以帮助学习者明晰其语义差异;另一方面,对于整体词汇能力较高的学习者,我们应向其说明,一语词与二语词语义常常是不对等的,并建议他们在学习汉语词时多从整句中提取词义,而不是仅仅记住其对译词,应多查汉语词典,而不是双语词典。

2) 二语教师应该对同素单双音词语在语体色彩和韵律制约规则上的差异具有清楚的认识,对可能致误的固化表达也要有所准备,并针对不同水平的学习者,采用不同的教学对策。对于初级学习者,教师在讲授词语时无须对比相应的单(双)音词语,也不必讲解其语体色彩和韵律制约规则的差异,但如果学生混用了单双音词语,则务必予以纠正。同时,针对汉语中介语语料库中出现的同素单双音词语的高频混用(如“\*北京城市、\*上课时候”等NP双+N双类),教师应对正确形式有所强调和拓展练习,以预防偏误的发生。对于具备了一定语感和语言知识的中高级学习者,教师可以对课文中涉及到的易混淆单双音词语进行辨析,提供单双音词语的常用搭配项并让学习者练习使用,并从易到难地分阶段地讲解单双音词语在语体色彩和韵律制约规则上的差异。

## 【参考文献】

### 〈单行本〉

- 北京语言学院语言教学研究所编,『现代汉语频率词典』,北京语言学院出版社,1986。
- 董秀芳,『词汇化:汉语双音词的衍生和发展』,四川民族出版社,2002。
- 刘叔新·周 荐,『同义词语和反义词语』,商务印书馆,1992。
- 刘叔新,『现代汉语同义词词典(第四版)』,中国社会出版社,1987/2011。
- 张 博,『不同母语背景的汉语学习者词语混淆分布特征及其成因研究』,北京大学出版社,2016。
- 中国社会科学院语言研究所词典编辑室编,『现代汉语词典(第7版)』,商务印书馆,2016。

### 〈论文〉

- 程 娟·许晓华,「HSK单双音节同义动词研究」,『世界汉语教学』,2004(4)。
- 冯胜利,「韵律语法理论与汉语研究」,『语言科学』,2007(2)。
- ,「论语体的机制及其语法属性」,『中国语文』,2010(5)。
- 季 瑾,「HSK甲级单双音同义动词部分不可替换的类型分析」,『语言教学与研究』,2005(5)。
- 李临定,「动词分类研究说略」,『中国语文』,1990(4)。
- 李荣宝·彭聘龄·李崑,「双语者第二语言表征的形成与发展」,『外国语』,2000(4)。
- 刘春梅,「留学生单双音同义名词偏误统计分析」,『语言教学与研究』,2007(3)。
- 刘智伟,「含同一语素的同义单双音节动词研究」,北京师范大学博士论文,2005。
- 骆健飞,「论单双音节动词带宾的句法差异及其语体特征」,『语言教学与研究』,2017(1)。
- 申旼京,「韩语背景学习者汉语词语混淆的母语影响因素研究」,北京语言大学博士论文,2011。
- 王洪君,「音节单双、音域展敛(重音)与语法结构类型和成分次序」,『当代语言学』,2001(4)。
- 吴 颖,「同素近义单双音节形容词的差异及认知模式」,『语言教学与研究』,2009(4)。
- 于 洋,「CSL学习者同素同义单双音名词混淆分布特征及其成因」,『语言教学与研究』,2015(6)。
- 张 博,「同义词、近义词、易混淆词:从汉语到中介语的视角转移」,『世界汉语教学』,2007(3)。
- 张 博·针对性,「易混淆词辨析词典的研编要则」,『世界汉语教学』,2013(2)。
- 张国宪,「单双音节动作动词词语用功能差异探索」,『汉语学习』,1989(6)。
- Bialystok, E, "On the relationship between knowing and using linguistic forms", *Applied Linguistics*, 1982, 3(3)。
- Jiang, N, "Lexical representation and development in a second language", *Applied Linguistics*, 2000, (1)。
- Levelt, W. J. M, *Speaking: From Intention to Articulation*, Cambridge, MA: Bradford, 1989。

【논문초록】

|  |       |  |       |             |       |             |
|--|-------|--|-------|-------------|-------|-------------|
| 키워드<br>Key Words   | 국문    | 단/쌍음절 어휘 혼용, 중국어 제2언어 학습자, 분포특징, 생성기제, 중국어 복합사 특정 구조   |       |             |       |             |
|  | 영문    | Lexical errors, Confusable words, Homomorphemic, Monosyllabic and disyllabic, Chinese second language learners |       |             |       |             |
| <div><div>The Distribution Characteristics and Generative Mechanism of Homomorphemic Monosyllabic and Disyllabic Confusable Words by CSL</div><div>Fu Dong-Dong · Park Byung-Seon</div><div><p>In modern Chinese, homomorphemic monosyllabic and disyllabic words are one of the very special groups of word clusters which are easily confused by Chinese as a second language learners. Previous studies have not comprehensively examined this confusion phenomenon, and there is a lack of relevant literature involving psychological analyses. Firstly, this article comprehensively surveys the Chinese Intermediate Language Corpus and identifies 27 groups of highly confusable words. Secondly, this article analyses the distribution characteristics of the homomorphemic monosyllabic and disyllabic words, finding strong commonality in the distribution of learners' native language backgrounds, with overwhelming bidirectional misuse and close lexical-semantic relations. It also finds that verbs are mainly misused in stylistic collocations, and nouns are frequently misused in both prosodic and stylistic collocations, while adjectives are often misused in semantics. Besides, the article analyses the generative mechanism from the psychological perspective of Chinese as the second language learners. It argues that L1-lemma mediation and the lack of prosodic and stylistic information of words in the psychological lexicon of learners are the main reasons for the confusion. This article argues that the semantic confusion of words is difficult to overcome, while prosodic and stylistic confusion are even more difficult to overcome. Last but not the least, the article offers targeted pedagogical suggestions.</p></div></div> |       |  |       |             |       |             |
| 저 자  | 제1저자  | 부동동 / 付冬冬 / Fu Dong-Dong   |       |             |       |             |
|  | 교신저자  | 박병선 / 朴炳仙 / Park Byung-Seon  |       |             |       |             |
| 논문작성일  | 투 고 일 | 2024.01.31.  | 심 사 일 | 2024.02.26. | 게재확정일 | 2024.03.11. |