

说者身份信息对口语词汇加工的影响*

殷舒琦¹ 穆凯代斯·艾尔肯¹ 沈泰宇¹ 李利² 余可可^{**1} 王瑞明^{**1}

(¹“儿童青少年阅读与发展”教育部哲学社会科学实验室, 华南师范大学心理学院/心理应用研究中心, 广州, 510631)

(²华南师范大学国际文化学院, 汉语学习与国际推广省重点实验室, 广州, 510631)

摘要 研究采用长时重复启动范式, 操纵学习和测验材料中说者身份的一致性, 考察说者身份信息对较浅的词汇通达(实验1)和较深的概念理解(实验2)的影响。实验1词汇决定任务发现, 对于学过的词, 说者身份一致条件的正确率显著高于不一致条件, 表明说者身份信息和语言信息以整合的方式影响词汇通达。实验2类别决定任务发现, 对于未学过的词, 说者身份一致条件的正确率显著高于不一致条件, 表明说者身份信息以独立的方式影响概念理解。基于上述结果和先前的相关理论, 研究尝试提出了包含身份信息加工的口语词汇加工的新观点, 有助于更具社会性和生态性地解释口语词汇的认知加工。

关键词 口语词汇加工 身份信息 语言信息 词汇通达 概念理解

1 前言

口语交流是人际沟通的重要形式。为保证口语交流的顺利进行, 听者需要准确地把握说者的说话内容(也称为语言信息)。在这一过程中, 听者往往面临着诸多由说者变异性带来的挑战, 例如说者身份的变化(从当前的说者变为另一个说者)和不同的情绪状态(说话语调悲伤、兴奋或平静)。这些与说者身份和情绪状态相关的信息被统称为非语言信息(Belin et al., 2004; Scott, 2019)。虽然传统的口语研究已对语言信息的加工问题做了许多探讨, 但这些研究大多忽视了非语言信息对语言信息加工的可能影响, 导致对语言信息加工的探讨缺乏社会性和生态性(赵荣等, 2016; Hickok & Poeppel, 2015)。例如, 不同性别和不同说话风格的说者在说同一个词或同一句话时所表达的语言信息可能是不同的。近年来, 研究者们逐渐意识到非语言信息在口语交流中的重要性(胡砚冰, 蒋晓鸣, 2023; 明莉莉, 胡学平, 2021)。听者如何快速利用非语言信息准确理解说者的说话内容, 引起了研究者的广泛关注(余可可等, 2021; Luthra, 2024)。其中, 说者身份信息是否以及如何影响口语词汇加工是一个关键问题(Kapnoula & Samuel, 2019; McLennan & Luce, 2005)。对这一问题的探讨有助于揭示更具社会性和生态性的口语词汇加工机制, 为口语交流提

供参考。

然而, 当前有关这一问题的研究存在争议。抽象观和情节观分别提出了不同的观点来解释说者身份信息是否以及如何影响口语词汇加工。其中, 抽象观的观点可分为早期的抽象观和发展的抽象观。早期的抽象观认为, 说者身份信息等非语言信息是一个噪声源, 为了减少干扰, 听者需要过滤这些信息, 只留下最基本的音素、词汇等抽象单元储存在心理词典中, 说者身份信息对口语词汇加工不会产生影响(Bowers, 2000; McQueen et al., 2006; Samuel, 2011)。例如, Orfanidou等人(2011)采用重复启动范式与词汇决定任务, 发现说者身份变化(性别变化)不会影响真假词汇的重复启动效应。Rodd等人(2013)采用词汇联想任务探讨说者身份变化(性别变化)对歧义词加工的影响。实验包含启动和测试两个阶段。启动阶段先给被试呈现包含歧义词的句子(男女说者各半), 接着呈现一个探测词, 被试需要判断探测词与句子语义是否相关, 以确保被试对歧义词进行了语义加工。随后, 被试需要在测试阶段完成词汇联想任务, 即听到歧义词后(此时均是女性说者), 尽快说出与之语义相关的词汇。研究者通过分析被试说出的词汇更偏向于歧义词的哪一个意思, 将其与歧义词在启动阶段的语义进行比较, 得到一致/不一致的结果, 再计算出一致占

* 本研究得到国家社会科学基金项目(23CYY049)的资助。

** 通讯作者: 余可可, E-mail: kkyu@m.scnu.edu.cn; 王瑞明, E-mail: wangrm@scnu.edu.cn

DOI:10.16719/j.cnki.1671-6981.20260306

总体的比例,即一致性比例。结果发现,启动和测试阶段说者相同和不同条件下的一致性比例不存在显著差异,表明词汇的语义启动效应不受说者身份的影响。

一些研究者认为,早期的抽象观忽视了说者身份信息对口语词汇加工的可能影响,因此提出了发展的抽象观,认为听者在加工口语词汇时,除了形成与语言信息相关的词汇表征和概念语义表征之外,还可以形成说者身份信息等非语言信息的表征,这些非语言信息通过独立的路径影响口语词汇加工(Belin et al., 2004; Blank et al., 2014)。这一观点也得到了研究证据的支持。例如, Cai 等人(2017)要求被试判断不同口音歧义词的意义,发现如果词汇是英式英语口语,被试倾向于判断为英式英语的词汇意义;如果是美式英语口语,被试则倾向于判断为美式英语的词汇意义。这一结果说明被试会根据口音来判断词汇意义,反映了说者身份对口语词汇理解的影响。此外,研究者还设置了处于英式和美式口音之间的中性口音条件,发现在英式英语口语区块(词汇为英式英语口语和中性口音),被试倾向于将中性口音的歧义词判断为英式英语的意义。而在美式英语口语区块(词汇为美式英语口语和中性口音),被试倾向于将中性口音的歧义词判断为美式英语的意义。这进一步表明,听者能够基于口音建立说者的身份表征,当中性口音的歧义词激活多个意义时,听者会根据所建立的身份表征从中进行选择。研究者由此认为,说者的身份表征会通过独立的路径影响听者对词汇意义的选择倾向。Davies 等人(2022)先让被试在学习阶段建立说者和特定说话风格的联系(偏向用隐喻意义/偏向用字面意义)。接着在测验阶段,让被试听不同说话风格的说者所说的多义词(同时含有隐喻意义和字面意义),同时呈现四张图片给被试(一张图片与隐喻意义有关,一张图片与字面意义有关,另外两张图片为干扰项),要求被试选择与这个词的意义有关的图片。结果发现,听者在建立了说者与特定说话风格的联系后,即使这些多义词并未在学习阶段加工过,听者在测验阶段都会受到说话风格的影响,倾向于将多义词理解成与说者特定说话风格相关的意义。例如,当被试听到偏向用隐喻意义的说者说的多义词时,倾向于选择与隐喻意义有关的图片。这一结果也说明,听者可以建立起说者身份的独立表征,这种独立的身份表征会影响听者对未听

过的多义词的理解倾向。

另有一些研究却认为,说者身份信息和语言信息以整合的形式影响口语词汇的加工。研究者也把这种观点称为情节观(Cooper & Bradlow, 2017; Goldinger, 1996; McLennan & Luce, 2005)。Kapnoula 和 Samuel(2019)首先在学习阶段要求被试学习新异词和与之对应的图片,其中一半新异词总是与特定的说者关联(相关条件),另一半新异词则没有这种关联(不相关条件),然后在测验阶段要求被试判断新异词对应的图片。结果发现,相比于不相关条件,被试在相关条件下总是能够更快地判断出与新异词对应的图片。这一结果说明,被试在学习阶段建立起了说者身份和词汇语义的整合表征,这种整合表征促进了被试在测验阶段对词汇语义的理解。在 Clapp 等人(2023)的研究中,被试需要连续听一系列口语词汇,并判断每个词是否在之前出现过。结果发现,被试判断相同说者说出的词汇显著快于不同说者。这表明,被试在连续听词汇时形成了说者身份和词汇的整合表征,词汇的说者身份发生变化会造成与这种整合表征之间的冲突,影响对词汇的判断。这两个研究的发现都支持了情节观的观点。

基于以上的回顾与分析,尽管早期和发展的抽象观都认为说者身份信息和语言信息是独立表征的,但是早期的抽象观认为说者身份信息不会影响口语词汇加工,而发展的抽象观则认为说者身份信息以独立的方式影响口语词汇加工。与之不同,情节观认为身份信息和语言信息的表征是整合的,两种信息以整合的方式影响口语词汇加工。虽然目前主要的研究证据都支持说者身份信息会影响口语词汇加工(Cai et al., 2017; Clapp et al., 2023; Davies et al., 2022; Kapnoula & Samuel, 2019),但是在影响机制这一问题上,发展的抽象观和情节观的解释存在争议。

对比分析两个观点的研究证据,我们认为存在以下两方面的问题。首先是说者身份信息的操纵。身份信息包含物理(例如年龄和性别),特质(例如可信度和吸引力)和社会(例如教育水平和专业性)等多个维度(Lavan, 2023; Lavan et al., 2024)。不同的研究实际上操纵了不同维度的身份信息。在支持发展的抽象观的相关研究中, Cai 等人(2017)操纵的是说者的口音,口音反映的是说者的地域背景,属于身份信息的社会维度; Davies

等人（2022）操纵的是说者的说话风格，说话风格反映说者的思维方式，属于身份信息的特质维度。在支持情节观的相关研究中，Kapnoula 和 Samuel（2019）操纵的是物理维度中的性别变化；Clapp 等人（2023）使用了两位男性说者的录音，同样操纵了身份信息的物理维度。操纵不同维度的身份信息可能导致上述不同的研究结果。最近一些研究指出，听者对说者性别等物理维度的识别是相对较快的，但对特质和社会维度的识别则相对较慢，且存在个体差异（Lavan, 2023; Lavan et al., 2024）。对于身份信息的操纵，相比其他维度而言，物理维度的操纵可能更快且更为有效。因此，有必要通过操纵身份信息的物理维度，再次探讨其对口语词汇加工的影响。

其次是不同研究所考察的口语词汇加工过程。Cai 等人（2017）和 Davies 等人（2022）考察的都是加工较深的语义理解过程。这些研究支持的发展的抽象观可能更适用于解释说者身份信息对较深层次的加工过程的影响。尽管 Kapnoula 和 Samuel（2019）的任务涉及了图片语义，但这种任务的加工相比于 Cai 等人（2017）和 Davies 等人（2022）是较浅的，因为这一任务不需要加工词汇的多种语义。在 Clapp 等人（2023）的研究中，被试需要对词汇进行再认，可能只需要通达词形等词汇表征，并不需要深入加工词汇的语义，反映的也是较浅的加工。进一步考虑 Kapnoula 和 Samuel（2019）和 Clapp 等人（2023）的研究，情节观可能适用于解释说者身份信息对较浅层次的词汇通达过程的影响。因此，发展的抽象观和情节观可能并不冲突，二者分别适合于解释不同深度的口语词汇加工（Cutler et al., 2010; Goldinger, 2007; Ma et al., 2023）。

综合上述两方面的分析，本研究旨在通过两个实验，具体考察在不同任务要求的情况下，说者身份信息对不同深度的口语词汇加工过程的影响。研究借鉴了双语认知研究中的内隐长时重复启动范式（李利等，2006; Zeelenberg, 2003），通过操纵学习和测验阶段语音材料的说者身份一致性，考察被试对测验阶段目标刺激的判断是否受到学习阶段说者的影响。在这一范式中，被试并不清楚实验目的、启动刺激和目标刺激之间的联系，可以较好地考察被试无意识、内隐的认知加工，尽可能地降低被试运用策略的可能性。两个实验均通过操纵说者性别这一身份信息的物理维度来设置说者身份一致和不

一致条件。

具体来说，实验 1 采用词汇决定任务，被试需要对口语词汇进行真假词判断，该任务对加工的要求较浅，侧重考察词汇通达这一过程。实验 2 采用类别决定任务，被试在词汇通达的基础上，还需要建立词汇的概念表征才能判断其是生物词还是非生物词，该任务对加工的要求较深，侧重考察概念理解这一过程。基于内隐长时重复启动范式的逻辑和本研究的实验设计，如果结果支持情节观，那么应该只在学过的词上发现说者身份一致的促进作用，因为听者只能对学过的词建立起说者身份和词汇/概念的整合表征。如果结果支持发展的抽象观，那么在学过和/或未学过的词上都可能发现说者身份一致的促进作用。一方面，听者可以利用学习阶段建立的独立的说者身份表征来加工学过和未学过的词。另一方面，听者在学习阶段已经通达了学过的词的词汇/概念表征，他们可能在测验阶段利用已通达的词汇/概念表征来完成实验任务，在这种情况下，可能并不会在学过的词上发现说者身份一致的促进作用。如果结果支持早期的抽象观，那么不管是对于学过还是未学过的词，说者身份一致都不会有显著的促进作用。在此基础上，如果实验 1 和实验 2 的结果不同，说明说者身份信息对口语词汇加工的不同过程会产生不同影响，早期和发展的抽象观与情节观可能适用于解释不同深度的加工过程。

2 实验 1

2.1 实验目的

实验 1 采用内隐长时重复启动范式和词汇决定任务，考察说者身份信息是否以及如何影响口语词汇加工的词汇通达过程。

2.2 研究方法

2.2.1 被试

基于 G*Power 使用手册（Faul et al., 2007），在计算样本量时可使用中等程度的效应量。以往与说者身份加工相关的研究（Ma et al., 2023）也采用了中等程度的效应量来计算样本量。因此，结合本实验设计，预计效应量 $f = .25$ ，要达到 .8 的统计效力，需要的总样本量为 34 人。实验共招募 88 名被试（23 名男性，年龄范围 18~25 岁，均值为 20.93 岁，标准差为 2.02 岁），随机分为两组。被试均为右利手，听力正常，母语为普通话，无神经性或精神疾病。所有被试均自愿参加实验，并签署知情同意书，

实验结束后获得一定报酬。实验已获得所在单位伦理审查委员会的批准。

2.2.2 实验设计

实验 1 采用两因素混合实验设计。被试间变量是说者身份（一致、不一致），被试内变量是词汇类型（学过、未学过），因变量是被试进行词汇判断的正确率和反应时。

2.2.3 实验材料

选取 72 个普通话双字词作为实验材料，包含 36 个真词与 36 个假词（音频材料见 <https://osf.io/zjn3s/>）。真词选自《现代汉语词典（全新版）》（2020）的常见双字词（如“衣服”），均为名词（词频为 124.47 ± 182.27 /百万词）。假词的构建方式是选用真字进行组合，形成可拼读但无意义的双字假词（如“床认”）。实验材料由 2 位母语为普通话的本科生（一男一女）录制。男性录音者（年龄：20 岁）音频材料的平均基频为 144.13 ± 32.48 Hz，女性录音者（年龄：19 岁）音频材料的平均基频为 188.14 ± 33.71 Hz。录音者普通话标准流利，均取得普通话二甲等级证书。录音者录制时均使用中性语调，正式录制前经过多次练习以保证录制的音频正确、流畅。录音采用 Cool Edit Pro 软件（Syntrillium Software），音频采样率设置为 44.1kHz。录制后，使用 Praat 软件（Boersma & Weenink, 1992）将所有音频材料的音量标准化为 70 dB SPL。

为了确保实验材料符合实验要求，在实验材料录制后，采用李克特 7 点评分，招募 10 名不参与本实验的被试对音频材料的熟悉度（“非常不熟悉”为“1”，“非常熟悉”为“7”）、可懂度（“完全不理解”为“1”，“完全能理解”为“7”）和具体性（“非常抽象”为“1”，“非常具体”为“7”）进行评定。评定结果显示，真词的熟悉度、可懂度和具体性均显著大于假词（ $ps < .001$ ），表明真词

与假词有很好的区分度。此外，不管是真词还是假词，学过的词的词频（仅真词）、熟悉度、可懂度、具体性与未学过的词均不存在显著差异（ $ps > .05$ ）。

2.2.4 实验程序

实验分成学习阶段和测验阶段，两阶段之间休息 1 分钟（李利等, 2006; 莫雷等, 2005）。实验开始前向被试详细介绍实验任务，确保被试充分理解任务内容后再进入学习阶段。学习阶段材料有 12 个真词、12 个假词和 24 个填充词（12 个真词和 12 个假词）。测验阶段材料包括学习阶段学习过的 12 个真词、12 个假词和未学习过的 12 个真词、12 个假词（与学习阶段的填充词不同）。其中，在说者身份一致组，被试在学习阶段和测验阶段听到的均为男性录音者的音频材料；在说者身份不一致组，被试在学习阶段听到男性录音者的音频材料，测验阶段则听到女性录音者的音频材料。表 1 呈现了材料安排的示例。每个试次开始，先向被试呈现 500ms 的注视点，提醒被试集中注意力，接着呈现 800ms 的音频刺激，被试在听到音频刺激后即可通过键盘上的“F”或“J”键按键判断听到的音频是真词还是假词。被试必须在 3s 内完成按键判断，按键与对应的判断在被试间平衡。为避免被试产生反应定势，相邻试次间设置了 600/800/1000ms 的空屏间隔（Jia et al., 2015; Yu et al., 2022）（见图 1）。实验共 96 个试次，每个条件 24 个试次，各条件随机呈现，实验总时长约 6 分钟。

2.3 实验结果

测验阶段的正确率和反应时结果见图 2。在 R 统计软件（R version 4.1.3）中使用 lme4 软件包（Bates et al., 2015）和 lmerTest 软件包（Kuznetsova et al., 2014）进行数据分析。对于正确率数据，采用广义线性混合效应模型（Generalized Linear Mixed-effects Model, GLMM）进行分析。对于反应时数据，首先剔除反应不正确的试次，接着进行 Kolmogorov-

表 1 实验材料安排示例（实验 1 以真词“衣服”和“汽车”为例、实验 2 以生物词“老虎”和“山羊”为例）

		学习阶段	测试阶段	
			学过	未学过
实验 1: 词汇决定任务	说者身份一致	男“衣服”	男“衣服”	男“汽车”
	说者身份不一致	男“衣服”	女“衣服”	女“汽车”
实验 2: 类别决定任务	说者身份一致	男“老虎”	男“老虎”	男“山羊”
	说者身份不一致	男“老虎”	女“老虎”	女“山羊”

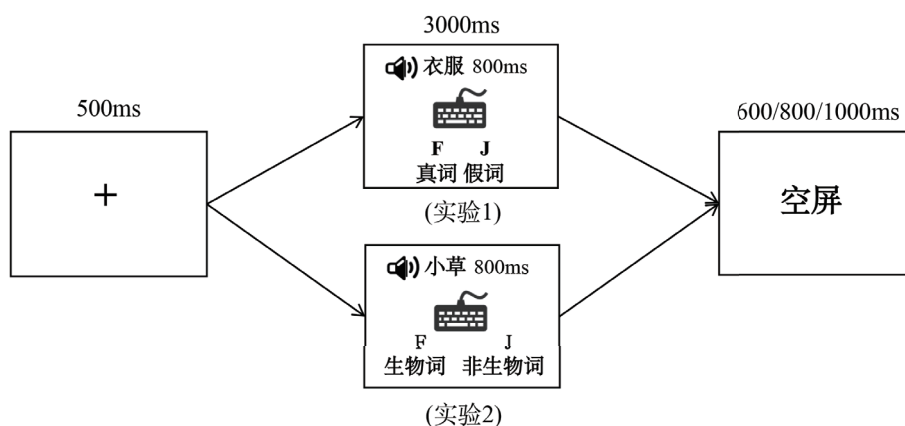


图1 实验流程图

Smirnov 检验,发现数据呈现非正态分布(总体分析: $D = .996$; 真词: $D = .999$; 假词: $D = .994$; $ps < .001$), 将其进行自然对数转换 ($\ln(rt)$) 使数据趋于正态分布之后,再使用线性混合效应模型 (Linear Mixed-effects Model, LMM) 进行分析。事后检验均用 lmerTest 数据包中的 emmeans 函数实现 (Lenth, 2021), 多重比较使用 Tukey 法对 p 值进行校正。

数据分析时首先根据实验设计对测验阶段的正确率和反应时数据进行分析(总体分析), 考虑到真词和假词的加工机制和策略可能存在差异(姜路遥, 李兵兵, 2023; 张钦, 张必隐, 1999; Orfanidou et al., 2011), 再分别对真词和假词的数据进行分析(真词和假词分析), 以探索说者身份信息对真词和假词词汇通达过程的影响。

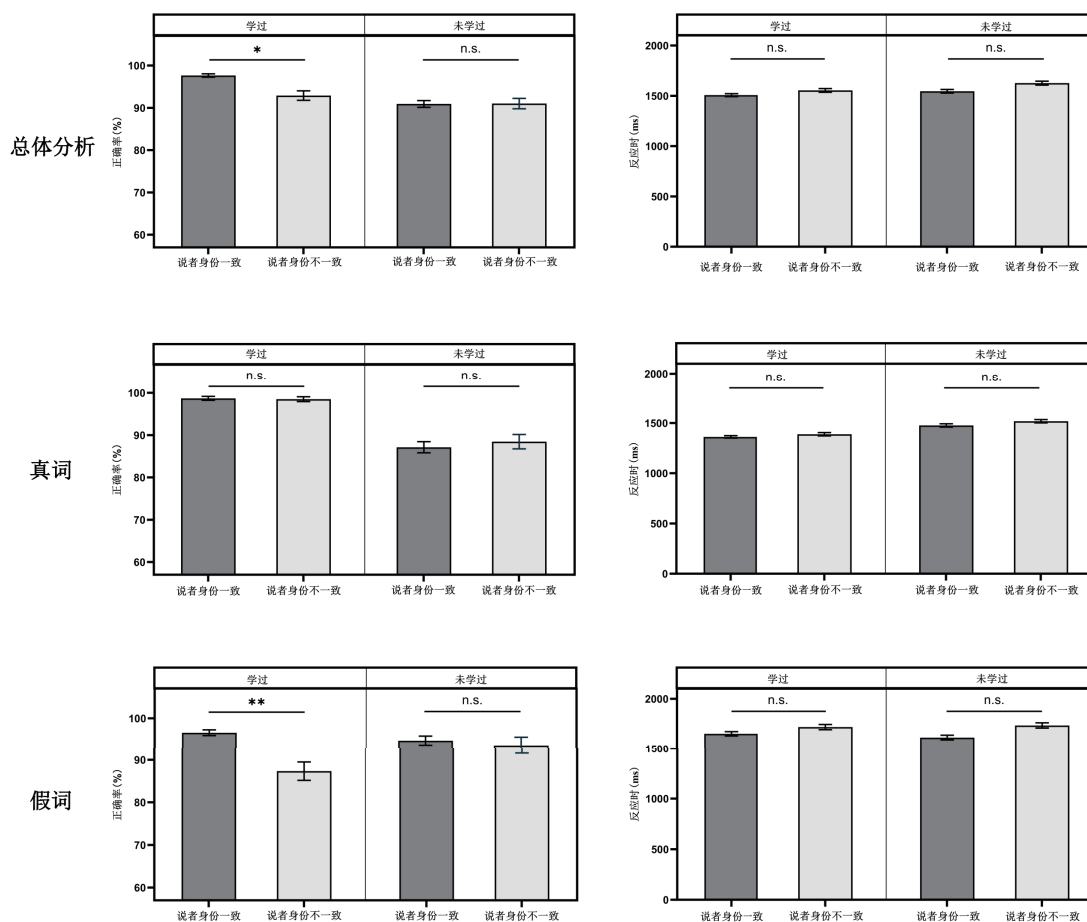


图2 实验1词汇决定任务各条件的正确率(%)和反应时(原始反应时, ms)

注: * $p < .05$; ** $p < .01$; *** $p < .001$; 误差棒为标准误

2.3.1 总体分析

以说者身份和词汇类型为固定效应,被试和实验材料为随机效应,对测验阶段进行总体分析。正确率最佳拟合广义线性混合效应模型和反应时最佳拟合线性混合效应模型为:正确率~说者身份*词汇类型+(1|被试)+(1|项目);反应时~说者身份*词汇类型+(1|被试)+(1|项目)。

正确率分析结果显示,说者身份主效应显著($\beta = -1.18, SE = .55, Z = -2.14, p < .05$),一致的正确率显著高于不一致;词汇类型主效应显著($\beta = -1.36, SE = .53, Z = -2.58, p = .01$),学过的词的正确率显著高于未学过的词;说者身份和词汇类型的交互作用不显著($\beta = .97, SE = .71, Z = 1.36, p > .05$)。根据本实验的研究逻辑,尝试对不同词汇类型条件下,说者身份一致和不一致的正确率进行比较。结果显示,对于学过的词,说者身份一致的正确率显著高于不一致($\beta = 1.18, SE = .55, Z = 2.14, p < .05$),而对于未学过的词,说者身份一致和不一致不存在显著差异($\beta = .22, SE = .50, Z = .44, p > .05$)。

反应时分析结果显示,说者身份主效应不显著($\beta = .03, SE = .04, t = .91, p > .05$);词汇类型主效应不显著($\beta = .03, SE = .03, t = .85, p > .05$);说者身份和词汇类型的交互作用不显著($\beta = .02, SE = .05, t = .34, p > .05$)。根据本实验的研究逻辑,对不同词汇类型条件下,说者身份一致和不一致的反应时进行比较。结果显示,不管是对于学过的词($\beta = -.03, SE = .04, t = -.91, p > .05$)还是未学过的词($\beta = -.05, SE = .04, t = -1.35, p > .05$),说者身份一致和不一致的反应时都不存在显著差异。

2.3.2 真词

以说者身份和词汇类型为固定效应,以被试和实验材料为随机效应,对测验阶段的真词进行分析。正确率最佳拟合广义线性混合效应模型和反应时最佳拟合线性混合效应模型分别为:正确率~说者身份*词汇类型+(1|被试)+(1|项目);反应时~说者身份*词汇类型+(1|被试)+(1|项目)。

正确率分析结果显示,词汇类型主效应显著($\beta = -2.51, SE = .99, Z = -2.54, p < .05$),学过的词的正确率显著高于未学过的词。说者身份的主效应不显著($\beta = -.21, SE = 1.08, Z = -.19, p > .05$)。说者身份和词汇类型的交互作用不显著($\beta = -.19, SE = 1.35, Z = -.14, p > .05$)。根据本实验的研究逻辑,

尝试对不同词汇类型条件下,说者身份一致和不一致的正确率进行比较。结果显示,不管是对于学过的词($\beta = .21, SE = 1.08, Z = .19, p > .05$)还是未学过的词($\beta = .40, SE = .86, Z = .46, p > .05$),说者身份一致和不一致的正确率都不存在显著差异。

反应时分析结果显示,词汇类型主效应显著($\beta = .08, SE = .03, t = 2.37, p < .05$),学过的词的反应时显著快于未学过的词。说者身份的主效应不显著($\beta = .02, SE = .03, t = .56, p > .05$)。说者身份和词汇类型的交互作用不显著($\beta = .01, SE = .05, t = .11, p > .05$)。根据本实验的研究逻辑,尝试对不同词汇类型条件下,说者身份一致和不一致的反应时进行比较。结果显示,不管是对于学过的词($\beta = -.02, SE = .03, t = -.56, p > .05$)还是未学过的词($\beta = -.02, SE = .03, t = -.71, p > .05$),说者身份一致和不一致的反应时都不存在显著差异。

2.3.3 假词

以说者身份和词汇类型为固定效应,以被试和实验材料为随机效应,对测验阶段的假词进行分析。正确率最佳拟合广义线性混合效应模型和反应时最佳拟合线性混合效应模型分别为:正确率~说者身份*词汇类型+(1+词汇类型|被试)+(1|项目);反应时~说者身份*词汇类型+(1|被试)+(1|项目)。

正确率分析结果显示,说者身份主效应显著($\beta = -1.45, SE = .51, Z = -2.84, p < .01$),一致的正确率显著高于不一致。词汇类型主效应不显著($\beta = .07, SE = .57, Z = .12, p > .05$)。说者身份和词汇类型的交互作用显著($\beta = 1.62, SE = .72, Z = 2.25, p < .05$)。简单效应检验的结果发现,对于学过的词,说者身份一致的正确率显著高于不一致($\beta = 1.45, SE = .51, Z = 2.84, p < .01$);对于未学过的词,说者身份一致和不一致的正确率不存在显著差异($\beta = -.17, SE = .62, Z = -.27, p > .05$)。

反应时分析结果显示,说者身份($\beta = .05, SE = .04, t = 1.27, p > .05$)和词汇类型($\beta = -.02, SE = .03, t = -.63, p > .05$)的主效应均不显著。说者身份和词汇类型的交互作用不显著($\beta = .03, SE = .04, t = .59, p > .05$)。根据本实验的研究逻辑,尝试对不同词汇类型条件下,说者身份一致和不一致的反应时进行比较。结果显示,不管是学过($\beta = -.05, SE = .04, t = -1.27, p > .05$)还是未学过的词($\beta = -.07, SE = .04, t = -1.99, p > .05$),说者身份一致和不一致的

反应时均不存在显著差异。

2.4 讨论

实验1首先在总体、真词和假词上都发现学过的词的表现优于未学过的词,表明学过的词出现了稳定的重复启动效应。在总体分析上,对于学过的词,说者身份一致的正确率显著高于不一致条件;对于未学过的词,说者身份一致和不一致的正确率不存在显著差异。进一步的分析发现,假词的结果与总体分析相同,但在真词中,不管是正确率还是反应时,对于学过和未学过的词,说者身份一致和不一致条件均不存在显著差异。从真词和假词的结果来看,说者身份信息一致的促进作用主要体现在学过的假词。在本实验中,真词为常见双字词,假词是由两个真字,而非两个假音(不符合拼读规则)组合而成的双字假词。基于这样的材料选取,结合词汇决定任务的要求,不管是真词还是假词的加工,听者均需要进行词汇通达,才能做出准确的判断(张钦,张必隐,1999)。虽然假词最终无法通达语义,但听者依然进行了词汇通达这一加工过程。那么基于实验逻辑,假词的结果在一定程度上说明了在词汇通达的过程中,说者身份信息是存储在记忆表征中的,它可能和语言信息形成整合的表征,以整合的方式影响词汇通达,支持了情节观。另外,从真词和假词的正确率来看,对于学过的词,真词的正确率比假词高,而对于未学过的词,真词和假词的正确率则没有明显差异。这可能是由于学过的真词的语义信息对测验阶段的词汇通达产生了自上而下的促进作用,听者不一定需要利用学习阶段形成的说者身份表征来进行加工。而对于未学过的词,听者在学习阶段并未建立语义信息的表征,在测验阶段也就没有观察到真词的促进作用。实验1主要考察的是词汇通达过程,实验2继续探讨说者身份信息是否以及如何影响口语词汇加工的概念理解过程。

3 实验2

3.1 实验目的

实验2采用内隐长时重复启动范式和类别决定任务,考察说者身份信息是否以及如何影响口语词汇加工的概念理解过程。

3.2 研究方法

3.2.1 被试

同实验1。

3.2.2 实验设计

同实验1。

3.2.3 实验材料

选取72个普通话双字词作为实验材料,包含36个生物词与36个非生物词(音频材料见<https://osf.io/zjn3s/>),均选自《现代汉语词典(全新版)》(2020)的常见双字词(如生物词“竹子”,非生物词“杯子”)。音频材料的录制与处理均同实验1。其中,男性录音者(年龄:20岁)音频材料的平均基频为 $142.71 \pm 37.05\text{Hz}$,女性录音者(年龄:19岁)音频材料的平均基频为 $187.15 \pm 38.76\text{Hz}$ 。为了确保实验材料符合实验要求,采用李克特7点评分,招募10名不参与本实验的被试对音频材料的熟悉度(“非常不熟悉”为“1”,“非常熟悉”为“7”)、可懂度(“完全不理解”为“1”,“完全能理解”为“7”)以及具体性(“非常抽象”为“1”,“非常具体”为“7”)进行评定。结果显示(见附录2),生物词的词频、熟悉度、可懂度和具体性与非生物词无显著差异($ps > .05$)。不管是生物词还是非生物词,学过的词的词频、熟悉度、可懂度和具体性与未学过的词也不存在显著差异($ps > .05$)。

3.2.4 实验程序

实验2同样分成学习阶段和测验阶段,两阶段之间休息1分钟。学习阶段材料有12个生物词、12个非生物词和24个填充词(也包括12个生物词和12个非生物词)。测验阶段包括学习阶段学习过的12个生物词、12个非生物词和未学习过的12个生物词、12个非生物词(与学习阶段的填充词不同)。表1呈现了材料安排的示例。每个试次开始,先向被试呈现500ms的注视点,提醒被试集中注意力,接着呈现800ms的音频刺激,被试在听到音频刺激后需要按键判断听到的音频是生物词还是非生物词(键盘上的“F”或“J”键,按键与相应的判断在被试间平衡),按键判断需要在3s内完成。为避免被试产生反应定势,相邻试次间也设置了600/800/1000ms的空屏间隔(见图1)。实验共96个试次,每个条件24个试次,各条件随机呈现,实验总时长约6分钟。

3.3 实验结果

实验2各条件的正确率和反应时见图3。数据分析方法与实验1相同。正确率最佳拟合广义线性混合效应模型和反应时最佳拟合线性混合效应模型为:正确率~说者身份*词汇类型+(1|被试)+(1|

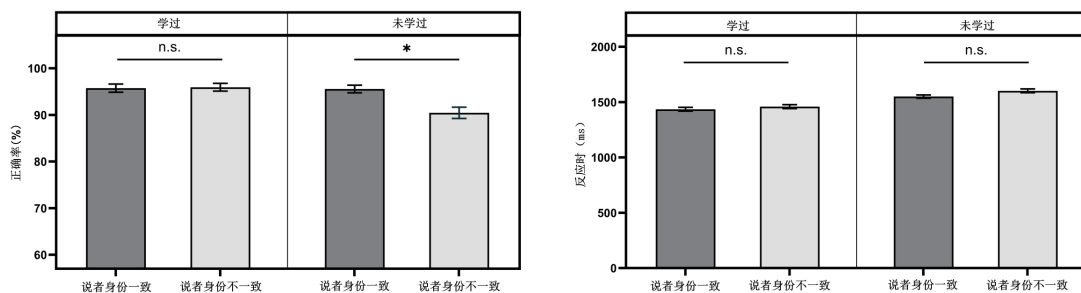


图3 实验2类别决定任务各条件正确率(%)和反应时(原始反应时, ms)

注: * $p < .05$; ** $p < .01$; *** $p < .001$; 误差棒为标准误。

项目); 反应时 \sim 说者身份 * 词汇类型 + (1 + 词汇类型 | 被试) + (1 | 项目)。

正确率分析结果显示, 说者身份主效应不显著 ($\beta = .36, SE = .55, Z = .65, p > .05$); 词汇类型主效应不显著 ($\beta = .38, SE = .50, Z = .76, p > .05$)。说者身份和词汇类型的交互作用显著 ($\beta = -1.50, SE = .70, Z = -2.15, p < .05$)。简单效应分析结果显示, 对于学过的词, 说者身份一致条件与不一致条件的正确率不存在显著差异 ($\beta = -.36, SE = .55, Z = -.65, p > .05$); 而对于未学过的词, 说者身份一致条件的正确率显著高于不一致条件 ($\beta = 1.14, SE = .54, Z = 2.11, p < .05$)。

反应时分析结果显示, 词汇类型主效应显著 ($\beta = .08, SE = .01, t = 3.72, p < .001$), 学过的词的反应时显著快于未学过的词。说者身份主效应 ($\beta = .01, SE = .01, t = .85, p > .05$)、说者身份和词汇类型的交互作用不显著 ($\beta = .01, SE = .01, t = .29, p > .05$)。根据本实验的研究逻辑, 尝试对不同词汇类型条件下, 说者身份一致和不一致的反应时进行比较。结果显示, 不管是学过 ($\beta = -.02, SE = .03, t = -.85, p > .05$) 还是未学过的词 ($\beta = -.03, SE = .03, t = -1.25, p > .05$), 说者身份一致条件和不一致条件的反应时均不存在显著差异。

3.4 讨论

实验2在反应时上观察到了学过的词的重复启动效应。更重要的是, 在正确率上, 对于未学过的词, 一致的说者身份起到了显著的促进作用。这一结果表明, 听者在学习阶段建立了说者身份的独立表征, 在测验阶段利用了这种独立的身份表征来判断未学过的词的概念。而对于学过的词, 并未发现一致的说者身份对其类别判断的促进作用, 说明当听者在学习阶段建立了词汇的概念表征之后, 无论测验阶段说者身份是否变化, 听者都能够利用已建立的概

念表征来对学过的词进行类别判断。根据实验2的结果, 在概念理解过程中, 说者身份信息 and 语言信息是独立表征的, 身份信息以独立的方式对口语词汇加工产生影响, 支持了发展的抽象观。

4 总讨论

4.1 说者身份信息对词汇通达的影响

实验1采用加工较浅的词汇决定任务考察说者身份信息对词汇通达过程的影响。实验结果发现, 说者身份一致对学过的假词的词汇通达有显著的促进作用。这与先前的研究发现一致, 说者身份的一致性会提升听者对学过的词的识别准确性, 但不影响未学过的词 (Clapp et al., 2023; Kapnoula & Samuel, 2019)。这些结果都表明, 说者身份信息和语言信息以整合的方式影响词汇通达, 支持了情节观的观点 (Cooper & Bradlow, 2017; Goldinger, 1996; McLennan & Luce, 2005)。然而, 这一结果与Orfanidou等人(2011)的研究结果并不一致。该研究同样采用词汇决定任务, 但并未发现说者身份信息一致性对词汇通达的促进作用。我们推测这可能与实验范式和注意偏向的不同有关。在实验范式上, Orfanidou等人(2011)采用的重复启动范式考察了较为即时的启动效应, 词汇的第一次呈现和第二次呈现之间的间隔仅为30秒左右。本研究采用的长时重复启动范式则考察了较为延迟的启动效应, 词汇的两次呈现时间间隔至少在1分钟。由此来看, 可能只有在加工比较缓慢的时候, 说者身份一致性才会影响词汇通达 (McLennan & Luce, 2005)。另一个可能的原因是注意偏向的差异, 不同的注意偏向可能会调节说者身份信息和语言信息的加工 (Ma et al., 2023)。Orfanidou等人(2011)在实验中明确要求被试忽略说者的变化, 这可能使得听者会更明确地偏向于关注词汇的语言信息, 造成说者身份

信息的影响减弱。而本研究并未明确要求被试忽略说者的变化，没有刻意引导被试的加工偏向，反映了在更为内隐的加工过程中，说者身份信息对词汇通达的影响。

4.2 说者身份信息对概念理解的影响

实验2采用了加工较深的类别决定任务考察说者身份信息对概念理解过程的影响。与实验1的结果不同，实验2发现说者身份的一致性提升了对未学过的词的判断准确性，而对学过的词的判断准确性则没有显著影响。这一结果更符合发展的抽象观（Belin et al., 2004; Blank et al., 2014）。说者身份信息和语言信息的表征是独立的，两种信息呈现并行加工的模式，使得说者身份信息以独立的方式对概念理解产生影响。Cai等人（2017）认为，当语音信号被初步编码之后，身份信息和语言信息就开始并行加工，并分别以不同的路径对后续的加工过程产生影响。本研究利用实验任务分离了词汇通达和概念理解过程，结合实验1的结果，说者身份信息对概念理解的影响与词汇通达并不相同，这也体现了听者在不同任务要求下加工口语词汇的灵活性。在词汇通达过程中，听者依赖于整合的说者身份信息和语言信息表征；而在概念理解过程中，听者可以利用独立的说者身份信息表征来理解词汇概念。此外，实验1与实验2均操纵了身份信息的物理维度（性别）。实验1的结果支持了情节观，与以往支持该观点的研究在身份信息维度的操纵上是一致的（Clapp et al., 2023; Kapnoula & Samuel, 2019）。而在以往支持发展的抽象观的研究中，Cai等人（2017）和Davies等人（2022）分别操纵了身份信息的社会维度和特质维度。虽然实验2对身份信息的操纵与他们不同，但是这种不同说明，发展的抽象观的观点不受身份信息维度的限制，无论身份信息是在物理、特质还是社会维度上发生变化，都会以独立的方式影响较深的概念理解。

4.3 考虑说者身份信息的口语词汇加工新观点

本研究两个实验的结果分别支持了情节观（实验1）和发展的抽象观（实验2），说明发展的抽象观和情节观确实分别适合于解释不同深度的口语词汇加工。虽然也有其他一些研究者认为需要综合考虑这两种观点（Cutler et al., 2010; Goldinger, 2007; Ma et al., 2023），但本研究对此提供了直接的证据支持。综合两个实验的结果，本研究也尝试通过整合发展的抽象观和情节观，提出考虑有说者身份信

息的口语词汇加工新观点。在语音信号输入后，听者首先通过声学加工提取出不同的语音特征，包括元音、辅音、声调等音素在内的语言信息特征以及音高、音色等说者身份信息特征。随后，听者会依据任务要求的加工深度形成灵活的语言信息表征和身份信息表征。在词汇通达等较浅的加工中，说者身份信息和语言信息形成整合的表征，两者以整合的方式影响词汇通达。在概念理解等较深的加工中，说者身份信息和语言信息则各自形成独立的表征，听者一方面可以基于语言信息的表征进行概念理解，另一方面也可利用说者身份表征来帮助理解词汇概念。

4.4 研究不足与展望

虽然本研究为口语词汇的认知加工提供了一定参考，但在研究方法上仍存在一定的局限性。首先，两个实验的学习阶段均呈现的是男性说者的音频材料，在测验阶段呈现的是男性和女性说者的音频材料。为避免学习阶段可能产生的性别效应，未来研究应考虑学习阶段音频材料的性别平衡问题。其次，Orfanidou等人（2011）使用了具象词作为实验材料，而实验1的真词同时包括了具象词和抽象词。尽管我们在数据分析时将词汇项目纳入了随机误差，对词汇的具体性进行了一定的统计控制，实验材料间的差异还是可能会影响研究结果。未来研究可对具象词和抽象词做对比分析，考察词汇的具体性对说者身份信息影响词汇通达的调节作用。再者，实验1真词的词频高于实验2，这可能会影响实验1和实验2结果的比较。虽然实验1主要是在假词上发现说者身份一致的促进作用，未来研究还是应该更严格控制实验材料的词频。此外，两个实验正确率的标准差均较大，说明可能存在较大的个体差异。未来研究可进一步探讨个体差异对口语词汇加工的影响。最后，两个实验采用的是认知行为实验，只能观察到被试的离线反应，对于认知机制的推测较为有限。未来可以考虑用脑电和磁共振等认知神经科学技术进一步探讨这一问题，从而更为细致地刻画口语词汇加工的认知过程。

5 结论

说者身份信息影响不同深度的口语词汇加工过程。在较浅的词汇通达中，说者身份信息和语言信息以整合的方式影响词汇通达；在较深的概念理解中，说者身份信息以独立的方式影响概念理解。综

合考虑发展的抽象观和情节观有助于更具社会性和生态性地解释口语词汇的认知加工。

参考文献

- 汉语大字典编纂处. (2020). *现代汉语词典*. 四川辞书出版社.
- 胡硯冰, 蒋晓鸣. (2023). “信”以传信, “疑”以传疑? 基于人声线索的可信度编码与解码. *心理科学*, 5, 1057-1066.
- 姜路遥, 李兵兵. (2023). 汉语听觉阈下启动效应: 来自听觉掩蔽启动范式的证据. *心理学报*, 4, 529-541.
- 李利, 莫雷, 王瑞明, 罗雪莹. (2006). 非熟练中—英双语者跨语言长时重复启动效应. *心理学报*, 5, 672-680.
- 明莉莉, 胡学平. (2021). 人类噪音加工的神经机制——来自正常视力和盲人的脑神经证据. *心理科学进展*, 12, 2147-2160.
- 莫雷, 李利, 王瑞明. (2005). 熟练中—英双语者跨语言长时重复启动效应. *心理科学*, 6, 10-15.
- 余可可, 周亚聪, 刘秉怡, 蔡涵涵, 王瑞明. (2021). 听话者对说话者噪音中语言学信息和副语言学信息的加工. *心理研究*, 1, 29-36.
- 张钦, 张必隐. (1999). 词汇决定任务中的策略因素. *心理科学*, 1, 75-76.
- 赵荣, 王小娟, 杨剑峰. (2016). 声调在汉语音节感知中的作用. *心理学报*, 48(8), 915-923.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48.
- Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences*, 8(3), 129-135.
- Blank, H., Wieland, N., & von Kriegstein, K. (2014). Person recognition and the brain: Merging evidence from patients and healthy individuals. *Neuroscience and Biobehavioral Reviews*, 47, 717-734.
- Boersma, P., & Weenink, D. (1992). *Praat: Doing phonetics by computer* (Version 6.2.06). [computer software]. <https://www.fon.hum.uva.nl/praat/>
- Bowers, J. S. (2000). In defense of abstractionist theories of repetition priming and word identification. *Psychonomic Bulletin and Review*, 7(1), 83-99.
- Cai, Z. G., Gilbert, R. A., Davis, M. H., Gaskell, M. G., Farrar, L., Adler, S., & Rodd, J. M. (2017). Accent modulates access to word meaning: Evidence for a speaker-model account of spoken word recognition. *Cognitive Psychology*, 98, 73-101.
- Clapp, W., Vaughn, C., Todd, S., & Sumner, M. (2023). Talker-specificity and token-specificity in recognition memory. *Cognition*, 237, 105450.
- Cooper, A., & Bradlow, A. R. (2017). Talker and background noise specificity in spoken word recognition memory. *Laboratory Phonology*, 8(1), 1-15.
- Cutler, A., Eisner, F., McQueen, J. M., & Norris, D. (2010). How abstract phonemic categories are necessary for coping with speaker-related variation. *Laboratory Phonology*, 10, 91-111.
- Davies, C., Porretta, V., Koleva, K., & Klepousiotou, E. (2022). Speaker-specific cues influence semantic disambiguation. *Journal of Psycholinguistic Research*, 51(5), 933-955.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175-191.
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5), 1166.
- Goldinger, S. D. (2007). *A complementary-systems approach to abstract and episodic speech perception*. In 16th International Congress of Phonetic Sciences, Saarbrücken, Germany.
- Hickok, G., & Poeppel, D. (2015). Neural basis of speech perception. *Handbook of Clinical Neurology*, 129, 149-160.
- Jia, S., Tsang, Y. K., Huang, J., Chen, H. C. (2015). Processing cantonese lexical tones: Evidence from oddball paradigms. *Neuroscience*, 305, 351-360.
- Kapnoula, E. C., & Samuel, A. G. (2019). Voices in the mental lexicon: Words carry indexical information that can affect access to their meaning. *Journal of Memory and Language*, 107, 111-127.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1-26.
- Lavan, N. (2023). The time course of person perception from voices: A behavioral study. *Psychological Science*, 34(7), 771-783.
- Lavan, N., Rinke, P., & Scharinger, M. (2024). The time course of person perception from voices in the brain. *Proceedings of the National Academy of Sciences*, 121(26), e2318361121.
- Lenth, R. (2021). *Emmeans: Estimated marginal means, aka least-squares means* (R package version 1.8.2). [computer software]. <https://CRAN.R-project.org/package=emmeans>
- Luthra, S. (2024). Why are listeners hindered by talker variability? *Psychonomic Bulletin and Review*, 31(1), 104-121.
- Ma, Y., Yu, K., Yin, S., Li, L., Li, P., & Wang, R. (2023). Attention modulates the role of speakers' voice identity and linguistic information in spoken word processing: Evidence from event-related potentials. *Journal of Speech, Language, and Hearing Research*, 66(5), 1678-1693.
- McLennan, C. T., & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning Memory and Cognition*, 31(2), 306-321.
- McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, 30(6), 1113-1126.
- Orfanidou, E., Davis, M. H., Ford, M. A., & Marslen-Wilson, W. D. (2011). Perceptual and response components in repetition priming of spoken words and pseudowords. *Quarterly Journal of Experimental Psychology*, 64(1), 96-121.
- Rodd, J. M., Lopez Cutrin, B., Kirsch, H., Millar, A., & Davis, M. H. (2013). Long-term priming of the meanings of ambiguous words. *Journal of Memory and Language*, 68(2), 180-198.
- Samuel, A. G. (2011). Speech perception. *Annual Review of Psychology*, 62(1), 49-72.
- Scott, S. K. (2019). From speech and talkers to the social world: The neural processing of human spoken language. *Science*, 6461, 58-62.
- Yu, K., Chen, Y., Yin, S., Li, L., & Wang, R. (2022). The roles of pitch type and lexicality in the hemispheric lateralization for lexical tone processing: An ERP study. *International Journal of Psychophysiology*, 177, 83-91.
- Zeelenberg, R., & Pecher, D. (2003). Evidence for long-term cross-language repetition priming in conceptual implicit memory tasks. *Journal of Memory and Language*, 49(1), 80-94.

The Role of Speaker's Identity Information in Spoken Word Processing

Yin Shuqi¹, Aierken Mukaidaisi¹, Shen Taiyu¹, Li Li², Yu Keke¹, Wang Ruiming¹

¹ Philosophy and Social Science Laboratory of Reading and Development in Children and Adolescents, Ministry of Education, & Center for Studies of Psychological Application, School of Psychology, South China Normal University, Guangzhou, Guangdong, 510631)

² School of International Culture, The Key Laboratory of Chinese Learning and International Promotion, South China Normal University, Guangzhou, Guangdong, 510631)

Abstract Considering the speaker's identity information provides a more social and ecological explanation of the cognitive processing of spoken words. However, whether and how speaker's identity information affects spoken word processing is controversial. The abstractionist view (including the early and developmental abstractionist views) and the episodic view hold different opinions on this issue. Moreover, previous studies have employed different experimental tasks that provide different evidence for these views. Based on our analyses of these previous studies, we propose that existing views may each be suitable for explaining different processes in spoken word processing. It is necessary to examine the role of speaker's identity information in spoken word processing requiring different processing depths. Based on this background, the present study focused on whether and how speaker's identity information affected lexical access and conceptual comprehension in spoken word processing. Addressing these issues can help us better understand spoken word processing.

The present study conducted two behavioral experiments and adopted the classic long-term repetition priming paradigm to minimize possible interference from explicit experimental tasks. Specifically, Experiment 1 adopted a lexical decision task to examine whether and how speaker's identity information affected lexical access in spoken word processing. Eighty-eight participants were recruited for the experiments and randomly divided into two groups (speakers' identities were consistent vs. inconsistent). The experiment contained learning and test phases. In the consistent group, participants would hear stimuli spoken by a male in both the learning and test phases; in the inconsistent group, participants would hear stimuli spoken by a male in the learning phase and by a female in the test phase. The experimental materials consisted of 36 real words (e.g., "/yi1fu2/", which means clothes in English) and 36 pseudowords (i.e., pronounceable but meaningless nonwords, e.g., "/ju4hong2/"). Participants needed to judge whether the auditory word was real or pseudo. Experiment 2 adopted a category decision task to examine whether and how speaker's identity information affected conceptual comprehension in spoken word processing. The participants and design were the same as Experiment 1, with 36 biological words (e.g., "/xiao3cao3/", which means grass in English) and 36 non-biological words (e.g., "/qian1bi3/", which means pencil in English) as experimental materials. Participants needed to judge whether the auditory word was biological or non-biological.

In Experiment 1, the performance of learned words was better than that of unlearned words, indicating a stable repetition effect. More importantly, in the overall analysis (including real words and pseudowords), for learned words, the accuracy of the consistent condition was significantly larger than the inconsistent condition; for unlearned words, there was no significant difference between the consistent and inconsistent conditions. Further analysis revealed that the results for pseudowords were the same as the overall analysis, but for real words, there were no significant differences in either accuracy or reaction time between the consistent and inconsistent conditions for both learned and unlearned words. In Experiment 2, the response times of learned words were significantly shorter than those of unlearned words, suggesting the repetition effect of learned words. However, in contrast to Experiment 1, the accuracy of the consistent condition was significantly larger than the inconsistent condition for unlearned words, while there was no such difference for learned words.

Speaker's identity information influences the processing of spoken word differently depending on the processes. Specifically, speaker's identity consistency facilitation for learned words in the lexical decision task suggested that the representation of the speaker's identity was integrated with linguistic information and would affect lexical access integrally, supporting the episodic view. In contrast, speaker's identity consistency facilitation for unlearned words in the category decision task suggested that the speaker's identity and linguistic information would be represented separately and affect conceptual comprehension independently, supporting the developmental abstractionist view. Integrating the developmental abstractionist and episodic views helps us better understand spoken word processing.

Key words spoken word processing, identity information, linguistic information, lexical access, conceptual comprehension