

《心理学报》审稿意见与作者回应

题目：汉语口语词汇产生中的多重音韵激活：单词翻译任务的 ERP 研究

作者：张清芳 钱宗愉 朱雪冰

第一轮

审稿人 1 意见：

意见 1：文中提到虽然较多研究在印欧语系中探测到了非目标项的音韵激活，但采用单词翻译任务“在荷兰语-英语和西班牙语-卡特兰语两种不同的双语情境下均未观察到任何的音韵效应，未探测到非目标项的多重音韵激活（Bloem & La Heij, 2003; Bloem, Boogaard, & La Heij, 2004; Navarrete & Costa, 2009）”。如果这一任务在印欧语系中都难以探测到非目标项的音韵激活，在汉语中选取这一任务是否恰当？研究结果是否不足以支持“汉语和印欧语系的口语产生分别遵循了独立两阶段模式和交互激活的模式”这一结论？文中还提到“汉语属于表意文字，语义和音韵节点间的联结较弱，而印欧语系中语义和音韵节点间的联结较强”，这一解释较模糊，为什么书面文字的特点会影响口语产生的机制？

回应：感谢审稿专家的问题。目前印欧语系中采用单词翻译任务考察背景图片音韵激活的共有三篇文献（Bloem & La Heij, 2003; Bloem, Boogaard, & La Heij, 2004; Navarrete & Costa, 2009）。在前两项 Bloem 等的研究中，均采用荷兰语-英语双语者和相同的实验材料均未发现多重音韵激活。Navarrete 和 Costa（2009）认为需要对此阴性结果的可靠性进行验证，因此他们在另一组被试中重复了上述实验。在实验 1a 中实验材料的音韵相关比例与 Bloem 等的研究相同，要求双语被试在忽略背景图片的情况下对呈现的单词进行翻译，结果发现音韵相关的背景图片会加快翻译潜伏期。在实验 1b 中降低了音韵相关的比例后未发现在翻译潜伏期上的显著差异，但发现了音韵无关条件较音韵相关条件被试的错误率显著增加。两个实验采用单词翻译任务在印欧语系研究中探测到了非目标项的音韵激活。针对汉语口语产生过程，Zhang 和 Zhu (2016)采用相同实验设计和相同的单词翻译任务考察了汉语中背景图片的音韵激活，在行为结果上不论音韵相关比例为 50%还是 25%汉语单词翻译任务中都不存在背景图片的音韵激活，而 Navarrete 和 Costa（2009）发现在卡特兰语和西班牙语翻译过程中存在多重音韵激活。汉语的结果与印欧语系的结果在单词翻译任务中不同。为了进一步考察汉语口语词汇产生中的多重音韵激活，我们采用了更为敏感的时间分辨率高的事件相关电位

技术对这一问题进行研究。文中表述有误，我们已经做了修改（见 P.3）。

单词翻译任务是否适合探测多重音韵激活？在单词翻译任务中，给被试呈现一个英文单词和一幅图片，要求被试将英文单词翻译成汉语。在实验设置上，图和词二者容易分离，避免了 PPI 范式所带来的注意混淆或注意不足的问题。单词翻译过程包括词汇识别、（双语）概念激活、词条选择、音韵编码，以及发音执行等过程。同时已有研究发现当非目标项存在视觉输入信息时更容易探测到音韵激活（Oppermann, Jescheniak, & Schriefers, 2008），因此采用视觉呈现的单词翻译任务有利于探测到非目标项的多重音韵激活（见 P.4-5）。本研究中的行为和 ERP 结果中都一致地发现了被试对背景图片进行了加工，提取到了背景图片的语义信息，证实了实验设计的有效性，结果发现了经典的语义效应而没有出现音韵效应。在结果部分，我们增加了贝叶斯分析，结果发现了实质性证据支持零假设， $BF_{01} = 3.54$ ，表明在（假定不存在音韵效应的）零假设下出现当前数据的可能性是在（假定存在音韵效应的）备择假设下可能性的 3.54 倍，以证实背景图片的音韵信息确实没有得到激活（见 P.9）。

有关印欧语系和汉语在多重音韵激活上的不同结果，单个研究的发现确实不能肯定地得出“汉语和印欧语系的口语产生分别遵循了独立两阶段模式和交互激活的模式”这一结论。尽管如此，我们综合多个研究的发现和对比（汉语：张清芳，杨玉芳，2006；Zhang & Zhu, 2016；Zhang, Zhu, & Damian, 2018；Zhu, Damian, & Zhang, 2015；Zhu, Zhang, & Damian, 2016；印欧语系：Damian & Martin, 1999；Starreveld & La Heij, 1995, 1996；Dell’ Aquca, et al., 2010）得出了上述对比性的结论（见讨论 4.1 部分），该结论仍然需要更进一步的验证。

为什么会出现不同的加工模式？我们试图从语言系统的特点进行解释。汉字是表意文字，语义和音韵之间的联结松散，印欧语系为拼音文字，语义和音韵之间的联结相对紧密。口语词汇产生包括了词汇选择和音韵编码两个阶段，语义激活和音韵激活分别发生在这两个阶段。在汉语口语词汇产生过程中，由于语义激活和音韵激活的联结较弱，所以非目标词的音韵激活较弱，而在印欧语系中，由于两者之间的联结较强，能够探测到非目标项的音韵激活。原来的解释模糊不具体，我们修改了相应的部分（见 P.14）。

参考文献：

Oppermann, F., Jescheniak, J. D., & Schriefers, H. (2008). Conceptual coherence affects phonological activation of context objects during object naming. *Journal of Experimental Psychology-Learning Memory and Cognition*, 34(3), 587-601.

意见 2：是否有特殊原因导致剔除奇异值的标准前后不一致？行为数据采用 3 个标准差，相关分析中采用 2.5 个标准化残差。

回应：感谢审稿专家的建议。为统一剔除奇异值的标准，我们按照 3 个标准化残差剔除奇异值，重新对反应时和 ERP 波幅进行了分析，其结果模式与之前相同。结果修改详见“3.3 反应时和 ERP 波幅的相关”蓝色字体部分（见 P.12）。

意见 3：行为数据的统计检验中不少效应只达到被试分析显著，是否可进一步采用 minF’ 进行检验？行为数据在项目分析中难以达到显著，这是否由材料的某些属性导致？是否可以进一步探讨背后原因？

回应：感谢审稿专家的建议。在选用特定语言材料的心理语言学研究领域，有研究者质疑采用被试分析（F1）、项目分析（F2）所得到的结果能否推广至所有的被试和所有的材料（Raaijmakers, 2003; Raaijmakers, Schrijnemakers, & Gremmen, 1999）。Clark（1973）认为把语言当成固定效应的做法是错误的（the language-as-fixed-effect fallacy），并提出了同时考虑被试变异和实验材料变异的 minF’ 检验。Baayen, Davidson 和 Bates（2008）认为采用作为 F1 和 F2 结合的 minF’ 检验过于复杂，在心理语言学领域使用度不高，因此提出使用基于最大似然估计的混合线性模型作为替代。

目前，大量的心理语言学研究都采用了混合线性模型来分析数据。本研究修改稿中采用混合线性模型分析方法，同时考虑被试和项目所产生的随机效应，考察了自变量对因变量的影响。具体结果详见正文“3.1 行为结果”部分（见 P.7-8）。

参考文献：

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with the language-as-fixed-effect fallacy: Common misconceptions and alternative solutions. *Journal of Memory and Language*, 41, 416–426.

Raaijmakers, G. (2003). A further look at the language-as-a-fixed-effect fallacy. *Canadian Journal of Experimental Psychology*, 57, 141–151.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412.

意见 4：相关分析中的 p 值是否经过校正？图 2 的地形分布图和图 3 中显示中中区应该是 ERP 差异波幅最大的兴趣区，但其差异波幅与反应时差值的相关不显著。这是否表明相关分析的结果并不稳定？

回应：感谢审稿专家的问题，我们同意您的看法，相关分析的结果是不稳定的。口语产生的 ERP 研究较少，我们是尝试着进行反应时和 ERP 波幅的相关分析，目的是考察波幅差异的

认知涵义，并为以后这方面的研究提供比较的基础。修改稿中的 p 值都经过了 FDR 校正，结果发现中后区显著的负相关 ($p = 0.006$) 经 FDR 校正后边缘显著 ($p = 0.054$)，但在中后区还是可以观测到反应时和 ERP 波幅之间的线性趋势。对于不稳定的相关分析结果我们在“4.2 语义促进效应”部分增加了讨论，并指出了相关分析结果的不稳定性 (见 P.15)。

意见 5: 另有一些细节问题已在原文中高亮标出，请查看。

细节 (1): “研究者发现设置与目标词存在语义关系的线索词或图片会缩短图画命名【单词翻译?】的潜伏期”

细节 (2): “被试的任务是把所呈现的英语单词翻译呈【成?】荷兰语 (母语)”

细节 (3): “(查自《现代汉语通用词表》(2003)【, 2003】)”

细节 (4): “位于左眼上下框【侧?】的电极记录垂直眼电 (VEOG)”

细节 (5): “进一步的简单简单效应分析表明”

细节 (6): “音韵相关条件与音韵无关条件命名【单词翻译?】潜伏期差异不显著”

细节 (7): “语义相关条件下命名潜伏期被试分析显著快于语义无关条件”

细节 (8): “在 6 个时间窗口中分别做简单简单效应分析”

细节 (9): “已有的研究结果一致表明汉语的口语词汇产生过程遵循独立两阶段的模式”

细节 (10): “翻译任务的反应时越短【语义促进效应越强?】”

细节 (11): “较少有研究采用词语翻译任务考察口语词汇出声【产生?】过程”

回应: 非常感谢您的仔细阅读和建议, 我们已对您所指出的细节问题逐一进行了更正和修改。

其中细节 (4): “位于左眼上下框【侧?】的电极记录垂直眼电 (VEOG)”, 由于本研究 VEOG 电极点放置的位置为左眼眼眶上和眼眶下正中间的位置, 故未修改为“侧”。

.....

审稿人 2 意见:

研究问题有意义, 研究方法恰当, 对汉语语言产生具有显著的理论意义。修改意见如下:

意见 1: 实验仅分析 200 ms 后的 ERP 波幅, 建议加上 0-200 以及 baseline 的 ERP 分析结果。

回应: 感谢审稿专家的建议。我们对 0~200 ms 以及 baseline 的 ERP 数据进行了方差分析。

在 -200~0 ms 的时间窗口内, 重复测量方差分析结果发现, 主效应、二重、三重和四重交互作用均未达到显著, $p_s > 0.09$ 。在口语产生的相关研究中, 研究者较少分析刺激呈现前的

ERP 数据，大多关注刺激呈现后的 ERP 结果，因此我们在修改稿中未加入基线窗口内的分析结果。

根据审稿专家的建议，我们分析了 0~200 ms 时间窗口内的结果，补充在“3.2 ERP 结果”部分（见 P.9-10），该时间窗口内的 9 个兴趣区均未出现显著的音韵效应或语义效应。

意见 2：建议阐明 ERP 分析的时间窗口是如何选取的。

回应：非常感谢审稿专家的建议。在修改稿中，我们将分析的时间窗口选择为 0~600 ms，主要是基于以下考虑：（1）在口语产生的研究中均采用刺激锁时的分析方法，即从刺激呈现时间（0 ms）开始对 ERP 进行分析（Costa et al., 2009; Dell' Acqua et al., 2010; Python et al., 2018; Zhu et al., 2015）。（2）由于口语产生涉及肺、喉、声带等肌肉和发音器协同运动，发音所产生的运动伪迹会影响后期 EEG 信号，为避免运动伪迹的影响，研究者通常不会分析靠近发音阶段时间窗口内的波形。Ouyang 等（2016）通过实验证明与发音相关的伪迹出现在发音前的 300 ms 或更早之前。在 Jost, Radman, Buetler 和 Annoni（2018）关于单词翻译的 ERP 研究中，二语翻译成一语任务的平均反应时约为 1200 ms，为避免运动伪迹的影响，Jost 等选取的时间窗口为 0~700 ms。本研究中被试平均单词翻译潜伏期为 935 ms，因此我们选择对 0~600 ms 时间窗口内的波形进行分析（相同的时间窗口选取见 Costa et al., 2009; Zhu et al., 2015）。此外，在 ERP 数据处理过程中，我们删除了反应时在三个标准差以外的试次（平均潜伏期为 935ms，标准差为 65.8，3 个标准差以外的最短潜伏期为 738 ms），保证所留下的试次发音没有早于 600 ms，排除了命名较快试次发音运动的影响。我们将选择时间窗口的原因增加到了正文的“3.2 ERP 结果”部分（见 P.9）。

参考文献：

Ouyang, G., Sommer, W., Zhou, C. S., Aristei, S., Pinkpank, T., & Rahman, R. A. (2016). Articulation artifacts during overt language production in event-related brain potentials: Description and correction. *Brain Topography*, 29(6), 791-813.

意见 3：在一些时间窗口发现了相关类型的主效应（在 200~300 ms 时间窗口内，相关类型主效应显著），如何解释？

回应：感谢审稿专家的问题。根据本研究的实验设计，相关类型这一变量分为两个水平，分别是音韵和语义，即在同一张背景图片上，所呈现英文单词的汉语对译词与背景图片之间存在语义（相关或无关）、音韵（相关和无关）两种关系。在 200~300 ms 时间窗口内显著的相关类型主效应表明语义关系和音韵关系对波幅产生了不同的影响，表现为语义关系较音韵关

系波幅更正 ($1.57-1.30 = 0.27 \mu\text{V}$)。

由于本研究考察的主要问题是背景图片的语义或音韵信息能否得到激活,即我们更加关注简单效应的结果,在语义条件下,相关与无关是否存在差异,在音韵条件下,相关与无关是否存在差异,简单效应分析的结果中未发现 200~300 ms 时间窗口内存在语义或音韵效应。因此未在正文中对 200~300 ms 时间窗口内存在的相关类型主效应进行讨论。

意见 4: 结果发现的相关性主效应 (相关性主效应边缘显著), 如何解释?

回应: 感谢审稿专家的问题。本研究中的相关性变量包含两个水平, 相关和无关。在相关水平下, 英文单词的汉语对译词与背景图片存在语义或音韵上的联系, 而无关条件下, 汉语对译词与背景图片不存在语义或音韵上的联系。在 400~500 ms、500~600 ms 时间窗口内显著的相关性主效应和 300~400 ms 时间窗口内边缘显著的相关性主效应, 表现均为相关较无关波幅更正 ($1.20 - 0.82 = 0.38 \mu\text{V}$; $1.61 - 1.14 = 0.47 \mu\text{V}$; $1.71 - 1.50 = 0.21 \mu\text{V}$)。该结果进一步证实被试确实注意到了背景图片, 而不仅仅是完成一个单词翻译的任务, 所以背景图片与目标词是否存在关系会对 ERP 产生影响。由于我们重点关注的问题是语义条件下和音韵条件下相关性是否分别显著, 所以未在正文中对相关性的主效应展开解释, 而是重点对简单效应的结果进行了讨论。

意见 5: 如表 1 所示, 600-700 ms, 发现了 ROI 的音韵效应; 而作者似乎忽略这些效应。

回应: 感谢审稿专家的问题。参考已有研究, 我们建议靠近发音阶段的 ERP 数据容易受到肌肉运动的影响, 所以在修改稿中仅保留了 0~600 ms 的结果 (见对意见 2 的回复)。600~700 ms 时间窗口内的 ERP 数据容易受到运动伪迹的影响, 因此在修改稿中我们未分析这一时间窗口内的数据。

意见 6: 建议将实验材料的相关属性用 table 表述。

回应: 感谢您的建议, 已将“2.2 材料”部分有关实验材料属性的文字表述更改为表格形式呈现 (见 P.6)。

第二轮

审稿人 1 意见：

意见 1：“汉字是表意文字，语义和音韵之间的联结松散，印欧语系为拼音文字，语义和音韵之间的联结相对紧密。”作者仍然没有说明为什么书面文字的特点会影响口语产生的机制。口语习得发生在文字习得之前，为什么文字的特点会影响口语产生中语义和音韵之间的联结？

回应：感谢审稿老师的问题。在口语词汇产生的过程中，包括了概念准备、词汇通达和语音编码、发音等过程，说话者要激活概念，提取概念对应的词语的语义和音韵信息。在口语产生过程中，语义激活扩散至音韵激活，之后进行语音编码产生目标词语。汉语口语词汇产生的研究表明词汇选择和音韵编码两个阶段之间无交互作用，语义激活和音韵激活之间无时间上的重叠（Zhu, Damian, & Zhang, 2015; Zhu, Zhang, & Damian, 2016）。与汉语不同，印欧语系的语言中发现这两个阶段之间存在交互作用（Damian & Martin, 1999; Starreveld & La Heij, 1995, 1996），且语义激活和音韵激活几乎是同时产生的（Dell' Aquca, et al., 2010）。

汉语和印欧语系的口语产生分别遵循了独立两阶段模式和交互激活的模式，这主要涉及到语言加工过程中语义和音韵之间的联结强度。在汉语口语词汇产生过程中，语义和音韵之间的联结强度可能较弱，而在印欧语系中语义和音韵之间的联结可能较强。词汇产生与词汇理解过程都涉及到在心理词典中提取相应的正字法、语义和音韵这三类信息之间的联结及其加工。例如，在图画-词汇干扰实验任务中，采用图画命名任务，设置与目标字存在语义相关、正字法相关、音韵相关的干扰字以及无关字，以视觉方式呈现干扰字，结果发现了在图画命名这一经典的口语词汇产生任务中出现了显著的语义抑制效应、正字法促进效应和音韵促进效应，表明口语词汇产生过程中讲话者激活了目标字的字形信息，且促进了图画命名（汉语：Zhang & Weekes, 2009; Zhang, Chen, Weekes, & Yang, 2009; Zhao, La Heij, Schiller, 2012; 英语：Damian & Bowers, 2003）。在书写产生过程中，尽管最终输出的字形信息，研究者同样发现了书写产生过程中音韵信息的激活（汉语：Qu, Damian, Zhang, & Zhu, 2011; Wang & Zhang, 2015; 英语：Zhang & Damian, 2010）。

由于汉字是表意文字，用形旁表意是其显著的特征（Zhou & Marslen-Wilson, 1999），尽管汉字的声旁也能在一定程度上提示语音，但是汉字字形和语音之间并未形成紧密的对应关系（Zhou & Marslen-Wilson, 2000）。例如，正字法类似的字，其发音可以完全不同（例如“床”和“庆”）；正字法完全不同的字，发音却可以相似（例如“床”和“闯”）。由于这样的特点，使得词汇的语义和语音之间的联系比较松散。相比而言，拼音文字由于存在形音对应的规则，

形音的对应关系十分密切（陈宝国，王立新，彭聃龄，2003），这使得语义和语音之间的联结相对紧密。这些特点不仅会影响词汇产生过程（Zhu, Damian, & Zhang, 2015; Zhu, Zhang, & Damian, 2016），而且对词汇阅读过程产生影响（陈宝国等，2003）。在语言加工过程中，人们经过多次地在心理词典中激活语义和语音，形成了语言加工的联结网络，呈现出不同的联结强度。上述语言特点的不同导致了人们在使用语言的过程中语义和音韵之间的联结强度不同，激活的扩散程度也不同，导致了不同的口语词汇产生模式。

我们尝试从语言文字的特点对不同语言中不同的口语词汇产生模式进行了解释，这需要更进一步的对比研究进行验证。上一稿中的表述未能清楚地阐述我们的观点，在修改稿中对此做了详细的解释（见正文 P.11），文中引用的文献也已增添至列表中。

意见 2：LMEM 分析应报告模型的公式和各变量的编码方法（dummy coding, deviation coding...）。模型比较的统计参数也应报告。

回应：感谢审稿老师的建议。本研究关注的变量均为二分变量（相关类型：音韵、语义；相关性：相关、无关；重复次数：第一次、第二次），在建模过程中自动处理为哑变量（0, 1 编码）。我们主要关注固定因子效应是否显著，模型中各个固定因子的系数和显著性通过 β 值和 p 值进行报告。

多数采用混合线性模型进行统计分析的文献一般均未报告最优拟合模型公式（例如，Chen, O'Seaghdha, & Chen, 2016; Hutson, Damian, & Spalek, 2013; Janssen, Hernandez-Cabrera, van der Meij, & Barber, 2015; Wang, Wong, & Chen, 2018），但我们认为您所提出的呈现最优拟合公式是一个很好的建议，便于读者理解数据处理的过程。因此，我们增加了两个脚注呈现了相应的最优拟合模型的公式（见脚注 1，脚注 2）。此外，模型比较的统计参数（卡方值）已经补充在正文“3.1 行为结果”部分（见正文 P.5~6）。

参考文献：

Chen, J. Y., O'Seaghdha, P. G., & Chen, T. M. (2016). The primacy of abstract syllables in Chinese word production. *Journal of Experimental Psychology-Learning Memory and Cognition*, 42(5), 825–836.

Hutson, J., Damian, M. F., & Spalek, K. (2013). Distractor frequency effects in picture-word interference tasks with vocal and manual responses. *Language and Cognitive Processes*, 28(5), 615–632.

Janssen, N., Hernandez-Cabrera, J. A., van der Meij, M., & Barber, H. A. (2015). Tracking the time course of competition during word production: Evidence for a post-retrieval mechanism of conflict resolution. *Cerebral Cortex*, 25(9), 2960–2969.

Wang, J., Wong, A. W. K., & Chen, H. C. (2018). Time course of syllabic and sub-syllabic processing in Mandarin word production:

意见 3: “结果发现了实质性证据支持零假设 (Jeffreys, 1961), $BF_{01} = 3.54$, 表明在 (假定不存在音韵效应的) 零假设下出现当前数据的可能性是在 (假定存在音韵效应的) 备择假设下可能性的 3.54 倍, 即背景图片与目标词音韵相关或无关对翻译潜伏期的影响不存在差异。”读者可能对贝叶斯分析并不熟悉, 是否可以提供一些依据来说明 $BF_{01} = 3.54$ 的统计意义 (例如一般认为 BF_{01} 达到多少时支持零假设) ?

回应: 感谢审稿老师的建议。Jeffreys (1961, P.432) 提出的贝叶斯因子的决策标准如下: $BF_{01} = 1$, 没有证据; $BF_{01} = 1\sim 3$, 存在较弱证据支持零假设; $BF_{01} = 3\sim 10$, 存在实质性证据支持零假设; $BF_{01} = 10\sim 30$, 存在强证据支持零假设; $BF_{01} = 30\sim 100$, 存在非常强证据支持零假设; $BF_{01} > 100$, 存在极强证据支持零假设。本研究中统计结果显示 $BF_{01} = 3.54$, 表示零假设 (假定不存在音韵效应的) 成立的可能性是备择假设 (假定存在音韵效应的) 成立可能性的 3.54 倍, 存在实质性证据支持零假设, 表明背景图片与目标词音韵相关或无关对翻译潜伏期的影响不存在差异。文中对此做了相应的修改 (见正文 P.6~7)。

意见 4: 根据作者对结果的解释, “400~600 ms 时间窗口为单词翻译过程中概念的激活”。如果音韵编码发生在概念激活之后, 那么修改稿中选取 0~600 ms 时间窗进行 ERP 数据分析是否合理? 原稿中 600~700 ms 时间窗内的音韵效应是否有一定意义? 作者选取时间窗时所参考的 Zhu et al., 2015 中, 平均反应时为 800 ms 左右, 而本文中的平均反应时更长 (> 900 ms)。这是否也支持将数据分析的时间窗延长至 700 ms ?

回应: 感谢审稿老师的建议。本研究中被试单词翻译潜伏期的均值为 935 ms, 选择 0~700 ms 作为分析的时间窗口受到发音运动的影响较小, 我们接受您的建议, 在第二次修改稿中重新纳入了 600~700 ms 时间窗口内的数据分析结果 (见表 3)。数据分析结果表明, 在 600~700 ms 的时间窗口内, 出现了边缘显著的音韵效应和语义效应。根据数据分析的结果, 我们对图 2 进行了相应的修改, 修改后的图 2 在三个区域 (左中、中中、右中) 上呈现出显著的语义效应 (400~600 ms), 边缘显著的语义效应 (600~700 ms), 以及边缘显著的音韵效应 (600~700 ms)。

我们采用贝叶斯分析检验对备择假设的支持程度, 结果显示 (见表 4), 处于边缘显著水平的音韵效应对应的 BF_{10} 最小值为中中区域的 2.01, 最大值为右中区的 3.27, 表明存在比较可靠的证据支持了音韵效应在右中区的存在。处于显著和边缘显著水平的语义效应对应

的 BF10 值最小为左前区域的 0.93，最大为 500~600 ms 时间窗口内中中区域的 50.25，多个区域包括左中、左后、中前、中中以及右中区域的 BF10 值在 400~600 ms 时间窗口内都大于 10，表明有强证据支持了语义效应的存在 (Jeffreys, 1961, P432)。上述结果表明，尽管在神经层面上在某个特定兴趣区保留了多重音韵激活，但在行为结果上不足以产生多重音韵激活的效应，相比而言，语义效应的脑电激活模式分布范围广且效应强，在行为结果上也表现为显著的促进效应。

音韵效应在行为和脑电之间的不一致情况与一些已有口语词汇研究的发现类似 (Cai, Yin, & Zhang, 2020; Qu, Damian, & Kazanina, 2012; Zhang & Damian, 2019)，我们认为这是由于脑电指标反映了在线的实时加工，而反应时是最终输出的结果反映，会受到一系列加工过程的影响。基于已有研究的结果和本研究的分析，我们倾向于认为讲话者在汉语口语词汇产生过程中，其神经层次上仍然保留了对多重语音激活的敏感性 (相应讨论 P.11)。

意见 5：“我们选定 ERP 波幅分析的截止时间为刺激呈现后的 600 ms，以 100 ms 为间隔分割【分别】分析刺激呈现后 0~600 ms 内的 ERP 波幅”；“所以虽然本研究相关结果不稳定，但结合单词翻译任务中概念中介【?】的时间窗口应晚于图画命名任务下概念激活的时间窗口”。

回应：感谢您的认真审阅。在修改稿中我们改正了这两处错误 (见 P.7、P.13)。

审稿人 2 意见：建议发表。

回应：感谢您对修改稿的肯定。

第三轮

审稿人 1 意见：

意见 1：请将脚注 1 中的“ $RT \sim (\text{相关类型} + \text{相关性})^2 + (1|\text{participant}) + (1|\text{item})$ ”和脚注 2 中的“ $RT \sim (\text{相关类型} + \text{相关性} + \text{重复次数})^3 + (1|\text{participant}) + (1|\text{item})$ ”移入正文；而脚注中的其他部分可以删去。

回应：我们根据您的建议修改了文章 (见正文 P.5~6)。

意见 2：图 2 地形图旁的文字标注太小，请稍稍放大。以上两点修改后推荐发表。

回应：根据期刊对于图标的字体大小要求，我们修改了地形图上的文字标注（见正文 P.9）。

编委复审

编委复审意见：

该文经专家两轮审稿，审稿专家提出了非常详细而有建设性的意见，作者根据意见进行了修改、补充和解释，二审审稿人均同意该文发表。我同意两位审稿人的意见，认为该文为中文认知提供了一些新的数据，对该领域有一定的推进，同意推荐发表。

主编终审

主编终审意见：

本研究借助运用事件相关电位技术，对被试进行翻译命名任务时是否受到背景图片音韵或语义干扰词的影响进行了考察，具有一定理论和应用价值。本论文研究方法选用恰当，数据处理过程科学规范，多次修改后，本论文已经达到《心理学报》发表文章的相关要求。