

《心理学报》审稿意见与作者回应

题目：词汇识别中歧义词素语义加工：ERP 研究

作者：赵思敏 吴岩 李天虹 郭庆童

第一轮

审稿人 1 意见：该研究结合掩蔽启动范式和 ERP 技术，探讨词素加工过程中的时间进程，结果发现，词素语义信息在早期的 ERP 成分 N250 上就出现了效应，基于此研究者提出词素语义加工可以很早，在 N250 左右时窗就获得激活。该研究具有一定的理论意义，但是可能还存在着一些问题，具体如下：

意见 1：作者在文章中反复强调，在之前的研究中，研究者发现了 N250 反映的是早期的知觉层面的加工而不是语义加工，而作者在这个研究中通过实验设计发现，N250 反映的是语义而不是知觉层面的加工。作者可能需要在前言和讨论中增加更多地文献，支持 N250 可能不仅可能反映知觉加工，还有可能反映语义加工。否则仅仅以这个研究而推翻或者否定之前大量研究的证据，略显武断。另外，作者可能需要在讨论中解释，为什么大量的之前词素的 ERP 研究中，都无法看到 N250 的效应。

回应：多谢指正，非常非常感谢审稿人的宝贵意见。我们在行文中对 N250 的介绍和讨论确实有所欠缺。根据您的建议，我们在前言部分我们增加了关于 N250 与词素语义加工关系的相关文献（请参见 P1 和 P2 红色字体部分），在讨论部分我们进一步解释了 N250 效应（请参见 P15 第二段全部内容）。

意见 2：作者的实验方法部分存在一些需要澄清的地方，包括：

意见 2.1：实验材料的控制上：作者希望选择的词素，其包含的两个义项的频率是匹配的。然而这些词素是否都只具有两个义项？此外，研究者提到两个义项的平均频率差为 0.15，统计是否显著？再者研究者对两个义项之间的语义关系进行了评定，结果是否显著？另外，之前的研究发现，词素的通透性是会对词素加工产生影响。而这个研究中，作者编制的启动词以及目标词，其各自的通透性如何？作者是如何控制这个变量的？

问题一：然而这些词素是否都只具有两个义项？

回应：确实，材料中所用词素并非只具有两个义项，而是根据条件一步步进行筛选得来的。我们之所以没有选择只具有两个义项的词素，是因为符合条件的词素太少，汉语文字博大精深，一字往往具有多个义项，只具有两个义项的词素少之又少。选材的第一阶段，我们从最新版《新华字典》中选择出具有两个或两个以上义项的词素 366 个，将这些词素选入《常用语义调查问卷》中并要求被试报告看到该词素第一反应出来的含义，统计被试报告结果发现，同一个词素被报告出很多义项（大于或等于两个义项），随后我们选取两个义项报告率在 20%到 65%之间，且报告率之差在 30%以下的词素为目标词素，进而保证选取的义项对被试来说不生僻且使用频率相当。在文中已经添加相应内容的介绍，请参见 P4 实验设计和材料部分第一段红色字体内容。

问题二：此外，研究者提到两个义项的平均频率差为 0.15，统计是否显著？

回应: 还是与汉语词素本身一字多义的属性有关, 很少存在词素仅仅具备两个义项且报告率均为 50% 的绝对平衡情况。事实上, 就汉语词素本身而言, 一种情况是, 词素具有多个义项并且其中一个义项的报告率非常高 (如 75% 以上), 而其他义项报告率相对较低, 也就是说一个词素的众多义项中存在着一个报告率很高的主要含义。相对于这种情况, 另一个种情况是, 存在两个或两个以上含义报告频率相当的词素, 我们选取其中两个含义进行实验, 虽然也存在一定的频率差, 如最终为 0.14 的差异, 但这是汉语词素本身的自然情况。此外, 相对于语义相对平衡的情况, 以往研究对于歧义词素语义不平衡的情况已经存在一些探讨 (e.g., Tsang & Chen, 2010; 2013; Tsang Wong, Huang, & Chen, 2014)。在这里我们更关心语义频率之差相对较小时, 歧义词素众多语义相对平衡时, 歧义词素语义的激活情况, 是否更依赖于词素所处的复合词语境。所以这里的平衡是相对的、不是绝对的。

因此, 虽然我们可以将两个义项的平均频率进行配对样本 t 检验, 在统计上也是显著的 ($t(119) = 11.93, p < 0.01$), 但其实这种统计的意义并不是很大。一是因为这是汉语词素的自然情况。二是因为我们这里有 120 个词素, 远远大于大样本容量 30, 因而对 120 个词素进行配对样本 t 检验, 很容易达到显著, 但这并不能说明什么问题。事实上, 在实验中我们统一把两含义中报告频率相对较高的定义为含义 1, 报告频率相对较低的定义为含义 2, 所以含义 1 和 2 之间的差异一定程度上可以说是我们设置的结果。

但是, 确实是在行文中我们没有认真地关注此问题, 多谢指正, 文章中实验设计和材料部分我们修订了之前关于词素语义均衡的绝对说法, 请参见 P4 实验设计和材料部分第一段红色字体内容。

问题三: 再者研究者对两个义项之间的语义关系进行了评定, 结果是否显著?

回应: 因为有时汉字虽然有很多含义, 但是一些含义为另一些含义的引申, 所以为保证每个歧义词素的两个义项确实为不同含义, 有一定区分度, 我们进行两者语义区分度 6 点量表 (1= 很不相关, 6= 很相关), 要求被试评价两者语义区分度。因为问卷统计的是两个义项的语义区分度, 直接得到两个义项语义区分程度, 因此这里无需对两个义项做配对样本 t 检验。

问题四: 另外, 之前的研究发现, 词素的通透性是会对词素加工产生影响。而这个研究中, 作者编制的启动词以及目标词, 其各自的通透性如何? 作者是如何控制这个变量的?

回应: 多谢指正, 对此, 我们补充了《汉语词汇语义透明度评定问卷》, 对实验中涉及到的所有启动词和目标词进行 6 点量表评定, 对目标词素是否可以推断整词含义进行评价 (1- 非常透明, 6- 非常模糊), 每个词语由未参与正式实验的 15 人评定, 结果发现, 对启动词三种类型, 即含义 1、含义 2 和无关词做方差分析, 未达到显著水平 ($F(2, 238) = 1.425, p > 0.2$), 对目标词两种类型即含义 1 和含义 2 做 t 检验也未达到显著水平 ($t(119) = 1.334, p > 0.1$), 说明所选材料在语义通透性上是相互匹配的。

这部分内容已经添加进文章中 (参见 P5 第一段红色字体部分)。其中启动词和目标词各情况下的均值和标准差已添加到材料部分 (参见 P5 表 1 红色部分)。

意见 2.2: 实验程序: CRT 的刷新率多少? 实验之后有没有确认被试是否可以发现启动词? 因为在掩蔽启动实验中, 这个问题至关重要。作者在这个研究中, 重复了一次启动刺激, 在 ERP 研究中这并不是一个好的做法。能否将重复的两组数据分别进行统计分析, 看看重复是否会对数据 pattern 造成影响。

问题一: CRT 的刷新率多少? 实验之后有没有确认被试是否可以发现启动词? 因为在掩蔽启动实验中, 这个问题至关重要。

回应：多谢指正，非常感谢审稿人的宝贵意见。屏幕刷新率是 75Hz，即每一屏呈现 13.33 毫秒，已经添加进文章中（见 P6 第一段红色字体部分）。对于被试是否可以发现启动词，实验中我们是这样处理的，实验结束后会展开一个调查，询问被试是否发现在目标刺激前有什么异常，大部分被试均都发现快速闪现一个类似汉字的刺激，但请他们报告具体内容，一些同学可能会报告双字词中个别汉字内容，但是整个双字词具体内容是很难报告出来的。文中添加相应内容，请参加 P6 第二段红色字体部分。

但是，我们承认这里我们的处理确实有所欠缺，以往研究采用掩蔽启动范式时，经常采用语义分类任务(Semantic category task)，要求被试报告非目标刺激的动物名称。其中，会在启动刺激部分设置一定的动物名称，最终观察动物名称处于启动位置时被试报告的频率，以此作为具体数据说明启动刺激是否可见这一问题(e.g., Grainger, Kiyonaga, & Holcomb, 2006; Holcomb & Grainger, 2007)。我们最初确实考虑以同样方式处理启动刺激可见性问题，但是考虑到语义分类任务对词素语义效应可能存在放大作用，另外已有的关于汉语词素加工的行为研究都是采用真假词判断任务，所以为了和同类研究进行比较，以及获得真实的词素语义效应，本研究未采用如此设计。但是本实验流程严格遵守以往已经发表的文章实验流程（e.g., Wong, Wu, & Chen, 2014），包括对启动刺激可见性的调查，希望可以为本研究提供一定支撑。

问题二：作者在这个研究中，重复了一次启动刺激，在 ERP 研究中这并不是一个好的做法。能否将重复的两组数据分别进行统计分析，看看重复是否会对数据 pattern 造成影响。

回应：首先，这里需要澄清的一个问题是，所谓的启动刺激重复一次，是指对于真词和假词来说，我们采用了相同的启动刺激。在真词内部启动刺激并不存在重复。这里我们在文中未表述不清，已经更正表述，见 P5 第二段红色字体部分。

但之所以真词和假词采用相同的启动刺激，是延续之前研究的做法，为了保证在实验中假词与启动刺激的关系和真词与启动刺激的关系一致（e.g., Wong, Wu, & Chen, 2014），避免对被试造成真假词区分的类别效应。但是，我们承认这样的设置确实会对结果造成一定的影响。因此我们以启动刺激重复性为变量（第一次 vs. 第二次），对实验数据进行重新编码，因为我们只关心目标刺激是真词时，各变量的效应，因此，对真词的反应时和正确率分别进行 2（启动刺激重复性）*3（启动词类型）*2（目标词类型）重复测量方差分析，检验启动刺激重复性是否会引发不同的 pattern。反应时结果表明启动刺激重复性主效应并不显著($F(1, 18)=0.032, p > 0.8$)，启动刺激重复性与启动词类型或目标词类型两者或三者的交互作用均不显著($p > 0.3$)，正确率结果类似，启动刺激重复性主效应不显著($F(1, 18)=3.756, p > 0.05$)，启动刺激重复性与启动词类型或目标词类型两者或三者的交互作用均不显著($p > 0.1$)，由此可见重复启动对行为数据模式并未造成影响。但是，值得注意的是，我们这里仅分析了行为数据，而未分析脑电数据，是因为如果进一步区分启动刺激的重复性，那么每个条件的试次就剩 20 来个，如此少的试次在脑电分析中很难得出有意义的结果。但是值得庆幸的是，我们的行为结果没有反映出启动刺激重复性的任何效应。此外，我们计算了在每个条件下启动刺激的重复次数，发现每个条件下启动刺激第二次出现的频次都是一样的（20 次），这就保证了条件间的平衡，也可以间接地控制了启动刺激重复性在本实验中的干扰作用。

意见 2.3：数据结果：总体来说，作者发现词素语义在 N250 上出现效应，主要是通过同形同义与无关条件的对比发现的。而从作者报告的结果看，同形同义和同形异义对比，其实 N250 的结果并不是非常显著，甚至在目标词词素取含义 2 时，同形同义和同形异义之间在 N250 差异消失了。进一步地，作者并没有报告同形异义和无关条件之间的差异。如果按照作者前言的逻辑，N250 是可以反映知觉层面的加工，那么我们是否可以预测同形异义与无

关之间应该也在 N250 上有差异，而同形同义由于跟目标词之间既有形也有义的联系，因此效应会更大一些。作者可能一方面需要提供这方面的数据，另一方面可能需要对这个结果进行讨论，可能这样会对作者的结论提供更充分的证据。

回应：感谢审稿人耐心阅读以及所提出的宝贵意见，如您所述，在大时间窗口分析时，N250 的效应确实主要体现同形同义与无关条件的对比获得，如果以此为根据论述词素语义和 N250 的关系确实有所欠缺。但是，值得注意的是，在小时间窗口分析时，确实发现了同形同义与同形异议条件（150~200 ms 和 200~250 ms），以及同形异义和无关条件的差异（200~250 ms），无论含义 1 还是含义 2 为目标词，这一结果请参见 P10 第二段和第三段红色字体部分。这种线性变化的结果无疑是加强了我們之前的论断，词素语义可以调节 N250 的变化，但是词素语义是单独起作用还是和词素词形交互产生作用，这里我们还无法确定。另外，结合您的意见和建议，在讨论中我们修订和增加了对 N250 性质的讨论，请参见 P14 红色字体部分，和 P15 第二段全部内容。

审稿人 2 意见：论文选题有一定的理论意义，研究方法恰当。作者对 N250 的性质讨论得不够充分。

回应：非常感谢审稿人耐心阅读和所提出的宝贵建议，文中确实对 N250 的性质讨论不够充分。针对您的问题，我们分别在前言中增加了一些关于 N250 意义的介绍，增加了相关文献的介绍（请参见 P1 和 P2 红色字体部分），在讨论中修订并添加增加了一些关于 N250 性质的内容（请参见 P15 第二段全部内容）。

第二轮

审稿人意见：作者修改后的稿件，已经基本上回答了我之前提出的问题。因此同意接收。

回应：非常感谢审稿人的耐心阅读和所提出的宝贵建议！

第三轮 编委复审意见

编委复审意见：该研究结合掩蔽启动范式和 ERP 技术，探讨词素加工过程中的时间进程，结果发现，词素语义信息在早期的 ERP 成分 N250 上就出现了效应。在歧义词加工中，当歧义词素两含义使用频率相当时，词素语义会在早期加工阶段就发挥作用，反映在 N250 的变化上。该研究设计严谨，研究难度大。经过两轮审稿，审稿人提出了详细而有建设性的意见，作者也进行了详尽的答复，并将相关补充内容加入文章中，两审稿人均同意可以接收该文。

因为经过多轮修改，文章显得冗长，请作者将文章主体部分压缩到 8000 字左右。压缩后可以发表。

回应：文章中删除部分或更改部分说明

一、更改前后字数统计（算标点符号和空格，不算图和表）

	引言	方法	结果分析	讨论	结论	文字总数
更改前	3259	2514	2305	3107	103	11288
更改后	2487	1792	2186	2337	101	8903

二、具体体整内容

文章从头到尾都进行了语言上的简化,我们力图在不改变文章含义和保持阅读流畅性的基础上,用更精简的文字传达给读者,其中有几种较大的删除(3行或3行以上)在此单独说明:

1.引言部分体整内容

正文 P2 第二段“在目标词出现 250 ms 内(N250 的典型时间窗口内)也没有发现透明度效应, N400 反映整体词汇语义整合过程,典型的时间窗口为 300~500 ms;因此研究者认为,在词汇加工的早期阶段,只要拥有共享的词素词形,就足够引起词素启动效应,词素语义在词汇加工早期阶段不起作用,同时这样的结果也证明了词素语义确实能够引发 N250 的变化”。

P3 第二段“正如透明度范式下的研究,来自脑电的研究结果,即使采用掩蔽启动范式,多支持词素语义调节 N400 的变化,词素语义的作用晚于词形(e.g., Lavric et al., 2012)。由此可见行为的研究结果是不能简单地推论到 ERP 研究上”。

和 P3 第三段“即使在掩蔽启动范式下,反应时结果也不可能揭示各阶段的加工过程。而 ERP 结果不同,它可以分阶段地探讨时程问题,精细地监控各加工过程。事实上,即使掩蔽启动范式下的行为结果认为词素语义可以在早期发挥作用,也可能仅是作用在脑电结果的 N400 这一成分上。正如透明度范式下的研究,来自脑电的研究结果,即使采用掩蔽启动范式,多支持词素语义调节 N400 的变化,词素语义的作用晚于词形(e.g., Lavric et al., 2012)。由此可见行为的研究结果是不能简单地推论到 ERP 研究上,因此很有必要采用 ERP 技术探讨歧义词素的加工,从加工时程的角度,更精细地揭示歧义词素的各阶段加工过程,以及歧义词素语义的激活情况”。属于精简内容,或再前面已经提及或在后文讨论中涉及到,因此删除。

2.方法部分体整内容

正文 P3 最后一段,实验设计和材料中“但是需要注意的是,这里的歧义词素两个义项的使用频率是相对平衡,并不是理想化的绝对平衡。这与汉语词素本身一字多义的属性有关,很少存在词素仅仅具备两个义项且报告率均为 50%的绝对平衡情况。事实上,就汉语词素本身而言,一种情况是,词素具有多个义项并且其中一个义项的报告率非常高(如 75%以上),而其他义项报告率相对较低,也就是说一个词素的众多义项中存在着一个报告率很高的主要含义。相对于这种情况,另一个种情况是,存在两个或两个以上含义报告频率相当的词素,我们选取其中两个含义进行实验,虽然也存在一定的频率差,但这是汉语词素本身的自然情况”精简为“但注意这里的歧义词素两含义使用频率是相对平衡,存在一定的频率差,这与汉语词素本身一字多义属性有关,符合词素本身的自然情况”

正文 P5 第二段,实验程序中将“此外,为了确定启动刺激是否可见,实验结束后会展开一个调查,询问被试是否发现在目标刺激前有什么异常,被试均发现在目标刺激前快速闪现一个类似汉字的刺激,但请他们报告具体内容,一些同学可能会报告双字词中个别汉字内容,但是整个双字词具体内容是很难报告出来的,说明被试虽然意识到启动刺激的存在,但很难获得具体内容。”修改为“实验结束后要求被试回答在目标刺激前是否发现异常结果被试均发现异常,但大多报告不出具体的内容”

3.讨论部分体整内容

正文 P13 第一段,删除“在词汇识别早期加工阶段,词素语义是否可以和词素词形一样发挥重要的作用,这个问题一直颇具争议。一些研究发现词素语义可以在早期发挥作用,而另一些研究却否认了这样的结果。采用歧义词加工的几篇研究结果虽然比较一致,但都是来自行为的研究,缺少时程方面的证据,另外,同时也忽略了歧义词素各含义使用频率问题”,

因为这与前文引言部分重复，因此可删除。

P14 第二段，删除“事实上，这个结果也符合 reordered-access 模型(e.g., Duffy, Morris, & Rayner, 1988)关于歧义词加工的假设，即在歧义词词汇识别早期阶段，歧义词素各个含义都被有效激活，在后一阶段只有符合情境的信息才被留存在下来”，因为 reordered-access 模型在前后文均未涉及到，用该模型直接解释关于歧义词词汇识别略显突然。

P15 第三段，删除“从上面的讨论，我们可以看出，词素语义对词汇识别的作用是分阶段完成的。第一阶段是通过 N250 实现的，在歧义词两含义使用频率相当时，共享词素同形同义时，才更容易引发 N250 的变化，引发目标词共享词素语义的激活，接下来激活后的词素语义就顺利地整合到整个词汇的语义识别中。相反，前期词素语义没有得到很好的激活(共享词素同形条件)或者没有被激活(无关条件)，接下来的整词语义整合就变得困难，N400 就越大”，这部分内容在引言和讨论中都涉及到过，属于重复内容，因此可以删除。

4.参考文献休整内容

删除了“Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading ☆. *Journal of Memory & Language*, 27(4), 429–446.”，因为讨论中已经删除了 reordered-access 模型解释的部分，因此删除该文献。

第四轮 主编终审意见

主编终审意见：该研究旨在探讨中文多词素词汇中词素语义的加工时程问题，主要关注的是词素语义是否可以在词汇加工早期阶段发挥作用。作者首先论述了已有研究通常采用的语义透明度范式可能存在的问题，又指出中文词素一般存在多个不同含义的特点。据此，作者结合掩蔽启动范式和 ERP 技术，以歧义词素做为实验材料，结果证明词素语义在中文多词素词加工早期阶段就可以发挥作用。总体来看，该研究目的明确，实验设计合理，变量控制较为严格，得出了一定有意义的结论，尤其是进一步肯定了词素加工与 N250 的关系。并且参考审稿人意见，作者已经认真、系统地修改了原稿件的内容，基本达到了学报发表要求。但是，文章在行文上还存在一些问题，可能需要作者进一步优化调整：1. “引言”部分段落结构有些零散，可以适当调整，将论述同一问题的内容整合到同一段落中。2. 文字中的一些细节性格式问题需要进一步规范，以保持前后文一致，如括号的全角/半角形式，字母间、括号间的空格问题等。

回应：非常感谢主编的宝贵意见，已将原文引言部分的第二、三、四段合并，第五、六、七段合并，最终引言部分为 5 个段落。另外，已多次精读文章，文字中的细节性格式问题已重新校对，括号已都统一成半角形式，字母间和括号间的空格问题已改正，感谢主编意见！