

《心理学报》审稿意见与作者回应

题目：中文阅读预视加工中的词频和预测性效应及其对词切分的启示：基于眼动的证据
作者：苏衡, 刘志方, 曹立人

第一轮

审稿人 1 意见：

论文存在如下一些问题，供作者思考：

意见 1：实验一到实验三发现，各种掩蔽条件下均影响词频效应，作者推论，中文读者词汇加工存在汉字加工过程，实际上，该实验结果也可能说明，预视加工不能够达到词汇水平（至少不能达到出现词频效应的程度），以此推测“存在汉字加工”阶段，对本研究的数据而言，有点过度推论。

回应：

谢谢审稿专家的意见。李兴珊等人(2009,2011)提出的词切分模型是本研究所依据的理论背景之一，该词切分模型认为词汇识别的前期必然经过汉字识别阶段，还有证据表明汉字在中文词识别中有相对独立的作用(申薇，李兴珊，2012)。当前的研究结果应该说是与李兴珊等人的观点并不冲突，但是考虑到本研究中并未操控和影响汉字加工过程，因此专家认为推测“存在汉字加工阶段”是过度推论，是有道理的。

根据现有多个文献均表明尽管读者基于词汇理解中文文本，但中文阅读中确实存在“汉字加工”过程(李兴珊等人(2009,2011, 2012; 李兴珊, 刘萍萍, 马国杰, 2011 申薇, 李兴珊, 2012; Yan et al., 2006)，因此在“总讨论”部分，我们根据本研究的数据结果结合汉字加工过程进行了讨论。不过，为提高文章的严谨性，我们在将本版修改稿中的“结论”进行了严格的限制，将“存在汉字加工”的描述改变说法“不能达到词汇水平”。整个“结论”部分被修改成“中文读者对词 $n+1$ 的预视加工达不到词汇水平，读者较难通过字词识别切分词 $n+1$ ；中文读者根据词汇预期加工切分词 $n+1$ ”。

希望这个修改能够增加文章严谨，请专家审查。

意见 2：作者在引言开头部分，直接引入“E-Z 读者模型”和 SWIFT 模型将词汇加工区分为两个阶段，这两个阶段具体指什么内容，他们与本研究之间的是什么关系，作者应该有明确的描述。

回应：

谢谢专家的意见。E-Z 读者模型”和 SWIFT 模型对词汇加工的两阶段划分是本研究的重要的理论依据，稿件确实的这两个阶段的描述不够具体，在将“此理论问题”过渡至本研究内容方面不够自然、流畅。本版修改稿中的第一段、第二段我们修改并丰富了相关内容。具体做法是，在第一段的末尾增加如下内容：

“熟悉性检验仅涉及加工水平较低时词汇正字法模版识别；词汇通达提取词汇更加抽象的语音和语义表征的阶段(Reichle, Pollatsek, & Rayner, 2006; Schotter, Reichle, & Rayner, 2014)。SWIFT 模型采用“词汇激活(lexical activity)”的概念来描述实时的词汇加工状态，它认为，阅读知觉广度内的所有词汇组成一个词汇激活性质的集合，随着时间的推移这个集合的值实时地发生改变，词汇激活是持续时间的函数，词汇前加工阶段(lexical preprocessing)中，该词激活度将随着时间的推移逐渐达到最大值，词汇通达阶段(lexical completion)其激活度开始下降，激活度接近于 0 时，该词被成功识别(Engbert, Longtin, & Kliegl, 2002; Engbert, Nuthmann, Richter, & Kliegl, 2005; Kliegl, Nuthmann, & Engbert, 2006; Richter, & Kliegl, 2006; Kliegl, 2007; Kliegl, Risse, & Laubrock, 2007; Schad & Engbert, 2012)。”

在第 2 段的末尾增加如下内容：

“实验证据表明，拼音文字阅读中，目标词汇的词频和预测性影响对之的注视指标(首次注视时间、凝视时间和总注视时间)和眼跳指标(跳读概率和再注视概率)，说明词汇加工确实实时影响拼音文字阅读眼动控制过程(Rayner & Duffy, 1986; Rayner & Well, 1996; Kliegl et al., 2004; Rayner, Binder, Ashby, & Pollatsek, 2001);然而在细化词频和预测性如何影响词汇加工两阶段问题上，E-Z 读者模型提出的假设性解释认为，词频和预测性对词汇前期加工阶段影响程度低于其在词汇加工后期阶段中的影响(Reichle et al., 2006; Schotter et al., 2014); SWITF 模型则强调词频和预测性以不同方式影响最大词汇激活度(Richter, & Kliegl, 2006)。”

E-Z 读者模型和 SWIFT 模型都将词汇加工看作是句子理解、文字识别的基本单元和影响眼动过程的基本因素，因而词汇加工难度效应是构建眼动控制理论模型的重要参考线索。词汇的词频和预测性是影响词汇加工的两个基本因素，两个模型一致认定这两个基本因素都影响词汇加工的两个阶段。不过这个观点存在两个不足：(1)词频和预测性对词汇加工两阶段的影响还停留在假设层面，目前尚没有相关的实证研究；(2)词频和预测性对词汇加工两阶段的影响在中文阅读中与它们在西文阅读中的作用是否相同。解释这个问题对探讨中文词切分问题有重要参考意义；还有利于深化对词汇加工过程的理解。我们的文章是基于上述考

考虑逐步引出本次的“研究问题、假设”和实验，我们在本次修改稿的“前言”中从第四段开始，基于上述基本思路，铺垫展开论文。请专家审查。

意见 3: 作者在引言倒数最后一段，提出，“探讨中文阅读中词加工阶段还应考虑早期的汉字加工阶段”，然而，本研究中实际上并未操作与汉字加工直接相关的变量。请作者予以解释。

回应:

谢谢审稿专家的意见。本研究确实没有“操作与汉字加工直接相关的变量”，因此引言倒数最后一段提出，“探讨中文阅读中词加工阶段还应考虑早期的汉字加工阶段”是不合适的，上述描述显然不利于清晰阐述研究问题，因此在本次修改稿中，我们完全接受审稿专家的意见，直接删除了“探讨中文阅读中词加工阶段还应考虑早期的汉字加工阶段”的描述，另外还对文章的最后两段进行了大幅修改，优化了论文结构。

与此同时，考虑到大量文献证据支持“中文阅读中存在汉字加工过程”的观点，我们在修改的讨论部分，根据本研究结果并结合“汉字加工”进行了讨论，这个问题与对“问题 1”的回复相同。请专家审查。

意见 4: 词切分目前是热点问题，但是，作者对 2015 年的文献引用不全，作者应该予以补充。

回应: 谢谢专家的意见。在本文成稿时 2015 年的部分文献还不能被搜索到，受到修改意见时新搜索到的与本研究相关的两篇 2015 年文献均是李兴珊课题组的研究成果，这些成果主要包含两项内容:(1)考察眼跳的目标选择问题，比如 Ma 等人(2015)的研究显示，汉字特性影响眼跳落点，意味着中文读者在一定程度上基于汉字选择眼跳目标, Liu 等人(2015)的研究则发现，副中央凹的预览加工能够引导眼跳落点；(2)考察字词加工问题，比如 Li 等人(2015)的研究显示，词汇特性对眼动过程的影响程度甚于汉字特性的影响作用，尽管中文阅读的文字加工有其特殊性，但还是主要符合普遍性语言规律。考虑到第一项研究内容与本研究的主旨的关系差距较大，在本版的修改稿中，我们引用了 Li 等人(2015)的研究结果(引言第四段末尾)。与此同时，Liu 等人(2015)的研究结论能够解释实验二的部分结果，因此在修改稿的“3.3 讨论”我们基于这个参考文献对实验二的结果进行了讨论。

请专家审查。

意见 5: 论文多出出现“我们”第一人称的写法, 请修改。

回应: 谢谢专家的意见, 本次修改稿对这个问题进行了修改, 请审查。

意见 6: 表 5 中, 平均眼跳幅度, 请用汉字的个数作为单位。

回应: 谢谢专家的意见, 平均眼跳幅度是基于整体分析的眼动指标。考虑到审稿专家二质疑“基于句子整体数据分析的必要性”, 认真考虑后发现这个质疑很有道理, 以此本次修改稿中完全删除了整体分析结果。请专家审查。

审稿人 2 意见: 该研究采用眼动方法, 通过 4 个实验考察了中文读者在预视中所达到词汇加工阶段, 并以此探讨词切分问题。研究手段得当, 实验设计合理, 研究内容具有重要的理论意义。以下问题请作者思考或修改:

意见 1: 请将“文摘”改成“摘要”, 并调整摘要的内容, 明确说明研究采用的研究方法, 同时概括 4 个实验的研究结果, 给出主要的研究发现。

回应: 谢谢专家的意见。我们完全接受专家的修改建议将摘要修改成如下内容:

“研究采用眼动随动显示技术考察中文阅读预视加工中的词汇加工问题。前三项实验发现, 剥夺预视加工的掩蔽条件、正确提示词 n+1 右侧边界的掩蔽条件以及不能提示词 n+1 右侧边界的掩蔽条件都不影响词频效应, 说明中文读者对词 n+1 处文字的预视加工达不到词汇水平。由此可见, 中文读者较难通过自下而上的文字识别切分词 n+1。实验四考察剥夺预视加工的掩蔽条件、提示词 n+1 右侧边界的掩蔽条件对预测性效应的影响, 结果发现, 剥夺预视加工完全消除预测性效应, 提示词 n+1 右侧边界则减少预测性效应。说明对词汇的预期加工是中文读者切分词 n+1 的参考线索。综合四项实验结果可知, 词频和预测性以不同的形式影响中文词汇加工过程; 中文阅读中的文字加工阶段有其独特性。”请专家审查。

意见 2: 请在引言第一自然段的倒数第三、四句中, 分别在 SWIFT 模型和 E-Z 读者模型后加注文献。

回应: 谢谢专家的意见。根据专家的建议在本版修改稿中, 我们已经针对 SWIFT 模型和 E-Z 读者模型增注了文献, 与此同时审稿专家一的第 2 条意见, 我们对引言的第一自然段的末尾进行了修改, 请专家审查。

意见 3: 在引言最后一个自然段中，作者仅提到了实验 4，对其他几个实验没做任何说明，请加以补充。或者通过其他方式阐明 4 个实验主要关注的研究问题。

回应: 谢谢专家的意见。在引言的最后一段我们针对四个实验所依据的理论背景和假设进行的描述，具体内容如下：

“词频效应和预测性效应是词汇加工过程中的两个常见现象，它们反映两个不同性质的、独立的词汇加工过程，E-Z 读者模型和 SWIFT 模型都提出两个相对独立的函数对之进行描述(Rayner et al., 2004; Engbert et al., 2005; Schad & Engbert, 2012)。据此，本研究分别从词频效应和预测性效应两个角度考察预视切分词 $n+1$ 的加工形式问题。如果读者通过识别字词切分词 $n+1$ ，那意味着在预视中能够达到或者至少部分达到早期词汇加工阶段，基于这个假设，剥夺词汇预视会将预视加工所致的词频效应“推延”至注视加工中，因此掩蔽词 n 右侧词汇将导致词频效应增大，与此同时提示词 $n+1$ 右侧边界的会促进词汇加工，进而将减少词频效应。本研究的前三项实验检验上述假设，若没有产生符合预期的呈现条件变量与词频变量间的交互作用，则意味着读者不可能基于自下而上的字词加工执行切分词 $n+1$ 任务。基于同样逻辑，若中文读者通过自上而下的预期确定词 $n+1$ 的右侧边界，掩蔽词 n 右侧词汇(剥夺预视)将导致预测性效应增大，同时提示词 $n+1$ 右侧边界的掩蔽将减少预测性效应。实验四包含上述两种掩蔽条件，以此检验“读者通过预期切分词 $n+1$ ”的可能性。”

过上述描述引出研究的四个实验，请审查。

意见 4: 请作者说明进行“基于句子整体数据分析”的必要性。

回应: 谢谢专家的意见。“基于句子整体数据分析”是针对句子阅读时所有眼动事件(注视和眼跳)的分析(取平均值或者总和值)，由于研究四个实验所采用的掩蔽范式是针对句子中所有字词的(并非针对特定的目标字词)操作，从这个角度出发应该给出整体分析的数据。然而专家质疑整体分析的必要性也很有道理。综合思考后觉得给出整体分析结果的价值不大，理由如下：

1 根据本研究目的，我们需要通过参考目标词汇兴趣区域上的眼动数据，确定研究结论，整体分析数据结果与研究目标关系不大；2 在作者 2013 年发表的讨论“切分词 $n+1$ 过程是否存在”的文章中所采用的实验范式与本文完全相同(本篇小说探讨读者基于何种性质的词汇加工切分词 $n+1$)，该论文已经完整给出整体分析眼动数据，本论文再次报告整体分析没有

太大价值(详细情况见文献:刘志方, 闫国利, 张智君, 潘运, 杨桂芳.(2013). 中文阅读中的预视效应与词切分. *心理学报*, 45, 614-625)。

基于上述考虑, 给出整体分析结果不仅不能说明问题, 还会导致文章过长影响简洁性, 给读者理解带来干扰, 因此在本版的修改稿中我们剔除了所有实验的整体分析结果。请专家审查。

意见 5: 请作者检查 4 个实验的分析结果, 对于交互作用显著的效应, 补充简单效应分析, 并说明交互作用的意义。

回应: 谢谢专家的意见。前三项考察掩蔽范式对词频效应影响的实验的结果中, 只有实验二的跳读概率指标呈现方式与词频交互作用效应显著, “简单效应分析显示控制条件中词频效应显著($p < 0.05$), 词 n+1 右侧掩蔽条件中的词频效应不显著($p > 0.05$)。”这个交互作用显著的意义是“提示词 n+1 右侧边界导致跳读概率指标上词频效应消失”。鉴于, 没有证据表明中文读者基于词 n+1 执行相邻词间眼跳(Li, Liu & Rayner, 2011; Ma, Li, & Pollatsek, 2015; 李兴珊, 刘萍萍, 马国杰, 2011), 但在控制条件下高频词汇被跳读的概率高于低频词汇确实意味着在预视加工中高频词得到较高程度的预视加工, 提示词 n+1 右侧边界导致跳读概率减少(实验范式也导致跳读概率指标的词频消失), 这意味着自下而上地预视加工词 n+1 上的汉字确实可能会导致切分词 n+1。

然而, 从本次研究的所有实验上看, 控制条件中跳读概率的词频效应并不稳定(实验一和实验二的控制条件中跳读概率词频效应不显著), 掩蔽范式对词频效应的影响也不稳定(实验一和实验二的控制条件掩蔽条件在跳读概率指标上交互作用都不显著)。多项研究证据表明(包括本研究), 切分词 n+1 与某种性质的词汇加工有关(Yan, et al., 2010,2015; Shu et al., 2011; 张智君等人, 2012; 刘志方等人, 2013), 但信息加工在特定时刻只能达到特定的“加工程度”, 由此可以确定, 切分词 n+1 的成功与否不是绝对的“或有或无”的问题, 它是一个概率性问题。因此根据掩蔽范式对词频效应影响的数据结果, 合理的推论是: “中文读者较难通过自下而上的文字加工方式完成切分词 n+1 过程”。

实验四操控呈现条件和目标词汇的预测性, 结果在凝视时间和再注视概率两个指标上发现呈现条件与预测性的交互作用显著, 文中相应结果数据呈现部分详细分析的它们的简单效应。在本版修改稿的“5.3 讨论”中结合实验假设给出这种交互作用的意义, 详细描述如下:

“结果发现, 呈现条件和预测性交互影响凝视时间和再注视概率, 简单效应分析后发现, 相对于控制条件, 剥夺预视导致凝视时间和再注视概率的预测性效应完全消失; 提示词 n+1

右侧边界(词 n+1 右侧掩蔽条件)则导致预测性效应在凝视时间上减少,在再注视概率指标上消失。由此可见,提示词 n+1 右侧边界掩蔽条件对预测性效应的影响符合并支持预先假设。不过,剥夺预视的掩蔽条件对预测性效应的影响则与预先假设不符,但考虑到预测性效应反映主观预期对词汇加工过程的影响,是种较为高级的认知过程,根据交互激活理论(interactive activation models)的观点,预测性能够影响早期的低级层面的词汇加工过程(Morton, 1969),基于此若设定预测性只能影响正常阅读中的早期的词汇预视加工阶段,剥夺预视意味着势必减弱预期对词汇加工过程的影响,可见这个结果也是支持“中文读者基于预期切分词 n+1”推测的证据。”

审稿人 3 意见: 本论文试图通过考察不同预视条件下汉语读者的词切分问题,该论文对揭示汉语句子阅读中的词切分和相关模型构建具有一定的重要意义。但是该论文的问题提出、研究结果和结论等存在诸多问题,还需要修改完善。下面有一些不明白的地方,向作者请教。

意见 1: 该文提出的两个模型及其参考文献——E-Z Reader 和 SWIFT 模型,是“用于解释拼音文字阅读中的文字加工过程”的吗?这类模型还是为了解决眼动模式?。

回应: 谢谢审稿专家意见。E-Z Reader 和 SWIFT 模型是两个描述“拼音文字阅读眼动控制过程”的模型,这两个模型都包含两个相同的主要模块:“文字识别加工模块”和“眼动控制过程模块”。阅读理解必须要求这两个模块相互协作才能完成,因此“文字识别加工模块的内容”和“眼动控制过程模块的内容”都是上述模型必须考虑、解释的。这两个模型既是解释拼音文字阅读中文字加工过程的参考模型,也是解释阅读中眼动控制过程的参考模型。显然,对这两个模块的特点进行深入的考察和研究,是发展和修正阅读眼动控制模型的必然途径。本研究的重点在于考察“中文阅读中的文字加工特点”。估计在原版稿件中“引言”的前部,我们同时提出“文字加工过程”和“眼动控制过程”两个概念,不利于读者清晰快速理解。在本版修改稿中“前言”的第一段,我们进行了修改,参考本文的研究目标只提及“文字加工过程”这个概念,希望能够打消专家的疑问,提高论文的可读性。请专家审查。

意见 2: 第 4 页,第 2 段——“据此可以假设,如果在预视加工中读者可以完成或者部分完成第一阶段的词汇加工,那么剥夺预视加工应该能够扩大词频效应。同样,如果句子背景对目标词汇的预测程度能够影响对之的预视加工过程,那么剥夺预视加工同样亦会改变预测性

效应。”请问为什么“剥夺预视加工应该能够扩大词频效应”呢？能否更进一步的阐述？假若是这样的原因“读者能够在预视中完成或者至少部分完成早期的词汇加工阶段，因此剥夺词汇预视加工对掩蔽条件将增大注视阶段的词频效应”，对于数据结果会有怎样的预测？

回应：谢谢专家的意见。我们认为，如果中文读者基于字词识别切分词 $n+1$ 的话，那么剥夺预视加工能够扩大词频效应。导向这个推论的原因是：词汇识别是阅读理解的必要前提，正常阅读中读者可以对词汇进行预视加工，如果在预视中，读者可以通过字词识别过程切分词 $n+1$ ，那就意味着字词的预视识别加工程度能够达到词汇水平，这样剥夺预视就意味着读者必须将对特定词汇的预视加工程度转移至注视加工当中，此时便会导致词频效应增大。这样会导致结果中出现“呈现条件和词频间的交互作用显著”，基于同样的逻辑若没有发现剥夺预视加工能够导致词频效应增大（呈现条件和词频间的交互作用不显著），那就意味着中文读者在对词 $n+1$ 的预视加工中不能达到词汇水平（因为词频效应是反映文字加工达到词汇水平的重要参考线索），在此基础上可以进一步认为，中文读者较难通过字词识别加工切分词 $n+1$ 。我们在本版修改稿文章“引言”部分的第 3 段提出这个假设，在“引言”最后一段根据具体的实验操作，细化阐述这个假设。为逐步导出本文章的研究问题、实验假设和实验操作思路，我们对原稿进行修改，其中对“引言”最后一段的修改应该能够回答审稿专家的疑问：

“词频效应和预测性效应是词汇加工过程中的两个常见现象，它们反映两个不同性质的、独立的词汇加工过程，E-Z 读者模型和 SWIFT 模型都提出两个相对独立的函数对之进行描述(Rayner et al., 2004; Engbert et al., 2005; Schad & Engbert, 2012)。据此，本研究分别从词频效应和预测性效应两个角度考察预视切分词 $n+1$ 的加工形式问题。如果读者通过识别字词切分词 $n+1$ ，那意味着在预视中能够达到或者至少部分达到早期词汇加工阶段，基于这个假设，剥夺词汇预视会将预视加工所致的词频效应“推延”至注视加工中，因此掩蔽词 n 右侧词汇将导致词频效应增大，与此同时提示词 $n+1$ 右侧边界的会促进词汇加工，进而将减少词频效应。本研究的前三项实验检验上述假设，若没有产生符合预期的呈现条件变量与词频变量间的交互作用，则意味着读者不可能基于自下而上的字词加工执行切分词 $n+1$ 任务。基于同样逻辑，若中文读者通过自上而下的预期确定词 $n+1$ 的右侧边界，掩蔽词 n 右侧词汇(剥夺预视)将导致预测性效应增大，同时提示词 $n+1$ 右侧边界的掩蔽将减少预测性效应。实验四包含上述两种掩蔽条件，以此检验“读者通过预期切分词 $n+1$ ”的可能性。”

意见 3：该论文能否回答这个问题“中文读者采用何种线索或者基于何种性质的加工在预视加工中确定词汇边界”？如果部分可以，是如何回答这个问题的？

回应：谢谢审稿专家的意见。由于中文特殊的书写方式，读者必须通过某种形式的词汇加工确定词汇边界，基于多项研究显示，中文句子阅读中的词切分发生在词汇的早期加工阶段(Yan, Kliegl, Shu, Pan, & Zhou, 2010; Shu, Zhou, Yan, & Kliegl, 2011; Yan, Zhou, Shu, & Kliegl, 2015; Gu & Li, Liversedge, 2015; Gu & Li, 2015; 张智君, 刘志方, 赵亚军, 季靖, 2012; 刘志方, 闫国利, 张智君, 潘运, 杨桂芳, 2013), 本研究考察中文读者基于何种性质的词汇加工切分词 $n+1$ 。基于成熟的眼动控制模型对词汇加工的看法, 本研究只操作了影响词汇加工的两个变量: 词频变量和预测性变量, 由于操作的变量有限, 因此研究是部分地回答上述问题, 理由如下:

词频效应反映自下而上的文字加工过程(Fodor, 1983; Forster, 1979), 而预测性效应则反映在句子理解基础上的主观预测对词汇加工的影响, 因而反映自上而下的文字加工过程(Dambacher et al., 2006)。若中文读者基于自下而上的字词识别在预视加工切分词 $n+1$, 意味着对词 $n+1$ 的预视加工能够达到词汇水平(鉴于 E-Z Reader 模型和 SWIFT 模型对词汇加工加工阶段的描述, 对词 $n+1$ 的预视加工所能达到的程度必定是最低达到或接近于达到词汇加工的第一阶段); 阅读理解意味着句子中的词汇被完全识别, 因而剥夺预视加工条件则意味着, 相对于正常条件, 词汇识别被完全推进至注视加工当中, 因而相对于控制条件, 剥夺预视加工应该能够扩大词频效应(应出现显著词频与呈现方式的交互作用)。与此同时, 提示词 $n+1$ 右侧边界应该逻辑上会加速词汇加工速度, 进而应该会减少词频效应(此时同样应出现显著词频与呈现方式的交互作用)。但实验一和实验二的结果完全不符合上述假设的预测, 实验三的结果还能排除实验二的结果是由实验范式本身所致。因此前三项实验的结果一致证明, 中文读者不太可能通过自下而上的词汇识别切分词 $n+1$ 。

自上而下加工是与自下而上加工相对应的另一种加工形式, 这两类认知加工过程广泛存在于人类认知过程当中。如果中文读者不能通过自下而上的词汇加工切分词 $n+1$ 的话, 通过自上而下加工方式(比如, 预期加工)切分词 $n+1$ 是可能的词切分形式(否则意味着, 中文读者不太可能通过预期加工切分词 $n+1$; 或者意味着中文读者不可能在预视中切分词 $n+1$)。实验四的基本假设是: 如果剥夺预视的掩蔽条件和提示词 $n+1$ 右侧边界的掩蔽条件能够改变预测性效应, 则说明读者能够在预视中通过预期(对词汇的预期加工)切分词 $n+1$ 。实验四的结果支持上述假设, 因此我们认为预期加工是中文读者切分词 $n+1$ 的加工方式之一。

意见4:第5页,第2段,作者提到“中文读者能够在预视加工中执行词切分过程(Li & Pollatsek, 2011; Gu & Li, 2015; Yan, Kliegl, Shu, Pan, & Zhou, 2010; Shu, Zhou, Yan, & Kliegl, 2011; Yan,

Zhou, Shu, & Kliegl, 2015; 张智君, 刘志方, 赵亚军, 季靖, 2012; 刘志方, 闫国利, 张智君, 潘运, 杨桂芳, 2013)。”，能否将证据展开描述些？其中引用的参考文献真的能够提供证据吗？比如——这篇文章 Li & Pollatsek (2011)。

回应：谢谢专家的意见。我们对上述引文中的词切分展开描述如下：

Yan 与其合作者通过考察词汇上注视点的分布模式推测，副中央凹预视中读者能够在一定程度上切分词 $n+1$ (Yan, Kliegl, Shu, Pan, & Zhou, 2010; Shu, Zhou, Yan, & Kliegl, 2011; Yan, Zhou, Shu, & Kliegl, 2015)。还有证据表明，提示词 $n+1$ 的右侧边界能够积极影响读者眼动模式(张智君, 刘志方, 赵亚军, 季靖, 2012; 刘志方, 闫国利, 张智君, 潘运, 杨桂芳, 2013)，词 $n+1$ 上汉字位置顺序颠倒也影响注视持续时间(Gu & Li, 2015)。由于参考文献 (Li & Pollatsek, 2011) 中论述的是词汇信息对汉字加工的影响，该文献的主旨与预视加工中的词切分问题有点距离，因此本处剔除了该参考文献。请专家审查。

意见 5：在问题提出的最后一段，是否需要简要介绍下四个实验的逻辑？另外，“本研究的实验四考察中文读者通过预期执行切分词 $n+1$ 的可能性，实验四假设如果提示词 $n+1$ 右侧边界或者剥夺预视加工的掩蔽条件能够影响预测性效应，则意味着中文读者通过对词汇自上而下的预期加工执行预视加工中词切分过程。”前三个实验为什么不提？只写实验四是否唐突？最后，比如什么叫“实验 2 中——提示词 $n+1$ 右侧边界”？看完这篇文章后，才理解到作者传达的意思。因此，在前文需要给予说明。

回应：我们完全接受专家的宝贵意见，在本版修改稿的最后一段我们仔细给出了本研究四个实验的实验逻辑和假设。最后一段修改后的内容如下：

“词频效应和预测性效应是词汇加工过程中的两个常见现象，它们反映两个不同性质的、独立的词汇加工过程，E-Z 读者模型和 SWIFT 模型都提出两个相对独立的函数对之进行描述(Rayner et al., 2004; Engbert et al., 2005; Schad & Engbert, 2012)。据此，本研究分别从词频效应和预测性效应两个角度考察预视切分词 $n+1$ 的加工形式问题。如果读者通过识别字词切分词 $n+1$ ，那意味着在预视中能够达到或者至少部分达到早期词汇加工阶段，基于这个假设，剥夺词汇预视会将预视加工所致的词频效应“推延”至注视加工中，因此掩蔽词 n 右侧词汇将导致词频效应增大，与此同时提示词 $n+1$ 右侧边界的会促进词汇加工，进而将减少词频效应。本研究的前三项实验检验上述假设，若没有产生符合预期的呈现条件变量与词频变量间的交互作用，则意味着读者不可能基于自下而上的字词加工执行切分词 $n+1$ 任务。基于同样逻辑，若中文读者通过自上而下的预期确定词 $n+1$ 的右侧边界，掩蔽词 n 右侧词汇(剥

夺预览)将导致预测性效应增大，同时提示词 n+1 右侧边界的掩蔽将减少预测性效应。实验四包含上述两种掩蔽条件，以此检验“读者通过预期切分词 n+1”的可能性。”

请专家审查。

意见 6: 实验 1 中发现“掩蔽条件完全剥夺预览加工并未导致中文词频效应增大”是没有探测到？还是前提逻辑假设不正确？还是真的这样的结论“剥夺预览加工不能导致中文词频效应增大”。

回应:

本研究的假设是基于一个基本的背景事实：句子理解是建立在词汇被成功识别的基础之上，实验中各个掩蔽条件虽然影响句子阅读效率也影响词汇加工过程，但是各条件下回答问题的正确率不受到呈现条件的影响，这意味着句子中的词汇是被充分识别加工过的(只是相对于控制条件需要更多的时间而已)。因此，相对于正常呈现，剥夺预览意味着将“对文字预览加工程度“推延”至“注视”当中被加工。基于这个考虑，如果预览加工程度能够显示出词频效应，那么剥夺预览加工意味着将“正常情况下(控制条件)这个部分的词频效应”推延至“注视”加工中来，这样会导致高频词和低频词之间的差异量增大，这种情况就意味着词频效应的增大。

没有探测到“掩蔽条件完全剥夺预览加工并未导致中文词频效应增大”意味着两个可能：可能一，结论“剥夺预览加工不能导致中文词频效应增大”为真；可能二，剥夺预览加工导致中文词频效应量的增大较少，因而较难发现显著的交互作用(呈现条件和词频)，不过由表 1 的数据可知，以凝视时间为例，控制条件下低高词频的差异为 32 ms，在掩蔽条件下低高词频的差异却为 21 ms，根本没有增大的趋势，由此可以排除“可能二”。另一个能打消专家疑问的事实是：如果“蔽条件完全剥夺预览加工并未导致中文词频效应增大”是前提逻辑假设不正确；那么实验四剥夺预览加工对预测性效应的影响模式应该与实验一的结果一致，但是由表 4 的数据和结果可知事实并非如此。

意见 7: 实验 2 中引入“外源性注意”，能否给予进一步说明，为什么将此因素引入？之后的实验 3 又是如何排除的？

回应: 实验二发现相对于控制条件，词 n+1 右侧掩蔽条件减少了首次注视时间和凝视时间。对于这个现象存在两种可能的原因：一、外源性注意所致，词 n+1 突然出现这种视觉变化导致注视快速向右移动；二、词 n+1 右侧掩蔽条件提示词 n+1 的右侧边界促进切分词 n+1，

进而在影响随后的注视持续时间。如果是外源性注意导致实验二的结果，那就意味着实验二结果不能说明读者能够预视切分词 $n+1$ ，为排除这种可能，并收集支持前两项实验结论的其他证据，组织实验三考察不完全剥夺词 $n+1$ 上两个汉字的预视对词频效应的影响。

实验三采用与实验二相似的研究范式，该实验中的掩蔽条件都是采用眼动随动显示技术对注视点右侧文字进行的操作，不同的是，实验三不以词为单元显现汉字，而是以相邻的两个不属于同一词汇的双字为单元显现汉字。由于实验句子完全是由双字词构成(6 或 7 个双字词)，证据表明中文读者将首次注视落于双字词的左侧字的概率明显多于落于其右侧字的概率(Yan et al., 2009; 白学军等人, 2013); 因此这种操作可以实现两个作用: (1)不完全剥夺对词 $n+1$ 上两个汉字的预视加工; (2)没有词 $n+1$ 右侧边界的提示线索。如果实验三非正常呈现条件对眼动指标的影响模式与实验二不同, 则可以排除实验二结果由实验范式本身所引入的外源性注意所致; 与此同时呈现条件与词频变量之间不存在交互作用, 则是支持实验一和实验二的推论的另一个证据: 中文读者对词 $n+1$ 的预视加工达不到词汇水平。

意见 8: 后面的总讨论还需要再修改，为什么“词频主要影响注视时的加工过程(文字加工的较晚阶段)，而预测性则影响预视时的加工过程(文字加工的较早阶段)。”?

回应: 谢谢专家的意见。本研究发现“词频主要影响注视时的加工过程(文字加工的较晚阶段)，而预测性则影响预视时的加工过程(文字加工的较早阶段)”，这个结果与以往针对拼音文字阅读的结果并不一致。为了清晰论述这个现象我们在“6.2 词频、预测性对词汇加工的影响及其对词切分的启示”部分增加第二段内容如下:

“不过，在拼音文字阅读研究领域内，基于 ERP 的实验发现，词频影响 P200 的波幅，词汇的预测性则只影响 N400 的波幅，这表明词频对词汇加工的影响发生在较早阶段，而预测性影响词汇的晚期加工(Dambacher, Kliegl, Hofmann, & Jacobs, 2006)。在这个方面，拼音文字与中文不同的可能原因有三点：一、中文词汇加工前存在明显的汉字加工阶段，且每个汉字都有较强的独立性含义(申薇，李兴珊，2012; Li et al., 2009, 2011)，汉字加工在时间进程上推迟了词汇加工，因而词频影响文字加工的稍晚阶段可以理解。二、针对阅读中词汇加工过程的 ERP 实验并不能保证正常的预视加工，其实验结果并不能揭示对词汇的预视加工过程，因而并不能排除预测性同样影响早期词汇加工进程的可能。三、中文阅读中存在的词切分现象可能是导致两者不同的重要原因。”

意见 9: 第 21 页，本研究是如何支持“李兴珊等人(2009, 2011)提出的词切分模型”的? 能否

细致展开？

回应：谢谢专家的意见。我们在“讨论”部分的最后一段针对本研究的实验结果和推论“如何支持李兴珊等人(2009, 2011)提出的词切分模型”进行了较为详细的描述，主要包括 3 点：1、我们发现对中文词预视加工达不到词汇水平，这点与该模型强调存在汉字加工环节的观点一致；2、本研究显示中文读者通过预期(自上而下)加工切分词 $n+1$ ，可见“词切分与词汇加工相关”；3、鉴于词 $n+1$ 上汉字的预视加工达不到词汇水平(前三项实验结论)，提示词 $n+1$ 右侧边界减少预测效应，意味着预视中对词长信息的预期能够影响注视中的汉字加工；剥夺词 $n+1$ 上汉字的预视加工完全消除预测效应则意味着对词汇预期加工与对词 $n+1$ 上汉字的预视加工间存在交互作用；可见对于特定中文词汇而言，词汇加工与汉字加工之间并非单向联系，它们之间存在交互影响过程。当然本研究对李兴珊等人(2009, 2011)提出的词切分模型也有启示。比如我们发现中文词的词频和预测性对词汇加工过程的影响作用发生在不同阶段(词频主要影响较晚期文字加工，而预测性则主要影响较早期文字加工)，鉴于该模型还没有涵盖词频和预测性如何影响词汇加工的机制，这显然有助于完善模型。

基于上述考虑我们对“6.3 对现有模型的启示”第二段修改如下：

“在充分考虑汉字加工的基础上，李兴珊等人(2009, 2011)提出的词切分模型认为词切分与词汇加工密切相关，词汇加工建立的汉字加工的基础之上，汉字加工与词汇加工之间存在交互作用，这个词切分模型被不断的验证和修正(Liu & Li, 2014; Li & Shen, 2013; Ma, Li, & Rayner, 2014; Gu et al., 2015; Gu & Li, 2015)。同样，本研究对此模型亦是既有支持，也有启发：首先、本研究发现对中文词预视加工达不到词汇水平，这与该模型强调汉字加工阶段的观点一致。其次、本研究显示中文读者通过词汇预期(自上而下加工)切分词 $n+1$ ，可见“词切分与词汇加工相关”。再次、鉴于词 $n+1$ 上汉字的预视加工达不到词汇水平(前三项实验结论)，提示词 $n+1$ 右侧边界减少预测性效应，意味着词汇预期提示词长能够影响汉字加工；剥夺预视加工完全消除预测效应则意味着对词汇预期加工与对词 $n+1$ 上汉字的预视加工间存在交互作用；可见对于特定中文词汇而言，词汇加工与汉字加工之间并非单向联系，它们之间确实存在交互影响过程。最后，本研究显示词频和预测性对词汇加工过程的影响作用发生在不同阶段(词频主要影响较晚期文字加工，而预测性则主要影响较早期文字加工)，鉴于该模型还没有涵盖词频和预测性如何影响词汇加工的机制，这显然有助于完善模型。”

意见 10：前 3 个实验中，每个实验都只有两个实验条件，为什么不用 t-test 而采用单因素方差分析？

回应：由于整体分析是针对整个句子阅读时的眼动数据分析，呈现方式只有两个实验条件因此是可以用 t-test 分析的。审稿专家二在其第 2 条意见中提出采用整体分析的必要性。在认真考虑审稿专家二意见后我们认为，其实整体分析的价值不是很大（详细原因请查看本节页码 6 的对审稿专家二意见 4 的回复），因此在本版修改稿中我们将剔除了整体分析结果。

意见 11：所有数据结果的描述，都是提到各种效应显著等，能否修改为——差异显著的具体方向？比如“对于凝视时间，呈现条件主效应显著”，修改为：“掩蔽条件比控制条件下的凝视时间较长”，标明差异的方向性。另外，对于诸多不显著的结果，有必要一一列出吗？在数据部分，比如第 9 页 2.3 讨论部分，提到“呈现条件影响所有的眼动数据”能否写明是如何“影响”的？。

回应：对于数据结果的描述，只是提到各种效应显著，没有报告具体的差异显著的方向的原因有两个：一、基于词汇兴趣区域眼动数据的分析是两因素的重复测量方差分析，“主效应显著”这个统计结果也没有包含方向信息，为了表述严谨，没有直接确定差异的方向性；二、各个表格里面(表 1 至表 4)都给出了各呈现条件下的均值和标准差，差异的方向性一目了然，与此同时遵循阅读眼动研究的惯例，修改稿的文字描述部分没有直接给出差异的方向性信息。不过，在讨论部分（包括各个实验后的分讨论和总讨论）相关的描述中我们给出了相应指标差异的方向性。

对于诸多不显著的结果一一列出的原因是：基于词兴趣区域的眼动数据一般反映词汇加工情况，但单个眼动数据受多种因素的影响从而在反映心理过程上不够灵敏(闫国利, 熊建萍, 臧传丽, 余莉莉, 崔磊, 白学军, 2013), 因此我们报告多个能够反映词汇加工过程的眼动指标结果。

意见 12：非常关键的——“图 1 实验一呈现条件举例”介绍——本人在第一次阅读时是非常困惑的——是只有实验一吗？——本文阅读到后面才明白这个地方。

回应：非常感激专家提出这个意见，“图 1 实验一呈现条件举例”是由于作者笔误导致。图 1 描述的举例是四项实验的所有呈现条件举例，书写错误致使句子不能表达正确意思。在本版修改稿，我们已经将这个题目改成“图 2 四项实验中的各个呈现条件举例”（修改稿件中在这个图之前又增加了一个图）。

意见 13：实验材料、实验设计问题；第六页，试验一材料没有控制字频，很多研究已经表

明字频会影响眼动模式。实验已经如此，可以在讨论部分提到该材料的局限性。另外实验是 40 个句子，40 对高频低频词吗？第七页，实验设计部分第二段，“6 词句和 7 词句”最好作图说明，关键词在哪个位置？。

回应：前三项实验通过考查的是剥夺预视和提示词 n+1 右侧边界对词频效应的影响，进而推测词切分的参考线索问题。虽然很多研究表明字频会影响眼动模式，但字频效应与本研究目的的关系确实不大，因此前三项实验材料中的目标词汇并没有操控字频变量。审稿专家一认为前三项实验并没有操控汉字加工相关变量，因此在结论中给出“存在汉字加工阶段”是过度推论，因此根据前三项实验结果，最终的结论应该是“中文读者对词 n+1 的预视加工达不到词汇水平”，这个结论与“存在汉字加工阶段”并不冲突，在“总讨论”部分，我们对“汉字加工”进行了探讨。

前三项实验的目标词是 40 个高低词频对，每个词对中的两个词汇意思相近，它们都可以放置在同一框架句子中相同位置（都在句子的中心部位，比如第三或第四个词汇为目标词汇），遵照专家意见在本版的修稿中，我们给出了一个图示说明（见图 1）。

司令调集后方 战士 支援前线。	高频词条件
司令调集后方 官兵 支援前线。	低频词条件

图 1 框架句子包含目标词对图示(框架句子中斜体加粗词汇为目标词汇)

意见 14：第十二页，讨论部分，第一段“实验二通过掩蔽范式考察提示词 n+1 右侧边界对词频效应的影响”不清楚。“因此呈现条件主效应结果只能说明说明提供词 n+1 右侧边界改变了具体眼动模式”“这意味着提供词 n+1 右侧边界并没有促进对词 n+1 的预视加工”第十二页，第二段“提示词 n+1 右侧边界的掩蔽条件减少了词兴趣区的注视时间”同控制条件相比，为什么掩蔽条件下词 n 上的注视时间都变小了？第十五页，讨论部分“综合前三项实验的研究结果可以推测中文读者不太可能通过……”？。

回应：原稿中，对于“3.3 讨论”第一段描述不够清晰，导致专家的疑问。本版修改稿中第一段进行了全面的修改，修改后的内容如下：

“实验二考察提示词 n+1 右侧边界的掩蔽条件对词频效应的影响，进而推测预视切分词 n+1 问题。结果显示，提示词 n+1 右侧边界的掩蔽条件导致各项眼动指标变小(见表 2 数据)。首次注视时间、凝视时间和再注视概率减少反映词汇加工难度降低，但跳读概率减少却反映词汇加工难度增加(Rayner, 1998)，因此呈现条件导致各项眼动数据变小说明提供词 n+1 右

侧边界的掩蔽或许只是改变具体眼动模式，而未必促进了词汇加工。词频变量对各项指标的影响，除了在跳读概率上不显著外，其他指标上均达到显著或边缘显著水平；除跳读概率外，词频和呈现条件的交互作用在各项指标上均不显著，这意味着提示词 n+1 右侧边界的掩蔽条件消除对词 n+2 处的汉字加工，进而影响了基于该字的眼跳过程(Liu, Reichle, & Li, 2015)，但该条件并没有改变其他指标上词频效应同样也意味着读者较难通过自下而上的字词加工切分词 n+1(这与实验一推论基本一致)。”。

第十二页，第二段“提示词 n+1 右侧边界的掩蔽条件减少了词兴趣区的注视时间”同控制条件相比，掩蔽条件导致注视时间都变小的存在两个可能的原因：一、外源性注意所致，词 n+1 突然出现这种视觉变化导致注视快速向右移动；二、词 n+1 右侧掩蔽条件提示词 n+1 的右侧边界促进词切分，进而在影响随后的注视持续时间。此掩蔽条件改变读者眼动模式若是由于外源性注意所致，那就意味着实验二结果不能说明读者能够预视切分词 n+1，为排除这种可能，并收集支持前两项实验结论的其他证据，组织实验三。实验三，考察不完全剥夺词 n+1 上两个汉字的预视对词频效应的影响。

第十五页，讨论部分“综合前三项实验的研究结果可以推测中文读者不太可能通过……”？修改后的内容为：“前三项实验结果一致表明，中文读者对词 n+1 上文字的预视加工达不到词汇水平，由此可见，中文读者不太可能通过自下而上的字词加工确定词 n+1 的右侧边界。”

意见 15：第十八页，讨论部分，“呈现条件与词汇可预测性交互影响凝视时间和再注视概率，进一步的数据分析显示，掩蔽词 n 以右词汇意味着完全剥夺预视效应，该条件下词汇凝视时间和再注视概率完全消失，这意味着词汇预测性主要通过影响预视加工的方式影响中文读者的眼动过程；掩蔽词 n+1 右侧词汇能够提示词 n+1 的右侧边界，但该条件下再注视概率的预测性效应消失，高预测性词汇与低预测性词汇的凝视时间差异减少，这意味提示词 n+1 右侧边界减少了可预测性效应。”

回应：原稿中这段描述的逻辑顺序存在问题，这导致增加专家和读者理解文章的难度。本版修改稿中这段被修改成：

“实验四考察读者基于预期切分词 n+1 的可能性。研究事先假设，如果剥夺预视能够导致预测性效应增大，且提示词 n+1 右侧边界减少预测性效应，则意味着读者通过预期切分词 n+1。结果发现，呈现条件和预测性交互影响凝视时间和再注视概率，简单效应分析后发现，相对于控制条件，剥夺预视导致凝视时间和再注视概率的预测性效应完全消失；提示词

n+1 右侧边界(词 n+1 右侧掩蔽条件)则导致预测性效应在凝视时间上减少，在再注视概率指标上消失。由此可见，提示词 n+1 右侧边界掩蔽条件对预测性效应的影响符合并支持预先假设。不过，剥夺预视的掩蔽条件对预测性效应的影响则与预先假设不符，但考虑到预测性效应反映主观预期对词汇加工过程的影响，是种较为高级的认知过程，根据交互激活理论(interactive activation models)的观点，预测性能够影响早期的低级层面的词汇加工过程(Morton, 1969)，基于此若设定预测性只能影响正常阅读中的早期的词汇预视加工阶段，剥夺预视意味着势必减弱预期对词汇加工过程的影响，可见这个结果也是支持“中文读者基于预期切分词 n+1”推测的证据。”

意见 16：第十九页，第一段，“掩蔽条件完全剥夺预视加工并未导致中文词频效应增大”有可能是没有探测到，另外基本的假设前提是非正确？为什么预视剥夺，就会增加词频效应呢？需要进一步的澄清说明，不然会让读者猜测。

回应：这个问题与专家提出的第一部分意见中的第 6 条相同，在该处我们已经针对这个问题进行了回答。请参见相关回复。

意见 17：第十九页，第三段“预视加工中不能体现词频效应则意味着中文读者在预视中对文字的加工主要体现在汉字加工层面，而在注视加工才能达到词汇加工水平。”第二十页，第一段，“读者基于何种线索切分词 n+1 存在两种可能……”难以从实验设计上来回答该问题。第二十页，第二段，“为了验证中文读者是否依据经验通过自上而下的预期执行切分词 n+1”。有吗？“剥夺预视加工完全消除了可预测性效应”？会消除吗？第二十页，第三段，“词频主要影响注视时的加工过程(文字加工的较晚阶段)，而预测性则影响预视时的加工过程(文字加工的较早阶段)。”确定？第十六页，“在整体分析中，我们仅考察呈现方式对整体分析眼动指标的影响(表 8)。”为什么？是不是预测效应对整体眼动指标影响不显著？

回应：第十九页，第三段“预视加工中不能体现词频效应则意味着中文读者在预视中对文字的加工主要体现在汉字加工层面，而在注视加工才能达到词汇加工水平。”这个推论并不合适，在本版修改稿中相关内容修改为：“词频效应反映词汇也是中文阅读理解的基本加工单元，但预视加工中不能体现词频效应则意味着对文字预视加工不能达到词汇水平，而在注视加工才能达到词汇加工水平”

“第二十页，第一段，“读者基于何种线索切分词 n+1 存在两种可能……”难以从实验设计上来回答该问题。”本论文的实验设计确实并没有直接操控这两个变量(自上而下加工和自

下而上的加工), 但是基于阅读眼动研究领域现有的进展来说, 对词汇加工过程而言, 词频效应与预测性效应是两个相对研究成熟的问题。词频效应更多反映的是自下而上的文字加工过程, 而预测性效应更多反映了自上而下的预期对词汇加工过程的影响。自下而上与自上而下是广泛存在于人类的所有认知过程中, 中文特殊的书写方式意味着读者必须通过某种形式的认知加工完全词切分过程, 因此词切分问题逻辑上存在两种可能加工: 一、通过自下而上的字词加工切分词 $n+1$, 二、通过自上而下的经验预期切分词 $n+1$ 。这两种可能之间并非是相互排斥的, 它们可能都是导致词 $n+1$ 切分的加工过程, 也可能只有其中的一种加工导致词切分。基于这个考虑我们并没有通过实验设计来解决这个词 $n+1$ 的切分问题, 而是通过四项实验分别讨论这两个可能性(前三项实验讨论可能性一, 第四项实验讨论可能性二)。

意见 18: “第二十页, 第二段, “为了验证中文读者是否依据经验通过自上而下的预期执行切分词 $n+1$ ”。有吗? “剥夺预视加工完全消除了可预测性效应”? 会消除吗?”第二十页, 第三段, “词频主要影响注视时的加工过程(文字加工的较晚阶段), 而预测性则影响预视时的加工过程(文字加工的较早阶段)。”确定?

回应: 对于上述问题, 本研究中四个实验的结果都是确定的。在阅读眼动控制研究领域, 词频效应和预测性效应是反映词汇加工过程的两个主要因素, 这两个因素对词汇加工和眼动控制过程的影响都是实时、快速的。综合四项实验结果可见, 掩蔽条件对词频效应和预测性效应的影响模式是不同的, 这就排除了“研究所采用掩蔽范式在在词汇预视加工过程中不敏感”的可能性, 因此根据四项实验结果做出推论是合适、合理、可信的。

意见 19: 第十六页, “在整体分析中, 我们仅考察呈现方式对整体分析眼动指标的影响(表 8)。”为什么? 是不是预测效应对整体眼动指标影响不显著?

回应: 仅考虑呈现方式对整体分析眼动指标的影响是因为整体分析是针对句子所有眼动事件进行的分析, 而预测性效应则只是针对特定单个词汇区域上眼动数据的分析, 因而整体分析没有必要考虑预测性因素。认真考虑专家二对提供整体分析必要性的质疑, 在本版的修改稿中我们剔除了整体分析的研究结果。

专家在这部分意见主要是针对文章各个分讨论和总讨论的内容, 综合考虑三个专家的已经, 我们在这版的修稿中, 已经对这些内容进行了比较全面的改写, 不仅改变了原有的表述方式, 同时也改变讨论的结构, 基本是全面的修改, 请专家审查相关内容。

意见 20：第三页，摘要部分，“不是又外源性注意所致”。第四页，第一段，“这表明拼音文字阅读中词汇加工确实能够实时影响的眼动过程”。第四页，第二段“因而考察预视加工中的词频与预测性效应为解决上述理论问题提供突破口。”病句第六页，第一段，“男生 20 民”。第六页，最后一段，“带上眼动议头盔”，能否贴个眼动仪器头盔的图？第八页，文章图表 4 提到字号？“考察字号对上述各项眼动指标的影响(表 2)。”？第十页，“实验结束后可获得报酬 20¥。。”第十二页，最后一行，“在”“进而在影响随后的注视持续时间”。第十三页，第二段，“那么可以到达目的(1)”第十五页，讨论部分，(Li & Ma, 2012; Li & Pollatsek, 2011; Gu & Li, 2015; Yan et al., 2010, 2015; Shu et al., 2011; 张智君等人, 2012; 刘志方等人, 2013)“等人”是 APA 的格式吗？第十六页，第一行，“有”“不是有该条件引入的外源性注意所致”第十六页，“控制框架句子的通顺性和难以程度”“难以”。第十九页，第一段，“出”“因此只要中文读者在预视中对词 n+1 出文字的预视加工”。

回应：谢谢专家提出上述宝贵意见，原稿件中文字的错别字和病句较多，本版修改稿中我们仔细阅读，认真校对了书写错误。关于在文章方法部分处“贴个眼动仪器头盔的图”，由于 Eye Link 型眼动仪器比较常用，在阅读眼动研究领域目前文章中贴出相关图片，所以遵循惯例，本版的修改稿中没有给出相关图片。关于参考文献，APA 的格式没有相关的参考文献规定；心理学报上规定第二次中文参考文献写法格式是：“张三等，2013”。按照这个要求，本版修改稿中对参考文献进行了相应的修改。

再次非常感谢三位审稿专家对本文提出多项建设性意见，基于专家的修改意见和作者的重新思考认识，我们对本论文进行了全面的修改，修改过的内容采用**红色字体**标示出来。另外通过综合考虑三位审稿专家的意见，将论文的题目修改成《**中文阅读预视加工中的词频和预测性效应及其对词切分的启示：基于眼动的证据**》，我们认为这个题目更能概括论文的主题内容。

再次感谢三位审稿专家提出的宝贵意见。本次改动的地方较多，改动的篇幅比较大，请专家审查。

第二轮

审稿人 1 意见：作者已根据审稿意见做出了合理的修改，建议发表

审稿人 2 意见：作者认真回答了本人的意见，审稿人对其答复表示满意。同意发表。但是，论文还是存在一些小问题，建议作者认真检查：1. 在 4.1.1 被试部分，“宁波大学 40 名大学生参加本次实验，男、女各 20 人。被试的视力或矫正视力正常，之前均未参加过类似实验。实验结束后可获得报酬 20 元。”20 元请用汉字，不要用符号。2. 在 5.1.2 实验材料、实验仪器和实验程序，“严格控制框架句子的通顺性和难度(关于实验材料的具体细节请参见文献：白学军，曹玉肖，顾俊娟，郭志英，闰国利，2011)”。其中，“闰国利”应该为“闫国利”。以上的小问题，文章中还有，在提交正式稿件前，建议作者认真检查

回应：谢谢审稿专家的肯定，在本版修改稿中，我们认真核对检查了文章中的错误，对所发现的错误均做了修改。

审稿人 3 意见：

意见 1：主要是统计的问题，本人还是认为，报告统计结果包含“方向信息”会更好一些。还有个小问题“Li, X., Bicknell, K., Liu, P., Wei, W., & Rayner, K. (2015). Reading is fundamentally similar across disparate writing systems: A systematic characterization of how words and characters influence eye movements in Chinese reading. *Journal of Experimental Psychology: General*, 143, 895-913.”这篇引文的正式出版年代是“2014”

回应：谢谢审稿专家的意见，在本版修改稿中我们按照专家的意见报告统计结果的方向信息，将“呈现方式主效应显著”提法修改成“掩蔽条件导致凝视时间显著增加”。另外考虑到实验四的两个掩蔽呈现条件对眼动指标的影响趋势并不一致，因此在呈现实验四结果时仍采用“呈现方式主效应显著”提法。对于文中错误表示参考文献的年代，也都做了修正。请专家审查。

主编意见：

研究探讨了中文阅读预视加工中的词频和预测性效应及其对词切分的影响，对于阅读模型的完善具有一定的启示作用。建议作者在写作和逻辑上进一步斟酌完善。

意见 1：P21 引言部分的第一段和第二段之间补充合理的过度。作者解释了 E-Z 模型和 SWIFT 模型，但是却没有进行比较和区分，就跳到了“词汇的词频和文本背景对之的预测程度是衡量词汇加工难度的两项基本指标。”这样有点过于跳跃，难以合理地引出词频与预测效应。

回应：谢谢主编提出的宝贵意见。第一段中我们首先引出 E-Z 模型和 SWIFT 模型对词汇加

工两阶段的描述，主要是为了强调“两个模型都认为词汇加工可以被区分为两个子阶段，且尽管两个模型对词汇加工两阶段的具体描述不同，但它们都认为词汇难度(词汇的词频和文本背景对之的预测程度)是影响两个子词汇加工阶段的基本因素”。基于两个模型在词汇加工过程上的共识，本研究考察中文阅读中对词汇的预视加工所能达到具体程度问题，并在此基础上探讨词切分问题。本研究的主要目的不是区分 E-Z 模型和 SWIFT 模型在中文阅读中孰对孰错问题，因而我们认为没有必要对 E-Z 读者模型和 SWIFT 模型进行比较和区分。

然而，我们认真阅读原稿后发现，原稿的 P21 引言部分的第一段和第二段之间的过度确实不够合理。原稿的描写方式确实存在思维跳跃问题，这最终导致前言没有清晰表达作者的原有思路。在本版修改稿中，第二段开头部分被修改成为：“尽管上述两个模型对词汇加工两阶段的具体描述不同，但它们都认为词汇难度(词汇的词频和文本背景对之的预测程度)是影响词汇加工两阶段的基本因素。比如，E-Z 读者模型认为，熟悉性检验阶段和词汇通达阶段所需要的时间是词频和预测性的函数，词频和预测性越低两个子阶段所需时间越长(Rayner, Ashby, Pollatsek, & Reichle, 2004; Reichle et al., 2006; Schotter et al., 2014)。SWIFT 模型则假设词频和预测性影响最大词汇激活度，词频和预测性越低其最大激活度越大，对之的加工所需时间越多”。第二段的结尾部分被修改成为：“但这些证据仍不能确定词汇难度对词汇加工两个阶段的分别影响程度问题。”通过此修改希望能够理顺前三段的思路。

意见 2：E-Z 模型和 SWIFT 模型贯穿本文始终，建议作者做出更为详实的介绍，从 E-Z 模型和 SWIFT 模型之间的比较以及加工阶段效应的不同从而引出本文的论点。

回应：谢谢主编提出的宝贵意见。E-Z 模型和 SWIFT 模型贯穿本文始终，两个模型同样都主要包括两个模块的内容：一、词汇加工模块，二、眼跳系统模块；本研究的主要目的只涉及词汇加工模块的内容。尽管 E-Z 模型和 SWIFT 模型在关于词汇加工阶段方面的看法存在细节上的差异，但是本研究的目标并非检验两个模型在中文阅读中词汇加工过程方面孰对孰错问题。E-Z 模型和 SWIFT 模型都是在拼音文字阅读中的研究基础上建立起来的计算模型，尽管它们对词汇加工的具体看法存在分歧，但是两个模型间的共识也是非常明显的，比如，它们都假设词汇加工过程可以被区分成两个子阶段。基于它们对拼音文字阅读中词汇加工阶段上存在的共识，本研究考察中文阅读中词汇加工过程的特殊性，进而推测出“原用来解释拼音文字阅读中词汇加工过程的模型并不能直接用来解释中文阅读中的词汇加工过程”。结合已有的研究结果，本研究在总讨论部分提及李兴珊等人提出模型在解释中文阅读词汇加工过程方面的优势。总之，本版修改稿中我们增加了“E-Z 模型和 SWIFT 模型在词汇加工阶段

问题上所取得的共识”方面的描述，但根据本研究的主要目的没有比较 E-Z 模型和 SWIFT 模型之间不同。

意见 3: P21 摘要处“综合四项实验结果可知，词频和预测性以不同的形式影响中文词汇加工过程；中文阅读中的文字加工阶段有其独特性。”谈到了词频的影响，但是文中大部分结果都是词频效应不显著，而且作者的结论部分也只是谈到了预视对词切分的影响，建议进一步确认修改。

回应: 谢谢主编提出的宝贵意见。本研究中，我们主要通过考察对中文词汇的预视加工程度问题，推测词切分的加工性质问题。因而摘要处“综合四项实验结果可知，词频和预测性以不同的形式影响中文词汇加工过程；中文阅读中的文字加工阶段有其独特性。”的描述并不合适，在本版修改稿的中文摘要进行了修改，剔除了上述说明。为力图与本文的结论保持一致，上述文字被修改成“中文读者较难通过自下而上的文字识别切分词 $n+1$ ，自上而下的词汇预期则是切分词 $n+1$ 的加工形式”。

关于词频效应的显著性问题。虽然多个眼动指标显示其词频主效应不显著，这些眼动指标都是“首次注视时间、跳读概率和再注视概率”，需要指出的是，前三项实验的凝视时间指标上的词频效应都达到显著水平。本研究主要基于凝视时间这个指标做出推论，这是因为，首次注视时间、跳读概率和再注视概率相对于凝视时间在衡量词汇加工过程方面不够敏感。首次注视时间指标中混淆在某个词汇上只有单一注视点和多个注视点的情况，尤其是在有多个注视点时，读者通过多个注视点识别加工词汇，而首次注视时间显然不能全面衡量词汇加工过程；跳读概率和再注视概率都会损失相当的信息，从而使得它们对词汇加工过程的差异不够敏感(闫国利等人，2013)。

意见 4: P35 的第二段倒数第三行，“结合以往研究可以推测，中文词汇加工至少可以被区分出汉字加工阶段和词汇加工阶段，这对两阶段划分是个挑战，但其中的词汇加工阶段是否仍可以进一步划分为两个子加工阶段仍需要深入的研究。”该讨论如何与 E-Z 模型和 SWIFT 模型建立联系？区分又是什么？该结果可以说是研究者一个重要推论，因此有必要与以往的研究进行一个对比。

回应: 谢谢主编提出的宝贵意见。本研究是基于“E-Z 模型和 SWIFT 模型在词汇加工过程的两个子阶段”方面的共识考察中文阅读中对词汇的预视加工程度问题。针对词频效应的实验发现(前三项实验)，剥夺预视加工，提示词 $n+1$ 右侧边界都不能影响词频效应，这说明词频

效应主要发生在注视加工阶段,这个结果也意味着中文读者对词汇的预视加工达不到词汇水平。“6.1 词 n+1 的预视加工程度和词汇加工阶段探讨”的第二段,首先回顾以往涉及中文阅读中字词加工的研究文献,基于这些研究结论,第二段推测“中文词汇加工存在汉字加工阶段”,我们认为这个事实是不能被“E-Z 读者模型和 SWIFT 模型关于词汇加工两个阶段假设”所解释,是中文阅读中词汇加工的特殊性(当然,这个特殊性是相对于拼音文字而言)。

由于本研究中的实验条件并没有操控汉字加工过程,因而不能就“汉字加工”问题与其他已有的研究作对比,因而“6.1 词 n+1 的预视加工程度和词汇加工阶段探讨”的基本思路是结合以往的研究结论和本研究的实验结果探讨中文词汇加工阶段问题(一位外审专家质疑在“结论”中提及汉字加工的合理性,因而在第一版的修改稿中,我们将汉字加工问题放置在讨论中去阐述),并在此基础上说明“E-Z 读者模型和 SWIFT 模型关于词汇加工两个阶段假设不能直接解释中文词汇的加工过程”。并在“6.3 对现有模型的启示”的第二段讨论强调汉字加工模型(李兴珊等人(2009,2011)提出的词切分模型)在解释“中文词汇加工过程”方面的优越性。

为了清晰表达上述思路,我们将该段文字修改成“结合以往研究可以推测,中文词汇加工存在汉字加工阶段。E-Z 读者模型和 SWIFT 模型所提及的词汇加工的两个阶段本质上都是词汇水平的加工阶段,因而不能直接解释中文词汇的加工过程。”另外,在“6.3 对现有模型的启示”的第一段中将相应的文字修改成为“其次, E-Z 读者模型和 SWIFT 模型虽然都认定词汇的词频属性和预测性属性以不同的路径影响词汇加工过程,但它们都认为词频和预测性都影响词汇加工的两个阶段都有影响作用,针对中文阅读的研究显示,词频主要影响注视时的加工过程(文字加工的较晚阶段),而预测性则主要影响预视时的加工过程(文字加工的较早阶段),中文词的词频和预测性以影响不同词汇加工阶段,这显然与上述理论模型假设不一致。”我们通过这些修改希望能够阐述中文阅读与拼音文字阅读过程中词汇加工方面的不同。

意见 5: 仔细通读全文,修改错别字。如 p32 倒数第 2 段第三行的“首字字颇、尾字字颇和整词词颇。”应是“词频”。

回应: 谢谢主编提出的宝贵意见。本版修改稿中我们已经认真修改了全文中的错别字。文中所有被修改过的内容统一采用蓝色字体标出。