

《心理学报》审稿意见与作者回应

题目：中文阅读中的字词激活模式：来自提示词边界延时效应的证据

作者：刘志方, 张智君, 杨桂芳

第一轮

审稿人 1 意见： 本实验通过在不同延迟时间内提供 N 和 N+1 正确或错误的词界标识，比较错误标识影响发生的时程，尝试对中文阅读理解中词切分的阶段进行推论。文章涉及的是一个有意义的问题，但这一实验给人的总体感觉是操纵了很多东西，但为何进行操纵，文章所表现出来的思路不清楚，此外，整个研究设计的逻辑也存在问题，从而影响结果的可靠性。以下对可能存在的问题进行说明：

意见 1： 首先，从研究逻辑上来说，作者现有的分析方法其实很难说明是词 N+1 的预视对随后注视 N 的影响，还是 N 的加工受到预视 N+1 的影响，也就是时间顺序上的效应很难区分得清楚。作者现在是认定 N+1 在前，而后才是 N 的效应，但事实上 N+1 的错误标示效应完全可能影响到 N 的阅读，而现在的研究设计和分析方法并无法把这两种效应区分开，因此也就很难确定谁的影响在前，谁的影响在后，这种影响又是在字或是在词汇加工的层面上产生。

回应： 谢谢审稿专家提出的宝贵意见。

本研究是希望通过考察中文阅读中词切分的时间进程特点，推测中文阅读中的汉字加工与词汇加工的关系问题。针对此问题理论上存在三种可能的假设：向上反馈假设认为两者之间是单向联系，识别汉字才能导致词汇加工完成；整体假设强调中文词汇是以整词的方式通达的；交互激活假设认为，汉字加工与词汇加工之间并非单向联系，汉字加工是词汇加工的必经阶段，但是词汇加工还会反向影响汉字加工过程。根据其基本观点，三种可能的假设对词切分的时间进程存在不同的看法：根据向上反馈和交互激活假设都可以推测词切分存在时间进程，也即是，词切分应该有明显的时间进程，随着时间的推移词切分逐步完成。根据整体假设可以推断，既然词汇从一开始就从整体激活，那么词切分应该没有时间进程特点，因此提示词汇边界协助切分的实验操作不会出现延迟时间效应。然而，目前阅读研究领域内，在汉字加工与词汇加工层面上检验上述理论假设的实证研究尚不多见。

参照注视点位置，词汇加工可以被区分称预视加工和注视加工两个部分。本研究的两个

实验分别考察预视加工和注视加工中词切分的时间进程特点,解决上述理论难题。实验一考察预视切分词 $n+1$ 的时间进程特点。实验一(A)直接操控词 $n+1$ 的字体改变颜色的延迟时间,结果发现这些操作没有影响总阅读时间,但积极影响基于词兴趣区域的眼动指标(相对于控制条件,提示改变词 $n+1$ 的字体颜色导致平均凝视时间、再注视概率和再注视概率减少),且这种影响随着操作本身延迟时间的增加而减弱。这个结果与前人的研究一致,张智君等人(2012)认为,提示词 $n+1$ 右侧边界改变词 $n+1$ 成为下一眼跳目标的可能性,说明中文读者至少能够在一定程度上预视切分词 $n+1$ 。但是对于本实验操作而言,还存在其他可能的解释:

1. 外源性注意引导眼跳快速发动指向词 $n+1$,进而导致平均凝视时间和再注视概率减少;
2. 改变词 $n+1$ 的字体颜色同时也提示了词 n 的右侧边界,该实验操作是通过提示词 n 右侧边界影响眼动过程。实验一(B)和实验二(A)分别排除了上述两个可能。

严格来说实验一(A)并没有促进和阻碍词汇加工过程(因为总阅读时间没有改变),因而我们将专家的意见“词 $N+1$ 的预视对随后注视 N 的影响,还是 N 的加工受到预视 $N+1$ 的影响”解读为“对特定词汇而言,它被预视时其呈现方式显现了词汇边界影响注视其时的眼动过程(是否再注视,以及凝视时间);还是注视之时的眼动过程(是否再注视,以及凝视时间)受到了其右侧词汇呈现方式的影响”。实际上根据实验二(A)的结果可以澄清上述混淆,如果实验一(A)结果的原因是“注视之时的眼动过程受到了其右侧词汇呈现方式的影响”,那么“实验二(A)的中提示词 n 右侧边界的呈现条件对眼动模式影响的模式”应该与“实验一(A)的中提示词 $n+1$ 右侧边界的呈现条件对眼动模式影响的模式”相同,事实却并非如此;由此可见,提示词 $n+1$ 右侧边界的呈现条件是通过影响“与词 $n+1$ 相关眼动事件”导致实验一(A)结果的。

在文本中增加的空格若正确表示词汇边界不影响总阅读时间,却积极影响词兴趣区域内眼动数据;但是空格若提示错误的词边界信息则消极影响总阅读时间和词兴趣区内眼动数据;白学军与其合作通过这种“反向证明”的方式,证明中文阅读中词切分的存在(Bai et al., 2008, 2011, 2013; Shen et al., 2010, 2012; Blythe et al., 2012)。本研究考察预视中的词切分过程,也通过实验一(B)提示错误的词 $n+1$ 边界考察预视切分词 $n+1$ 过程,结果发现这些呈现条件也影响总阅读时间和词汇兴趣区域内眼动数据。然而,实验一(B)的结果仍然面临“词 $N+1$ 的错误标示效应完全可能影响到对词 N 的阅读”的可能,审稿专家认为“现在的研究设计和分析方法并无法把这两种效应区分开”,不过两个理由表明实验一(B)的词 $N+1$ 错误标示效应是通过影响到对词 $N+1$ 加工导致,而非通过影响对词 N 的加工所导致: 1.如果实验一(B)词 $N+1$ 的错误标示效应是影响对词 N 的加工所导致,那么应该看到实验二(B)的结果应该与实验一(B)相似(因为实验二(B)提示错误的词 n 边界只能是主要影响词 N 加工);但实

验一(B)提示错误词 $n+1$ 边界对总阅读时间的影响随延迟时间的增加呈现倒“U”趋势与实验二(B)完全不同, 这表明这两项分实验的提示错误词边界影响不同阶段、位置上的词切分过程。2. 证据表明副中央凹内的中文汉字的预视加工能够引导眼跳落点(Liu et al., 2015), 实验一(B)发现, 40 ms 延迟时间下提示错误的词 $n+1$ 边界导致读者采用较为保守的眼跳策略(更多地注视双字词的首字, 较少地注视尾字), 这说明提示错误的词 $n+1$ 的边界确实增加的预视加工难度。

根据初稿的描述, 我们试图通过实验一(B)的结果推测, 提示错误的词 $n+1$ 边界对词切分的影响发生在字加工层面还是发生在词汇加工的层面上。在专家的意见的启示下, 我们人为本论文实际上解决的是“汉字加工与词汇加工关系”问题, 不是考察“提示错误的词 $n+1$ 边界对词切分的影响发生在字加工层面还是发生在词汇加工的层面上”的问题。由于, 向上反馈假设和整体假设和实验一(B)的预测不同: 根据向上反馈假设, 提示错误的词 $+1$ 边界对阅读理解和词切分没有影响; 根据整体假设提示错误的词 $+1$ 边界对阅读理解和词切分的影响应该随着延迟时间的增加而减弱。在本版修改稿中, 我们重新组织了文章的前言, 分讨论和总讨论的结构, 在前言中根据已有理论假设(向上反馈假设、整体假设和字词交互激活假设)提出, 整理研究所要解决的理论问题, 并在解决如上理论问题目的的基础上阐述四项分实验所要操作的具体内容。各实验的分讨论和总讨论中我们剔除了“提示错误词边界对词切分的影响发生在字加工层面还是发生在词汇加工的层面上”相关说法, 并根据前言, 组织讨论内容。

意见 2: 其次, 眼动数据的测量包括时间和空间两方面, 目前作者报告的主要是时间的指标, 而对空间指标, 如落点位置, 很少进行分析。事实上, 在空间指标方面, 作者还应报告在词切分标志出现时, 其前一个关键区(及关键词中)有多少注视点、以及注视点在区(或词)中所在的位置、这一落点位置距离下一个词界标志的起点有多远等。这是因为, 干扰或促进词切分的标志所产生的结果其实可能与某一注视落点与 $n+1$ 颜色边界标示的距离有关。在不同的距离下, 边界标示所包含的信息进入中央凹的程度不一样, 故有可能产生不一样的影响。

回应: 眼动数据的测量包括时间和空间两方面, 目前我们主要报告时间指标, 却较少报告空间指标的原因有以下几点:

1. 本研究主要目的是通过考察中文阅读中的字词加工与词切分的时间进程特点检验向上反馈假设、整体假设和字词交互激活假设; 本研究考察的是字词加工过程, 证据表明时间指标能够很好地反映中西文字识别加工过程(Yan et al., 2006; Rayner et al., 2004; Rayner &

Duffy, 1986; Rayner & Well, 1996; Kliegl et al., 2004), 但眼动的空间指标(比如注视位置)受词汇加工因素的影响不大(Rayner et al., 2001)。因此, 鉴于本研究的目的我们没有采用落点位置等空间指标。

2. 词汇长度是影响眼跳落点的首要因素: 注视词汇越长其被再注视的概率就越大; 词 $n+1$ 越长其被成功切分的概率就越低, 首次注视点也非常容易落在词的首字上(Yan et al., 2010; 刘志方等人, 2011)。本研究采用双字词构成的句子正是为了控制了词长对落点位置的影响。至于词边界标志出现时, 其与当时注视点的距离尽管会影响眼动过程, 但完全由双字词构成实验句子材料可将上述因素作为系统误差平衡掉。尽管“干扰或促进词切分的标志所产生的结果其实可能与某一注视落点与 $n+1$ 颜色边界标示的距离有关”, 但分析这种相关与解决本文的研究目的作用不大。

3. 证据表明, 中文阅读中, 副中央凹内文字的预视加工能够影响眼跳落点(Liu et al., 2015); 因此眼跳落点能够反映预视加工难度问题, 为了检验提示错误的词 $n+1$ 边界是否增加预视切分词 $n+1$ 的加工难度, 分析眼跳落点指标(不同呈现条件下首字与和尾字的注视概率), 结果发现较早提示错误的词 $n+1$ 边界导致读者采用较为保守的眼跳策略(40 ms 延迟时间条件下读者更多地注视双字词的首字, 较少地注视尾字)。实际上, 再注视概率和跳读概率也能够反映眼跳落点数据, 本研究的四项分实验的结果中均有这两个指标描述。因此, 过多细致的提供眼跳落点结果信息会导致文章中包含过多的冗余信息, 可能会给读者理解本文带来困难。

总之, 考虑到文章的研究目的、实验控制方式, 以及文章的简洁性本文主要报告时间指标, 在适当位置报告少量的空间指标。

意见 3: 实验 1A 与 1B 应该当做是呈个混合设计来呈现研究, 因为作者感兴趣的是两个实验中不同延迟的不同趋势。实验二也是如此。在这两个实验中, 如果要强调 A 和 B 的趋势不同, 首先需要证明交互作用的存在。

回应: 本研究没有报告交互作用数值的原因如下: 首先, 实验一(B)是实验一(A)的排除实验。其次, 实验一(B)还有自己的实验目的, 向上反馈假设、整体假设和交互激活假设对“错误地显示词 $n+1$ 右侧边界线索如何影响阅读理解和词汇加工的预测不同, 这与其对实验一(A)的预测是两个相对平行、独立的内容。实验一(A)和实验一(B)是检验两个相互独立的假设, “两个分实验中不同延迟的不同趋势”不是作者主要的兴趣所在。再次, 对于实验二而言, 分实验(A)和分实验(B)同样是检验两个相互独立的内容, 因此没有必要证明交互作用的存在。

意见 4: 作者在分析数据时应该去掉第一个和最后一个兴趣区，因为第一个兴趣区没有接受预视，而最后一个兴趣区没有需要预视的信息，这两个区域的性质与其他区域不同，不能合并分析。

回应: 实验中句子中的所有词汇/汉字都参与实验处理，每个呈现条件中每个句子的兴趣区域数据都包含第一个兴趣区域和最后一个兴趣区域，对所有兴趣区域的数据进行做平均值处理，这两个兴趣区域内的眼动数据作为一种系统误差被平衡在所有呈现条件中，因而将句子中所有兴趣区域内的眼动数据作平均值不会影响实验结果。与此同时，鉴于本研究主要参照总阅读时间衡量字词加工过程，总阅读时间包含对句子中所有字词的加工过程，因而在不排除首兴趣区和尾兴趣区域的眼动指标的情况下做平均值。这种做法在以往的研究中也较为常见，并且已经被接受。

意见 5: 整个句子的总阅读时间等指标对理解词切分发生在字词或句子水平这一精细的研究问题并没有帮助。此外，因为研究感兴趣的是不同指标随时间变化的动态情况，多用折线图（带误差线）来反映数据的趋势会更加清楚。

回应: 谢谢专家的意见。本研究将“总阅读时间”作为揭示本研究目的的主要参考标准,原因如下:

1. 句子中所有的字词都参与实验操作，在阅读句子时读者会采用“权衡平均凝视时间、跳读概率和回视次数的眼动策略”完成任务，因而基于词汇兴趣内的眼动数据难以有效真正地反映词汇加工过程，采用反映“所有汉字词汇加工”的指标更符合实际，只有总阅读时间能够反映“所有汉字词汇加工”，据此本研究以“总阅读时间”作为主要参考指标，以词汇兴趣区内眼动指标作为辅助性参考指标。

2. 本研究中所采用的句子都是经过评定后选择难度最低，通顺性最高的句子作为实验材料，考察阅读中词汇加工的研究领域也都是参照这个标准选择实验句子材料。本研究的所有实验都依照拉丁方的方式在各个实验条件中平衡每个句子，能够保证每个句子均能接受所有的实验处理方式，可见句子理解中涉及词汇加工以外的信息加工因素作为系统误差已经被平衡控制，因此总阅读时间能够客观地反映不同实验条件中的词汇加工、词切分差异。

3. 以往多项研究均采用总阅读时间指标反映词切分与词汇加工问题，比如，白学军与其合作这进行一系列的词切分研究中均以“总阅读时间”主要参考指标(Bai et al., 2008, 2011,2013; Shen et al., 2010, 2012; Blythe et al., 2012); 在句子中所有的字词都参与实验操作

的研究中,总阅读时间往往都是最重要的参考指标(Liversedge et al., 2004; Rayner et al., 2006; Blythe et al., 2009)。本研究通过操作句子中的所有字词的字体颜色改变方式考察词切分现象,据此本研究将“总阅读时间”作为主要的参考标准。

另外,审稿专家建议采用折线图(带误差线)来反映数据的趋势。鉴于若将所有指标都采用折线图表达的话,会导致文章页码增多造成浪费,因此本修改稿中将主要分析指标采用表格方式呈现,对“2.2.2 数据补充分析”中的“首字与和尾字的注视概率均值”采用带误差线折线图的方式反映数据。希望如此修改能够方便阅读。

意见 6: 在实验 1 的 1A 中,首次注视和凝视时间的效应趋势不一样,这两个指标的差异反应了什么,作者需要进行讨论和分析。实验二中的实验 2B 什么不做非词内位置的分析,而只在实验 1B 中进行相关的分析?

回应: 仅从“平均首次注视时间”和“平均凝视时间”两个指标来看,其随着延迟时间的增加呈现相反变化趋势;但有两个因素影响平均凝视时间(“首次注视的持续时间”和“是否再注视”),也即是,在其中一个指标数值保持恒定情况下,平均首次注视时间或者再注视概率增加都会导致平均凝视时间增加。然而,由表 1 的数据可知,相对于控制条件,尽管提示词 $n+1$ 边界导致平均首次注视时间增加,但却同时也导致再注视概率的减少;提示词 $n+1$ 边界的不同延迟时间条件下,平均首次注视时间和再注视概率都是随着时间的增加而增加,并最终导致平均凝视时间接近与控制条件。总之,由于“权衡平均首次注视时间和再注视概率眼动策略”的存在,本实验的词汇兴趣区指标则主要参考平均凝视时间指标。对这个问题我们在本版的修改稿中的“2.3 讨论”的第一段给出简单说明“提示词汇 $n+1$ 边界线索条件中平均首次注视时间的增加不足以抵消再注视概率的减少”。

“实验 2B 什么不做非词内位置的分析,而只在实验 1B 中进行相关的分析”的原因是:“实验一(A)发现,词 $n+1$ 字体颜色改变导致平均凝视时间减少,有两种可能的原因导致上述结果:可能一、提供词 $n+1$ 右侧边界所致;可能二、注视点右侧文字突然改变颜色产生的外源性注意引导眼球快速向右移动。为了排除可能二进行实验一(B)。实验一(B)以相邻但不属于同一词汇的两个相邻双字为单元改变相应汉字的字体颜色,为了排除外源性注意引导读者运动的可能性,有必要提供实验操作所参加的兴趣区内的眼动数据。”但“实验二(A)是改变注视点所在区域内字体颜色,不存在外源性注意引导注视点快速移动的可能性。”在本版修改稿中的上述两项原因分别在“2.2.2 数据补充分析”和“3.2 结果”中给出描述说明。

意见 7: 研究即使观察到错误词汇标示所产生的效应在延迟长或延迟短的时间下出现,也只能说明这种影响会出现得较晚或较早,而不应该说这种影响出现在词或字的层面,如 24 页倒数第二段,“如果对词 n+1 处汉字识别加工由词汇通达开始,那么应该看到在 40 ms 延迟时间点上干扰切分词 n+1 将对总阅读时间和眼动数据有最为严重的影响,而本研究结果却不支持如此假设。这说明,对词 n+1 处汉字识别加工是由字通达开始”,事实上,作者的研究只能证明干扰词汇的影响在稍后期出现,不能证明由字通达出现,因为实验并没有干扰字加工的操纵。其他还有不少地方出现类似的论述。

回应: 谢谢专家提出宝贵意见。近期多项研究表明中文阅读中的文字加工过程符合普遍的语言规律,也即是,词汇也是中文阅读理解的基本加工单元(Bai et al., 2008, 2011,2013; Shen et al., 2010, 2012; Blythe et al., 2012; Li et al., 2014)。然而还有研究表明,在词优效应框架中汉字加工有较强的独立性(申薇, 李兴珊, 2012),汉字加工同样影响中文阅读中的眼动控制过程(Yan et al., 2006; Li et al., 2014)。基于上述两项基本事实,本研究根据已有的理论假设(向上反馈假设、整体假设和字词交互激活假设),从字词加工的时间进程角度出发,考察汉字加工与词汇加工之间的关系问题,严格来说本研究的主要目的不是考察“汉字通达”和“词汇通达”孰先孰后问题。撰写初稿时,由于作者对这些问题的考虑不周,对问题的描述也不够清楚,受审稿专家所提问题的启发,在本版修改稿中,我们针对专家所提的内容进行改动、尽量避免相关描述。文章的前言,分讨论和总讨论都做了相应的修改。请审查。

意见 8: 讨论部分不应该再出现数据分析表和统计数据,这一部分应该放在结果中,如果是一个补充分析,可以在主分析后进行呈现,或如果数据太多,想增加报告内容,则应该在附录中进行呈现。

回应: 根据专家的意见,在综合考虑文章布局之后,我们将总讨论中出现的图和统计数据作为一种补充分析移至“**2.2.2 数据补充分析**”中,并在“**2.3 讨论**”增加相应的论述,本版修改稿的“总讨论”则参照提高概括度原则做了全面的布局和修改。

意见 9: 总体上,整个实验操纵变量的逻辑和理论预期不清,作者在进行延迟时间操纵、以及干扰或促进词切分的变量操纵上,均应在前言中说明操纵的原因,以及根据不同的理论观点,这种操纵对随后各测量指标可能产生何种影响,而不仅仅是根据差异进行解释。建议将讨论中谈到的,作者感兴趣的问题进行归纳,在前言最后部分列出来,并且说明出现什么样的结果能说明什么问题,或支持什么理论,这样整个研究思路会才容易

回应：根据专家的意见，由于对研究问题的描述不够清晰，从初稿来看确实给人感觉逻辑和理论预期不清。本研究前两项分实验(实验一)说明中文读者能够在预视中切分词 $n+1$ ，单向反馈假设或单整体假设都不能切分词 $n+1$ 的字词加工过程，预视中的切分词 $n+1$ 过程更加符合字词交互激活假设的预测。后两项分实验(实验二)可实现两个目的：1 排除实验一的结果是“实验操作影响与加工词 n 相关事件所致”的可能性；2 考察词 n 上字词加工过程符合向上反馈假设、整体假设还是字词交互激活假设的预测？综合四项分实验的结果可以确定中文阅读中的字词加工模式符合交互激活假设的预测。受到审稿专家意见的启发，我们经过反思后发现初稿描述并不能清晰表达文章的研究前提、研究目的、和研究结论。

在本版修改稿中，我们做了相关修改。在“前言”的最后部分列出所有解决的理论问题，并根据实验操作说明实验结果将如何证实结论。在讨论部分，我们根据前言中提出的问题进行了讨论，希望能够澄清整个研究思路。

审稿人 2 意见：作者还需思考如下几个问题：

意见 1. 第一，本研究的目的是考察中文阅读过程中的字、词加工和词切分的时间特性，所以当注视点位置发生改变时，研究者延迟改变词 $n+1$ 的颜色，延迟时间分别为 40ms、120ms、160ms。请在研究中清楚地说明选择这三个延迟时间的依据。

回应：谢谢审稿专家提出的宝贵意见。本版修改中我们增加了，选择 40ms、120ms、160ms 三种延迟时间的依据。具体做法是在“**2.1.5 实验设计**”增加一段内容，说明依据。具体内容如下：

词 $n+1$ 字体改变颜色的延迟时间选择 40/120/160 ms 依据在于：(1)考虑呈现刺激电脑屏幕刷新率以及电脑侦测注视发生时的系统延迟，从电脑接到“改变词 $n+1$ 字体颜色”指令到这个指令完成约 10-15ms 延迟，因此实验的实际延迟时间为 50-55/130-135/170-175 ms。(2)参照视觉信息从视网膜传至脑皮层所需时间为 55 ms 左右(刘志方等人, 2011)，据此最早延迟时间设置为 40 ms 时方能体现“延迟提示词边界线索”目的。(3)根据研究目的，中间点延迟时间若保证该延迟时间下词 $n+1$ 的预视加工正处于或者接近于词汇加工的早期阶段(比如，熟悉性检验或词汇前加工阶段，参照文献：Reichle, Pollatsek, & Rayner, 2006; Engbert, Longtin, & Kliegl, 2002; Engbert, Nuthmann, Richter, & Kliegl, 2005)更能实现预期效果；实验一将中间点延迟时间设为 120 ms 实际提示线索的延迟为 185-190 ms, 这个延迟时间能够保证对词 $n+1$

的预视加工接近于词汇加工的早期阶段。(4)设置的末尾点延迟时间也应该保证该延迟时间点上提示边界能够影响到词汇加工。根据“等差数列”原则，设置末尾点的延迟时间应该为 200 ms(此时实际提示线索的延迟为 265-270 ms)，但此时的预视加工基本结束，这个延迟时间点上提示线索对词切分基本不会产生影响。因而取 120 ms 至 200 ms 的中间值 160 ms 合乎要求。

意见 2第二，本研究的实验材料都是双字词组成的句子，共包括 60 个。为保证被试认真阅读，研究者设计了判断问题，但是为什么只有 27 个实验句子附带判断问题呢？请交待清楚。

回应：实验过程中安排判断题的目的在于考察被试是否认真阅读句子，被试回答问题的正确率在 85% 以上说明其是认真读完句子。正式实验材料总共 60 个句子，给每个句子都安排判断题会导致实验持续时间增长，这种做法没有太大必要，会导致实验效率下降。目前阅读眼动研究领域均采用穿插安置判断题的做法检验被试是否认真阅读句子，带有判断题的实验句子和没有判断题的实验句子被随机呈现，采用对判断题回答的正确率衡量被试阅读理解的认真程度。由于本实验中所采取的句子材料、判断题与我们以前已发表的论文所采用的句子材料、判断题完全相同，在本版修改稿中给出相应的参考文献(实验材料特征的详细描述参见文献：刘志方等人，2011，2013；张智君等人，2012)予以说明。

意见 3第三，中文阅读过程中的词切分问题是这一领域的热点问题，有很多相关研究。但是，本研究的总讨论部分没有结合这些研究成果展开讨论，只是报告了本研究的实验结果及其解释，深刻性和全面性有待提高。例如，本研究基于句子整体水平和兴趣区水平对眼动指标进行数据分析，但是两个水平的实验结果并不一致，本研究没有对此进行解释和讨论。

回应：谢谢审稿专家的意见。文章考察的是中文阅读中词切分问题，对于这个问题目前确实有很多相关研究。在初稿中，我们仅仅根据当前的实验结果进行解释和讨论，没有很好地参照现有其他人的研究结果。为了提高文章的深刻性与全面性，本版修改稿中我们注重检验向上反馈假设、整体假设和交互激活假设的孰对孰错，并结合李兴珊与其合作者的研究、白学军与其合作的研究以及 Yan 与其合作者的研究结果进行讨论。相对于初稿，本版修改稿的总讨论的结构和内容已经做了全面的改观。

对于“本研究基于句子整体水平和兴趣区水平对眼动指标进行数据分析，但是两个水平的实验结果并不一致，本研究没有对此进行解释和讨论”这个问题，我们实际上是主要通过总阅读时间说明研究问题，这是因为词汇识别是阅读理解正常进行的前提保障，实验句子中

的所有词汇均参与实验操控，因此句子总阅读时间(阅读句子所花费的时间)是衡量词汇加工的关键指标；鉴于基于句子整体分析的其他眼动指标在说明研究目的方面的价值不大，本版修改稿删除了(平均注视时间、注视次数和平均眼跳幅度指标结果)。我们仍然采用基于词汇兴趣区域的眼动数据的原因是，这些指标能够反映字词加工过程，但在提示词汇边界的情况下读者容易通过权衡凝视时间、跳读概率、再注视概率和回视次数的眼动策略完成阅读(张智君等人，2012)；因此本研究中采用 5 类基于词兴趣区的眼动指标(首次注视时间、凝视时间、跳读概率、再注视概率和回视次数)作为辅助性证据说明研究问题。

因此，尽管总阅读时间与词兴趣区域内眼动指标并不一致，但由于重点参照的眼动指标应该是“总阅读时间”，因此，这种不一致并不影响实验推论。在初稿中我们并没有对这个问题进行清晰解释说明，这导致审稿专家二的疑问，也可能导致审稿专家一认为实验结果不够可靠的原因之一。在本版修改稿中，我们对这个问题进行详细的交待说明。

意见 4.第四，结果表述方面的问题。在报告统计结果的 P 值时，可以采用报告 P 值的具体数值，也可以都采用大于或小于临界值的方式，尽量避免两种方式混合使用。

回应：谢谢专家意见。考虑到研究各个实验结果特点，本版修改稿中在报告统计结果的 P 值都采用大于或小于临界值(0.05)的方式,仅在边缘显著的情况下报告具体的 P 值。

审稿人 3 意见：本研究采用眼动随动技术考察汉语阅读中字、词加工和词切分的时间特征，通过操纵协助和干扰切分词的延迟时间，分别考察预视切分词 n+1 和切分词 n 的过程。研究发现了汉字加工水平到词汇加工水平的转折特点。该研究范式成熟，研究内容较有意义，研究结论具有创新性，写作思路较清晰，但有以下问题建议作者给出进一步的解释和说明。

意见 1. 建议在文中补充选择 40ms, 120ms 和 160ms 作为三个延迟时间段的理由，并结合本研究的设计提出在这三个延迟时间段条件下的研究假设。

回应：谢谢审稿专家提出的宝贵意见。由于这条意见与第二位审稿专家所提出的第 1 条意见完全相同，因此这里不再累述。

意见 2. 建议在文中充分分析实验一与实验二间的逻辑关系，以帮助读者更好地理解结论得出的过程。

回应：谢谢审稿专家的宝贵意见。初稿中我们对实验一和实验二之间逻辑关系的描述并不清晰，这可能是审稿专家一认为本研究“思路不清，存在逻辑问题”的主要原因。实际上，实验一和实验二间的逻辑关系完全可以回答审稿专家一所提的第1条意见。修改稿中，我们在“3实验二：识别与切分词 n 的时间进程特点”中通过澄清做实验二的目的，说明两项实验间的逻辑关系。

“首先，实验一不能澄清其中的操作是影响字体颜色变化区域内的文字加工所致，还是影响该区域内右侧的处于注视中心区域内文字加工所致，实验二操控注视区域内文字字体改变颜色的延迟时间澄清此问题。基于实验操作特点，实验二主要影响操控注视区域内文字的识别加工；因而如果实验一的操控是通过主要影响注视区域内文字加工所致，那么实验一操控条件对眼动模式影响的基本趋势应该与实验二相同，否则则说明实验一的操作是通过影响字体颜色变化区域内的文字加工所致。

其次，考察切分词 n 中的字、词加工特点。由实验一结果可知整体假设不能符合实际的词切分情况，词切分存在时间进程；如果词切分主要发生在预视加工阶段，那么提示词 n 边界的操作条件将不会出现显著的延迟时间效应(实验二(A))；实验二(B)进一步检验向上反馈假设、整体假设和字词交互激活假设在解释加工词 n 中的合理性。另外，相对于预视词 $n+1$ ，加工词 n 属于较晚期的词汇加工，实验二的两项实验将有助于揭示词切分后期加工特点。”

意见 3. 中文摘要建议精炼。

回应：本版修改稿的文摘进行了精炼，修改后的中文摘要是：“研究采用眼动随动显示技术，通过操控提示正确/错误词边界线索的延迟时间，考察汉语阅读过程中字词加工与词切分的时间特性。实验一发现，提示词 $n+1$ 边界不影响总阅读时间，但积极影响词兴趣区内的眼动数据且存在延迟时间主效应；提示错误的词 $n+1$ 的边界线索对总阅读时间的影响随延迟时间的变化呈现倒“U”型的变化趋势。实验二发现，提示词 n 边界不影响总阅读时间，但消极影响词兴趣区内的眼动数据且无延迟时间主效应；提示错误的词 n 边界对总阅读时间和眼动数据均产生消极的影响，且存在延迟时间的效应。综合两项实验的结果可以推测，单一的向上反馈假设和整体假设都不能全面解释阅读中的字词加工过程，汉字加工与词汇加工间存在多轮交互激活过程。”

意见 4. 建议加强本研究对以往研究和模型的呼应，使本研究意义更为明确。

回应：谢谢审稿专家的宝贵意见。在本版修改稿总讨论的末尾部分，我们增加了本研究对以往研究和模型的呼应。

本研究从两个方面深化白学军等人采用空格提示词汇边界的研究结论：“(1)在较早时间点上提示词 $n+1$ 右侧边界对眼动数据影响与在词间增加空格的效果相同，但通过改变字体颜色提示词 n 的右侧边界则达不到以上结果，说明词间空格是通过提示词 $n+1$ 右侧边界影响眼动过程。(2)较早时间点上提示错误的词 $n+1$ 边界不影响阅读理解，较早时间点上干扰切分词 n 则严重影响阅读，说明错误的词边界线索是通过妨碍词 n 加工影响阅读理解。

另外，本研究结果对现有对词切分的模型、假设方面的启示为：“首先，保证所有汉字可见的情况下，提示错误的词边界线索消极影响词汇加工与阅读理解说明词切分不可能是由仅仅单字被依次识别加工的后果，向上反馈假设和单向分析策略(该策略认定读者是通过识别汉字实现词切分)不能解释中文字词加工过程(吴俊等人, 2008; Wu, Slattery, Pollatsek, & Rayner, 2008)。其次，预视切分词 $n+1$ 不可避免地涉及单字加工过程，但预视加工与注视加工中词汇识别影响汉字加工，说明字词加工间存在交互激活的机制，这个发现与李兴珊等人提出的词切分模型的字词交互激活假设一致(李兴珊等人 2009, 2011)。不过，词汇识别与单字识别层次上存在可能的抑制形式，仍需要深入的研究。”

第二轮

审稿人 1 意见：作者依据三位审稿人的意见，认真、仔细地回答了疑问，修改了论文。建议发表。

审稿人 2 意见：通过对审稿人意见的思考和修改，文章的思路进一步清晰化，意义性也更明确，有一个小问题还需要作者考虑。本研究主要解决的问题是通过操控提示正确/错误的词 $n/n+1$ 边界的延迟时间检验三种理论假设，其研究内容与文章标题是否相符，本研究两个设计是否真正考察到字加工与词切分的时间进程，或者仅考察了中文阅读中字加工和词加工的关系性质。如果研究的重点放在字加工和词切分的时间进程上，那么研究应关注其切分发生的内部进程。

回应：非常感谢上述两位外审专家的肯定。

对于审稿专家二提出的“研究内容与文章标题是否相符”的问题，经过认真领会专家意见后，我们发展这个建议很有价值。文章初始的重点放在描述“字加工和词切分的时间进程”

上,初稿针对字词加工的时间进程的描述是整体性,大体的勾画。在一审中综合考虑三位外审专家的意见后,我们发现,将本研究的重点停留在“字加工和词切分的时间进程”的描述上,会导致文章的理论深度不够。考虑到,实验一和实验二中的分讨论都已经针对“字词加工和词切分的时间进程”进行了描述,因而在“总讨论”再次描述这个问题显得文章不够精炼,且缺乏理论深度。

实际上,通过描述“字加工和词切分的时间进程”还可以说明更为深刻的理论问题,因此,第一次修改稿中,我们将重点转移至通过“描述字词加工和词切分的时间进程”进而验证“向上反馈假设”、“整体假设”和“交互激活假设”的孰优孰劣问题。此时原有的题目显然不能概括修改稿的内容,仔细斟酌后我们将题目修改成《中文阅读中的字词激活模式:提示词边界线索延时效应的启示》,希望能够更好地概括文章的主旨、凸显文章的理论价值。

在提交第二版修改稿之前,我们再次邀请了一位具有心理学博士学位的读者对文章进行了多次挑剔性阅读,并在其提出意见的基础上再次针对论文进行了修改。本版的修改主要表现在三个方面,一、精简内容(尤其是前言和讨论中的内容);二、优化前言和讨论中各个段落之间的逻辑;三、优化段落内部句子间的逻辑结构。其中修改的具体地方如下:

将第一版修改稿前言中的第二段,进行了概括,并将之与第一段进行了合并。通过如此改动,以便概括提出“中文阅读理解过程符合普遍的语言学规律:以词汇为基本单元理解中文文本”。2. 第一版修改稿前言中的第三段在本版修改稿中变为“第二段”,我们对本段内容也进行了修改,改变的原段中句子间的逻辑结构,以便在引出“中文阅读理解中存在特殊性”的基础上,从整体上清晰自然引出本研究的目的。3. 本版修改稿从第三段开始,首先,引出研究的具体理论假设,然后将这些需要验证的理论问题,细化分布在预视加工和注视加工中词切分过程当中。针对预视中的词切分问题(本版修改稿的第四段),我们将现有的研究成果,进行了分层,提炼,以形成达到逐步深入的逻辑顺序。4. 在总讨论中,也按照“前言”所呈现的逻辑结构,对实验结果进行讨论、简化提炼内容和逐步深化了理论问题。

总的来说,第二版的修改稿,也进行了细致的修改,其中修改过的内容都用蓝色字体标示出来。请专家审查。

第三轮

审稿人1意见:作者根据前面数位审稿人的意见对文章进行了修改,研究结果可靠、有理论意义。但还存在一些细节问题,建议修后发表。1. 标题等中文信息修改后,英文标题、关键词等没有同步更新。2. 请作者认真仔细阅读全文,修整部分错漏或错误字句:如21页第

一段“最后，从词切分本身角度来看，Yan 与其合作者首次注视点的位置曲线推测，中文读者能够根据预视中词切分(Yan, Kliegl, Shu, Pan, & Zhou, 2010; Shu, Zhou, Yan, & Kliegl, 2011; Yan, Zhou, Shu, & Kliegl, 2015); ” 3.丢字现象：31 页第一行，“实验一”的“实”缺失。

回应：谢谢专家的肯定。本版修改稿已经根据专家的意见做了细致的修改。本版修改稿件的英文标题、摘要内容和关键词都针对中文内容做了修改。考虑到作者的英语水平有限，本篇论文若能被接受，我们将花钱请专业公司针对英文摘要进行翻译和母语化润色。对于文中错漏或错误字句，我们认真通过全文后进行了细致的修改。在本版修改稿中，我们继续精简语句，以增加文章的可读性。所有被修改之处都用绿色字体标示出来，请专家审查。

再次，感谢专家提出宝贵意见。

第四轮

编委意见：文章存在下述问题，请修改后复审。

意见 1.数据分析繁多，使人读后理不清头绪。建议方差分析只报告显著的数据，不显著的一笔带过，且方差分析时不加小标题，效应量符号的写法也不正确，应在符号加下标 p 。

回应：谢谢专家意见。我们起先考虑通过提供全面的眼动数据指标，相互印证系统地说明研究主旨。基于专家的提示，原稿中提供的五种基于词汇兴趣区域的眼动指标(平均凝视时间、平均首次注视时间、跳读概率、再注视概率和回视次数)确实提供了冗余信息。考虑到平均凝视时间、平均首次注视时间和再注视概率所反映的认知过程基本较为一致，根据它们对差异敏感性的特点，本版修改稿保留“平均凝视时间”指标，这样本版修改稿中的因变量采用 4 项指标：总阅读时间、平均凝视时间、跳读概率和回视次数。

同样，参照专家的意见，两项实验的结果分析部分中只报告了显著的数据，不显著的放置在靠前位置，并一笔带过。由于本研究采用被试分析和项目分析两类分析方法，为了标识两种分析方法，并参照以往的阅读眼动研究领域内的惯例，我们也保留“ F_1 ”(代表被试分析)和“ F_2 ”(代表项目分析)方式呈现结果。

在前一修改稿中，我们参照文献“郑昊敏，温忠麟，吴艳.(2011).心理学常用效应量的选用与分析. 心理科学进展 2011, 19 (12), 1868–1878.”用“ η^2 ”表示效应量(心理学报 2014 年中的多篇文章也是采用这种表示方式)。在本版修改稿中，我们参照专家的要求用“ η_p^2 ”表示效应量。请专家审查。

意见 2.推论繁多，大量的如果.....就.....，可能一、可能二、请简洁地写推论过程。

回应：本版修改稿中，我们已经根据专家的意见，简洁地写出了推论过程，并对整个论文进行了精简工作，请专家审查。

意见 3.27 页被试分析不显著、项目分析显著基本上可以认为没有效应，不用再作详细分析。

回应：谢谢专家意见。上版修改稿的 27 页码实验一(A)的平均凝视时间指标的方差分析在被试分析上不显著，在项目分析上显著，基于研究目的和数据特点这两方面的考虑，我们对之进行了事后检验。

首先，以往研究发现，**及时**提示词 $n+1$ 的右侧边界能够减少词兴趣区域内的平均凝视时间，根据这个结果，实验一(A)继续探讨切分词 $n+1$ 的时间进程，因此细致探讨**各延迟时间点上**提示词 $n+1$ 右侧边界对平均凝视时间的影响仍具有方法学价值。

两个原因延迟时间效应在被试分析上不显著，在项目分析上显著：(1)由图 1 可见，三种延迟时间(40/120/160 ms)的提示词 $n+1$ 右侧边界下平均凝视时间，有两种条件与控制条件间基本相似，这种均值差异模式应该是导致被试分析中没有延迟时间主效应的原因；另外(2)被试分析的统计力(statistic power)低于项目分析(被试数为 24，而句子数则为 80)也是导致如此结果的原因。

基于上述考虑，若仅仅基于呈现条件主效应在被试分析上不显著，就放弃详细分析，会淹没“发现中文读者可以预视切分词 $n+1$ ”的可能性，而这个事实已经得到多项证据支持(Gu & Li, 2015; Gu & Li, Liversedge, 2015; Yan et al., 2010, 2015; Shu et al., 2011; 张智君等, 2012; 刘志方等, 2013)，因而本版修改稿仍然保留平均凝视时间指标上的事后检验(事后检验也是基于被试分析)，检验发现相对于控制条件，40ms 延迟时间中提示词 $n+1$ 右边边界能够减少词汇兴趣区域内的平均凝视时间，说明及时提示词 $n+1$ 的右侧边界能够促进词切分。

意见 4.本研究想解决的问题太多，用的分析方法、分析指标也太多，显得很乱，请精简，如非必要，可以省略。

回应：谢谢专家意见。实际上本研究所要解决问题是：中文阅读中的字词激活模式问题。基于阅读中的眼动特点，研究中的两项实验通过分别检验预视阶段和注视阶段中的字词激活模式，以解决这个问题；前一修改稿基于这种区分试图以“分别提出两个问题”的方式描述研究所要解决的问题。经专家的提示，我们发现这种做法会误导读者。通过将原稿中的倒数第二段和倒数第三段进行整合，本版修改稿从整体上提出研究问题。为实现其目的，实验一(B)

提供了补充分析，补充分析方法是非常必要的。不过，参照专家的第一条意见，我们对所有分实验的分析指标和分析结果进行了精简。请专家审查。

本版修改稿中修改内容用紫色字体标示，再次，感谢专家提出宝贵意见。

第五轮

主编终审意见：我仔细阅读了稿件《中文阅读中的字词激活模式：提示词边界延时效应证据》(xb15-131)及其评审过程，同意评审编委和评审专家的意见，同意发表。关于文章的题目，中文题目“中文阅读中的字词激活模式：提示词边界延时效应证据”中的“提示词边界延时效应证据”好像有些表达不清楚，参照英文题目 The activation model of characters to word in Chinese reading:evidence from the delay word boundary effects，是否用“来自提示词边界延时效应的证据”更清楚？请作者考虑。

回应：谢谢主编提出宝贵意见，经过认真思考过后，我们觉得主编提出的题目修改建议更清楚，因此我们将这篇稿件的题目修改成为《中文阅读中的字词激活模式：来自提示词边界延时效应的证据》