

《心理学报》审稿意见与作者回应

题目：语境预测性对阅读中字词加工过程的影响：眼动证据

作者：刘志方; 全文; 张智君; 赵亚军

第一轮

审稿人 1 意见：

研究者采用眼动追踪技术探讨语境预测性对阅读中字词加工过程的影响，结果发现，语境预测性与整词频率和尾字频率存在交互作用，但是跟首字频率之间无交互作用。语境预测性作为影响阅读的一个重要属性，在很早就得到研究者的广泛关注，作者通过一系列的三个研究，以对其加工特点进行说明，选题很有意义。但是本研究仍存在以下问题。

意见 1：根据自检报告中的描述，作者并未提供效应量(effect sizes)，“当前研究结果使用 R 语言进行线性混合模型分析，限于该方法中效果量的计算还是一个难题，故未报告。”而事实上，效应量的数值计算已经成为很多杂志投稿的必备条件，如 *Journal of Experimental Psychology: Human Perception and Performance*，这也是心理学报将该要求放入自检报告的原因，所以希望作者能在修改稿中予以补充。

回应：效应量是指由于因素引起的差别，是衡量处理效应大小的指标，效应量指标不受样本容量影响。它表示不同处理下的总体均值之间差异的大小，可以在不同研究之间进行比较。本研究基于线性混合模型分析因变量数据，在 R 环境中使用 lme4 统计软件包进行分析，因而计算变量及其交互作用的效应量，应该在具体的线性混合模型中得以进行。目前为止，我们尚未看到在“线性混合模型中计算效应量”的统计思路和相关代码，但是我们提供了 95% 区间值，希望以此弥补效应量缺失的不足。由于我们并非统计专家，也不善于撰写电脑代码，如果专家有更好手段，还望不吝指导。

意见 2：同样，自检报告还要求读者计算样本量，以体现本研究的 power effect，但是作者并未提供。通过看文章可知，作者进行了大量的工作，每个实验的被试数均超过了 200 名，确实具有足够说服力，但是鉴于中文阅读中的高跳读率，所以还是按照杂志要求，根据相关前期研究文献来计算样本量，会有更大的说服力。

回应：本研究预期发现，语境预测性与频率变量交互影响中文词汇识别。以往多数研究并未发现语境预测性与词频交互作用显著。语境预测性与频率变量交互作用本身效应量小可能是

未有发现阳性结果的潜在原因,因而我们希望通过扩大样本量发现其存在的证据。参照审稿专家的意见,我们使用 G Power 软件计算了实验一(涉及语境预测性与词频变量间的交互作用问题)的所需样本量。计算样本所需量之前需要大致确定语境预测性与词频间交互作用的效应量,梳理已有文献可见语境预测性与词频的交互作用效应量均低于 0.1(卢张龙等,2008),据此计算得出实验一所需样本量最少为 282 人,为得到稳定结果是我们进一步补充的实验一被试数量,最终被试数量为 322 人。对于后两项实验(实验二和实验三)所需被试量而言,其被试数量已经接近 320 人,目前为止,我们尚没有看到语境预测性与字频变量交互作用的以往研究数据,但参照实验一的样本量计算值,实验二和实验三的样本数量可以接受,因而后续两项实验,未增加样本量。

意见 3: 在本篇文章中,有几个非常重要的概念,比如作者在研究目的中提出“研究包含三项实验,通过观察语境预测性与整词词频,词内汉字字频间交互作用,以探讨阅读中语境预测性影响词汇识别中的具体加工环节问题。”,在这里的“加工环节”是否有具体清晰的概念或者操作定义?根据作者的描述,可推知将加工环节区分为字的加工环节和词的加工环节,但是,字和词的加工环节是否有明确的区分?如何区分?作者是否可以给出明确的说明,并指出区分的依据。这样做会让研究的目的更清晰。

回应: 本文研究主题围绕“语境预测性所影响词汇识别的具体环节”。关于识别词汇的具体加工环节,不同的理论框架下有不同的解释和假说。比如,有研究将词汇加工看成 3 个阶段:前词汇加工(prelexical processing),词汇通达(lexical access),后词汇加工(post-lexical processing)(Forster, 1981; Fodor, 1983),后词汇加工被看作是与“将词汇整合到语句背景,以实现句子理解”有关。以往研究发现,语境预测性影响对目标词汇的语义整合过程,也即是后词汇加工。本研究关注的问题是,语境预测性是否影响较为早期的两个加工阶段:前词汇加工(prelexical processing),词汇通达(lexical access)。词汇通达常常以“是否出现词频效应”为参照标准。前词汇加工中包含多种信息加工成分,比如,视觉信息、正字法信息和语音信息的获得。E-Z 读者模型和 SWIFT 模型都将词汇识别划分成两个阶段,这两个阶段分别于前词汇加工、词汇通达有一定重合。不同的词汇加工阶段划分并不矛盾。

本研究结合中文词汇的书写和识别特点,将中文词汇识别区分成:字的加工和词的加工两个阶段(环节),这种划分是参照以往实证研究和理论假设。参照前人构建的理论模型和以及实证研究(Yan, Tian, Bai, & Rayner, 2006; Li, Rayner, & Cave, 2009; Li, Bicknell, Liu, Wei, & Rayner, 2014; 李兴珊,刘萍萍,马国杰,2011),我们以字频效应(字频变量对词上眼动数据的

影响)和词频效应(词频变量对词上眼动数据的影响)分别代表这两个加工环节。关于“加工环节的清晰鉴定”，我们认为，将词汇识别过程区分成字的加工和词的加工两个加工环节，仅是一种大致上的划分；目前尚不能做出让审稿专家满意的、具体明确清晰的概念或者操作定义，原因如下：首先从现有考察中文阅读中字词加工的实证研究来看，研究多以字频效应和词频效应代表字的加工环节和词的加工环节。其次，现有专门解释中文阅读词汇加工的理论模型并没有给出字的加工和词的加工的具体概念和操作定义，解释拼音文字阅读词汇加工的模式对本研究的借鉴意义不大(在词汇加工划分环节的概念和具体操作性定义问题上的借鉴意义不大)；识别拼音文字词汇中的字母完全受制于整词加工，识别中文词汇中的汉字则有不同的特点。再次，字的加工和词的加工的区别性还需深入研究，研究显示对于双字词而言，首字加工在很大程度上受制于整词的加工，而后字的加工则有一定的独立性，由此可知，字的加工(尤其是双字词中的首字)并不完全脱离于词的加工，将字的加工与词的加工完全区分/割裂出来并不现实。

更为重要的是，我们考察各层加工间的交互激活过程。考察语境加工与词的加工、字的加工间的交互作用，从而为理解中文阅读中的词汇识别机制提供一个视角。细致区分字的加工和词的加工的具体概念和操作性意义，对于澄清研究主题的作用不大。当然，随着研究的深入，给出字的加工和词的加工的具体概念和操作性意义，并在这个基础上深入探讨相关的研究主题，是有必要的。这是我们后续研究的努力方向。

意见 4：作者在前言的最后讲到：“研究语境预测性与频率变量间的交互作用将有利于检验认知模块理论与交互激活理论，前者认为各认知模块独立运行，故两因素间交互作用不显著(Fodor, 1983; Forster, 1979)；后者假定各层面认知相互影响，故预测两因素间交互作用显著(Morton, 1969; McClelland, 1987)。”由此可知，本研究的结果有着重要的意义，可以为两种理论的争论提供参考，这应该是本研究很重要的内容，甚至是意义所在。但是，这两个理论的具体内容作者缺乏相关的描述和介绍，建议补充。

回应：从大的研究方向来看，本研究的主题确实与“认知模块理论和交互激活理论之间的争论”有关。不过，我们的具体研究主题是，语境预测性与频率变量是否交互影响中文阅读背景中的词汇识别过程。在阅读研究领域，已经产生多个理论模型用于解释阅读中的词汇加工和眼动控制过程，这些理论模型在“语境预测性和词频变量如何影响词汇识别和眼动控制过程”问题上的假设，并不一致。我们是在中文阅读背景中检验“语境预测性和频率变量交互影响词汇识别”假设。解决认知模块理论与交互激活理论之间争论，有非常重要的理论价值，

考虑到认知模块理论与交互激活理论可能并非完全对立，而“考察中文阅读中语境预测性和频率变量交互影响词汇识别问题”也不足以解决“认知模块理论和交互激活理论之间的争论”。综上考虑，本版修改稿，主要参照具体研究主题(语境预测性和频率变量交互影响词汇识别问题)对引言进行重新组织和修改。在总讨论的末尾部分，参照认知模块理论和交互激活理论对研究结果和主题进行深化讨论和扩展，以增加论文的理论深度。

如果专家还有其他更好的论文撰写思想和思路，还望不吝指点。

意见 5: 根据作者在前言中的描述，已有英文研究和中文研究中均未发现词频和语境预测性之间的交互作用，所以作者对研究结果进行预期时，为何会认为存在该交互作用？即便结果如此，可是在前言中进行不同于以往研究的预期，会有点逻辑上的不通顺。“多项研究一致发现，在各项注视时间指标……上，语境预测性与词频之间的交互作用并不显著(Altarriba, Kroll, Sholl, & Rayner, 1996; Ashby et al., 2005; Rayner et al., 2001; Rayner, Ashby, Pollatsek, & Reichle, 2004; Kliegl, Grabner, Rolfs, & Engbert, 2004; Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006; Mielliet, Sparrow, & Sereno, 2007; Gollan, Slattery, Goldenberg, Van Assche, Duyck, & Rayner, 2011)。”“语境预测性与词频之间的交互作用在各项注视时间(首次注视时间、凝视时间和总注视时间)上均没有达到显著水平(卢张龙, 白学军, 闫国利, 2008)。而对于两者间交互作用是否在其他眼动指标上显著，则缺乏相关证据。”同样，当作者得到跟之前研究不同的结果时，需要进行原因解释，但是作者的解释需要更深入和恰当的依据，而不能仅仅是推测。如，将本研究与已有英文研究的不同归为语言的属性差异，“成年正常中文读者与拼音文字成年读者存在差异，应与中文书面语言认知的语言特殊性有关”；将本研究与已有中文研究的不同归为以前研究检验力较弱，“考虑到，检测不到并不代表不存在(Altman & Bland, 1995; Cervero, & Laird, 2000)。阅读中语境预测性与词频交互作用本身可能较难捕捉(Hand et al., 2010)，统计检验力较弱时，语境预测性与词频交互作用更不容易被观察。以往研究中所使用的实验材料和被试取样数量均不多(Rayner et al., 2004; 卢张龙等, 2008)，统计检验力较弱，以及分析的眼动指标有限，可能是导致语境预测性与词频交互作用不显著的重要原因”。作者应说明，不同语言之间何种的差异会导致出现不同的研究结果；作者应检验之前研究的检验力数值到底是多少，以支持其统计检验力较弱的推测，这样做会更有说服力。

回应: 原稿在第三段的开头直接指出“词频和语境预测性之间的交互作用不显著”，这种结果主要是在正常阅读技能的成年读者得到的。实际上，低阅读技能成年和失语症英语读者的语

境预测性与词频交互作用则达到显著水平(Ashby et al., 2005; Huck, Thompson, Cruice & Marshall, 2017), 因此“词频和语境预测性之间的交互作用不显著”的描述并不准确, 本次修改稿将第三段的开头改写成“在追踪读者眼动轨迹的自然阅读中(主要是拼音文字阅读), 研究者发现语境预测性与词频之间的交互作用并不明朗。多项研究发现, 健康成年读者的语境预测性与词频交互作用不显著, 但也有证据显示, 低阅读技能成年和失语症英语读者的语境预测性与词频交互作用达到显著水平。”

原稿第四段通过卢张龙等人(2008)的研究介绍中文阅读中“语境预测性与词频交互影响词汇识别”的研究结论。鉴于该研究并未发现语境预测性与词频交互作用显著, 不过对待这种结果需要慎重。这是因为当交互作用效应量较小时, 境预测性与词频间交互就难以发现, 容易犯“假阴性”错误。分析卢张龙等人(2008)的实验数据可以发现, 其中语境预测性与词频间交互作用的效应量都小于 0.1, 该研究中所使用的实验材料和被试取样数量均不多, 这种情况下语境预测性与词频间的交互作用较难捕捉。所以本研究拟通过增加被试数量, 进而继续探讨和确定语境预测性与词频交互作用显著的可能性, 及其不显著的稳定性。

本次修改稿, 针对前言进行了重新组织, 尽可能多地摒弃了像“成年正常中文读者与拼音文字成年读者存在差异, 应与中文书面语言认知的语言特殊性有关”之类的武断推测。若还有其他方面的问题, 请专家继续批评指正。

意见 6: 作者在实验二、三中, 分别讨论了双字词的语境预测性和首字、尾字字频的交互作用。在这里需要说明为什么采用的预测性是双字词的属性, 而采用的字频是单个汉字的属性, 这两种属性分属于不同的客体, 对分属于不同客体的属性之间的交互作用进行探讨是否可行? 请作者说明目的和依据, 并予以解释。

回应: 实验一操控的是双字目标词汇的语境预测性属性和整词词频属性。实验二和实验三操控的是双字目标词汇的语境预测性属性和目标词汇的首字字频属性/尾字字频属性。如果将目标词汇和其内组成汉字分别看成是不同客体, 那么可以说“预测性是双字词的属性, 而采用的字频是单个汉字的属性, 这两种属性分属于不同的客体”。不过, 本研究将目标词汇看作是一个整体(这个整体是由两个汉字组成), 组成目标词汇的两个汉字隶属于目标词汇这个整体; 因而其首字字频与尾字字频实际上也是目标词汇的属性, 只不过这个属性与整词词频属性是不同层次的属性而已。

本研究探讨的交互作用的实质是不同加工层面之间的交互激活。以双字目标词汇的语境预测性属性和整词词频属性(实验一操控的变量)为例: 语境预测性不是单个目标词汇的属

性，它的存在是以语境为前提；词频变量是目标词汇的整体属性。语境预测性属性在概念层级上高于整词词频属性，实验一考察不同层级属性变量(两个相邻层级变量)之间的交互作用。目标词汇的首字字频和尾字字频也是目标词汇的属性，只不过这两个属性在概念层级上低于整词词频属性，有研究考察过整词词频属性变量与字频属性变量间(这两个变量也是相邻层级属性变量)的交互作用(Yan, Tian, Bai, & Rayner, 2006)。实验二和实验三考察语境预测性属性变量和字频变量间的交互作用，这两个属性不是相邻层面的属性，前者高于后者 2 个层次(它们之间间隔词汇整体上的属性层)，研究考察两个不相邻上下加工层面之间的交互作用，这种做法在以往并不多见，它很大程度推进以往研究和理论。

考虑到，原稿的表述方式不易理解，本版修改稿着重强调首字字频和尾字字频都是目标词汇的首字字频和尾字字频。

意见 7: 本研究用了大量的篇幅进行了讨论，但是在前言中缺乏相关理论的解释和对结果的预测，因此导致对讨论内容理解的困难。比如，“英语阅读中语境预测性非常稳定地影响跳读概率(Balota et al., 1985; Schustack et al., 1987)，其对跳读概率的影响程度大于词频变量.....实验一同样发现中文阅读中的语境预测性效应与词频效应之间存在这种差异，由此可见，词频因素影响词汇加工在时间点晚于语境预测性因素在一定程度上是个跨语言的规律。”虽然如作者所述，语境预测性对跳读概率的影响程度大于词频变量，但是为什么认为词频因素影响词汇加工在时间点晚于语境预测性？一个是效应大小的问题，一个是发生时间早晚的问题，所以推论的依据是什么，还请解释。

回应: 原稿总讨论篇幅较多的主要目的是，从多个方面论述三项实验结果如何解决研究主题，我们通过讨论/说明多项眼动指标统计结果的心理学意义，从而多方面/多次层次地论述研究主题。阅读眼动研究的特点是，眼动追踪技术会提供多种因变量指标，这些眼动指标所代表的心理学意义有所重叠，但各个眼动指标分别侧重反映特定的心理意义(或者侧重于心理过程的特定阶段)。在本研究中，跳读概率仅是采用的多项因变量指标其中的一项，在引言中，我们并没有直接的、仅针对跳读概率因变量的假设预期，研究参照多项因变量指标统计结果，推测验证研究主题。针对所有因变量指标做预期假设并不现实，且这样做的话不利于读者把握研究主题。

目标词汇被注视之前，读者会对目标词汇进行预视加工，研究表明，读者会根据预视加工获取的目标词汇的部分信息决定是否跳读该目标词汇(此时目标词汇并未被完全识别)，对于特定词汇而言，预视获得其词汇信息的相对程度越高，其被跳读的概率就越大。语境预测

性影响目标词汇的跳读概率,说明语境预测性变量会影响对目标词汇的预视加工。词频不影响跳读概率,说明词频变量不会影响对目标词汇的预视加工。另有证据表明,当目标词汇被注视的时候,词频变量会影响对之注视的时间,当目标词汇处于被预视阶段时(注视点位于目标词汇左侧词汇上)词频变量则不影响注视时间。考虑到,对特定词汇的预视加工,在时间点上早于对其注视时的加工,因此可以根据语境预测性和词频变量影响跳读概率的效应差异,推测两个自变量影响词汇识别过程的相对发生时间点的顺序:语境预测性变量影响词汇加工的时间点,早于词频变量开始影响词汇加工的时间点。语境预测性变量和词频变量开始影响词汇加工的时间点不同,语境预测性与词频间交互作用不显著,词频效应是否呈阳性是“对词汇的加工程度是否达到词汇水平”的标准,由此可知,语境预测性并不直接影响文字加工中词的加工阶段。

原稿中的描述中方式可能会误导审稿专家和读者,因此我们针对相应描述进行了修改。

意见 8: 如前所述,作者对一些关键的概念缺乏恰当和准确的描述。如,“切词”这个概念来源于哪篇文献? Li et al.,已发表多篇文献均使用的是“词切分”,所以还请作者对关键概念进行精确使用,并标注参考文献。

回应: 本版修改稿中,“切词”概念一律修改成“词切分”,以便与以往研究一致。

意见 9: 建议作者在介绍 R 分析时,说明固定变量和随机变量。另建议,将统计分析结果列入表格,以便行文更加清晰。

回应: 本版修改稿详细说明了固定变量、随机变量、协变量和 slope 问题处理。分析模型中并未包括 slope 项,因为加入该项后,模型将不能收敛。本版修改稿中,将统计分析结果统一列入表格中。

意见 10: 文章的表述需要仔细修改,多处行文不通顺或者有误,不一一列举:“预测性是读者常利用语境线索之一。”;“表后跟随各项指标的统计分析结果,分析结果报告报告回归系数 b、标准误 SE 以及 t 值($t=b/SE$),”;文后材料举例也有错误;英文摘要同样需要修改,不需等文章接收后再修改。

回应: 本版修改稿将各处的书写错误进行全面修改。

.....
审稿人 2 意见:

该论文通过三个实验,系统探讨了中文阅读中语境预测性和词频及首、尾字字频之间的

关系。结果发现，语境预测性和词频以及尾字字频之间存在交互作用，表明语境预测性同时影响汉语词汇识别的字、词加工环节。研究成果对汉语阅读的认知加工机制有较大启示作用。如下建议供作者考虑：

意见 1： 研究者在讨论最后一部分论述了本研究结果对汉语阅读眼动控制理论的启示。建议作者在引言部分，尝试从理论争论的视角出发提出问题(而不单单是对以往文献的回顾与总结)，提升本论文的理论价值。

回应： 本版修改稿参照专家意见重新组织了引言部分。本研究通过探讨语境预测性与频率变量间的交互作用，探讨语境预测性影响中文词汇识别的词汇加工阶段问题。为引出研究主题，修改后引言基本思路的铺垫过程是：首先，在总结以往研究结果的基础上，总结提出语境预测性影响词汇识别问题(第 1-2 段)；其次，结合语境预测性与词频变量间的交互作用，引出几个眼动控制模型，及其在“语境预测性与词频是否交互影响词汇加工方面”的假设差异(第 3-5 段)，这些差异与本研究的主题密切相关；最后，结合中文特点，提出具体研究假设。总的来说，本研究的主题主要澄清阅读中语境预测性所影响中文阅读中词汇识别的具体加工阶段问题，在前言中放弃了解决“认知模块理论和交互激活理论之间的争论”的目的(这个问题是个非常普遍，更为深刻的理论问题，我们认为单单通过本研究的三个实验尚不足以解决)。不过，为加深研究的理论深度，我们在组后总讨论部分，根据本实验结果就“认知模块理论和交互激活理论”的基本观点进行了讨论。

意见 2： 作者在“引言”部分论述到，在拼音文字阅读中，起跳距离(launch site distance)影响语境预测性和词频之间的作用方式，当起跳距离较短时(副中央凹预视程度相对较大)，预测性和词频的交互作用开始出现。既然起跳距离是一个潜在变量，建议作者在结果分析时，将起跳距离作为一个协变量纳入到模型中。再或者，鉴于每个实验的被试数量较多(超过 200)，在数据量足够充分的前提下，作者可以考虑将起跳距离作为一个变量(分为较近和较远两个水平)，进一步挖掘数据。

回应： 鉴于本研究的主要目的是考察语境预测性与频率变量间的交互作用，因而我们参照专家提议，将起跳距离作为协变量纳入分析模型。考虑到，注视位置也会影响注视点的持续时间，所以本版修改稿针对数据进行了重新分析，分析模型中除了包含固定因子和随机变量外，还包含“起跳距离”和“首次注视位置”两个协变量。我们希望以此尽可能地剥离无关变量的影响，从而获得更为纯净的实验效应。

意见 3: 在做阅读眼动研究时，一般会对不同实验条件下阅读材料的通顺性和合理性进行匹配，尤其是在操纵语境预测性变量时，低预测语境的阅读语句一般看起来不是很合理，也不是很通顺。因此，作者非常有必要提供阅读材料通顺性和合理性的评定数据。

回应: 编制实验材料时，我们考虑到了低预测语境语句的合理性和通顺性问题，参照严格的标准最终从近 300 个句子中选择 40 个左右的合适句子。根据专家指示，修稿成稿之前邀请 20 名大学生完成对实验材料的评定，总的来说，实验材料的通顺性和合理性符合标准。本版修改稿增添了评定结果(见“**实验材料**”标题下内容)。

意见 4: 在结果部分，作者应首先报告阅读理解正确率，以保证读者都认真阅读并且理解了句子所表达的含义。在结果呈现之前，应报告所使用的模型，以及如何处理 slope 这些关键信息。

回应: 本版修改稿在报告结果之前，增加了阅读理解的正确率，也报告了使用的模型。

意见 5: 在结果部分，作者只是简单列举了每个因变量的分析结果，但并未在结果部分对每个结果说明了什么进行简单总结。由于结果很多，读者很难在很短时间内捕捉出本研究最为核心的发现。为了增加文章的可读性，建议作者：(1)以列表的形式总结每一个实验的分析结果；(2)在结果陈述时，可以对同样的效应进行总述。例如，“语境预测性的主效应在首次注视时间、凝视时间和总注视时间这三个时间类眼动指标达到显著。”

回应: 本版修改稿已经参照专家的意见进行了修改：将统计结果放置在表格中呈现，结果陈述时，对同样的效应进行了总述。

意见 6: 在结论部分，在呈现完主要结果之后，应写明研究结果表明了什么，或支持了什么。

回应: 参照以上意见，本版修改稿将结论部分内容改成“中文阅读中，语境预测性与尾字字频变量间的交互作用显著，说明语境预测性对词汇通达的影响，至少起始于字的加工环节。”

意见 7: 此外，本文存在很多处笔误及语句不通顺的地方，建议找本文其他作者严格校对一遍。主要体现在：第一，p15 第一段前两行，单看 t 值和 p 值来说，语境预测性效应不是边缘显著，而是显著($t = 6.916, p < 0.001$)；而尾字字频效应是边缘显著而不是显著($t = 1.938, p = 0.053$)。建议作者仔细核对每一个统计检验结果。第二，参考文献格式不规范，例如，第二条参考文献页码没有写全；第三条参考文献期刊名称后面应将“.”改为“;”。其他类似错误也

有很多，建议作者核对每一条参考文献。第三，语句不通顺处也有很多。例如，p4 第二行“语境预测性与词频交互目标词汇上的凝视时间”；p6 第一行“主要参加整词词频和单字字频”；p17 第 12 行，“高语境预测性会在削弱词频效应”；p19 第 8 行“在种趋势在后续注视时间...”。

回应：本版修改稿将各处的书写错误进行全面修改。

总回应：非常感谢两位审稿专家提出宝贵意见。为了稳定实验结果，参照审稿专家的提示，我们计算了所需样本量，并根据计算结果增加了实验一的被试数量。随后，为获得更加纯净的实验效应，我们重新分析了三项实验的结果，将起跳位置和首次注视位置作为协变量纳入分析模型。新的分析结果显示：实验一的语境预测性与词频变量间交互作用不再显著，同样实验二的语境预测性与首字字频变量间交互作用仍不显著。为了确定交互作用不显著的可靠性，针对实验一和实验二的注视时间因变量进行了贝叶斯分析，计算得到的贝叶斯因子结果支持原假设，即认为语境预测性与词频变量、首字字频变量间的交互作用不显著。我们根据新的实验结果，并参照两位审稿专家的意见，针对论文的前言和总讨论进行了大篇幅的修改，修改过的内容以蓝色字体呈现。

总的来说，本次修改的内容较多，如果还有其他方面的问题，还需专家继续批评指正。

第二轮

审稿人 1 意见：

意见 1：通过修改稿可以看出作者在原稿基础上进行了大量的工作，在上一稿中，我提出让作者提供本研究的 effect sizes/power effect，作者原本的每个实验至少 200 人，但是作者现将被试数增加至“为得到稳定结果是我们进一步补充的实验一被试数量，最终被试数量为 322 人”。其实本来 200 名被试我已觉得足够，只是缺乏杂志在自检报告中所要求的 effect size，如果杂志认为可行，我无异议，也不希望作者再增加被试。感谢作者的辛苦工作。但是还有以下问题请作者思考。

回应：感谢专家的意见。我们增加被试数量是基于如下考虑：

本研究涉及语境预测性和频率(词频和字频)类两种变量。语境预测性效应和词频效应是两个非常强烈的效应，它们广泛存在于各种词汇识别类任务中。两个变量影响词汇加工过程的方式不同。综合以往研究可知，语境预测性效应和频率效应量都比较大，若仅关注两个主效应的话，少量被试便可达到要求。但本研究重点关注是，语境预测性与频率变量之间的交

互作用。关于语境预测性与频率变量之间是否交互影响自然阅读中的词汇加工，相关的研究结果并不一致：有研究发现交互作用显著(Ashby, Rayner, & Clifton, 2005; Huck, Thompson, Cruice & Marshall, 2017; Hand, Mielliet, O'Donnell, & Sereno, 2010)，也有研究发现两者间交互作用不显著(Altaribba, Kroll, Sholl, & Rayner, 1996; Rayner, Ashby, Pollatsek, & Reichle, 2004; Kliegl, Grabner, Rolfs, & Engbert, 2004; Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006; Mielliet, Sparrow, & Sereno, 2007; Gollan, Slattery, Goldenberg, Van Assche, Duyck, & Rayner, 2011)。检测不到的效应并不代表它不存在(Altman & Bland, 1995; Cervero, & Laird, 2000)，这是因为当效应量本身很小时，这个效应就很容易被淹没在整体误差当中，此时通过扩大样本量方有可能检测到它。中文阅读中有 ERP 研究表明，语境预测性与词频交互影响 N1 波幅(Lee, Liu & Tsai, 2012)；但自然阅读中，语境预测性与词频交互影响眼动过程的作用则没有达到显著水平(卢张龙等人, 2008)。导致这种差异的重要的可能原因是，ERP 实验任务严格控制每个词汇的呈现时间，这样做能够很好限制随机误差，此时交互作用容易显现出来；但自然阅读中，读者注视特定词汇多少时间是较难控制的，随机误差较大时，实验效应就较难检测出来，此时增加被试量则可补救这种局限。综合以往研究可知，倘若语境预测性与词频交互影响词汇加工，其交互作用的统计效应量也是比较小的。由于本研究采用眼动技术，所以根据卢张龙等人(2008)的实验数据计算所需样本量较大，因此我们再次增加了样本数量以达到相应要求。关于实验二和实验三的被试量计算问题，考虑到尚查阅不到语境预测性与字频变量间交互作用的文献，故这两个实验的被试量选择也参照实验一。另外，增加被试数量后，我们参照专家的意见，控制了相应协变量进行计算。交互作用不显著的情况下，提供了贝叶斯分析。

意见 2：如作者所述“字的加工与词的加工完全区分/割裂出来并不现实。”所以我在上一稿中提出，“将加工环节区分为字的加工环节和词的加工环节”，这样是否合适？这样的区分到底是根据加工时程进行的区分还是加工机制的区分？

回应：从加工时程上看，视觉词汇加工大致区分为前词汇加工、词汇通达和后词汇加工三个阶段(Forster, 1981; Fodor, 1983)。各个时段内词汇加工的内容各不相同，从普遍意义上讲，前词汇加工包括视觉加工、词形知觉、语音/正字法提取等加工成分，词汇通达是词汇语义激活，此时读者完全识别目标词汇，后词汇加工包含词汇语义选择和语义整合(Forster, 1981; Lee, Liu & Tsai, 2012)。词频是词汇水平上的变量，词频效应显现是词汇通达的证据(Hudson & Bergman, 1985; Monsell, Doyle, & Haggard, 1989; Sereno & Rayner, 2000, 2003)；词汇水平

下某变量影响前词汇加工的某具体内容。

我们将词汇加工区分成字的加工和词的加工两个环节是基于中文词汇的特点、以往研究结论和 Li 等人提出的词汇加工模型。词的加工与词汇通达有关，字的加工则属于前词汇加工的具体内容。本版修改稿对字的加工进行限定，认为包含字的视觉分析、字形知觉和字的识别等。严格意义上，字的加工和词的加工不会完全割裂。根据交互激活理论，字的加工隶属识别词汇的早期加工阶段，它从一开始便会受词的加工的激活/限制，这意味着字的加工中存在词的加工，但在词汇加工早期字的加工相对占据主导地位，随着加工深入(时间推进)，词的加工才开始占据主导地位。根据以往研究结论得知词频影响词的加工，字频影响字的加工(其中包含字的视觉分析、字形知觉和识别等)，本研究探讨语境预测性是否也影响词汇通达和前词汇加工两个较靠前的加工阶段，根据以往研究结论选择整词词频、首字字频和尾字字频作为变量。总的来说，本研究通过观察语境预测性与频率变量之间的交互作用，为探讨阅读中词汇加工的语境效应提供了一个视角(当然探讨这个问题也存在其他视角)；我们希望通过这个视角中得出一个比较普遍的结论。最终发现，中文阅读中语境预测性与尾字字频交互影响前词汇加工(比如，尾字加工)；词汇加工进入通达阶段时，语境预测性与频率类变量的影响作用则相对独立。

意见 3: 在修改稿中，作者新增了一个非常好的数据分析，贝叶斯分析，但是作者的书写方式不够规范，建议作者参考已有相关文献，将所做的分析进行更精确的描述。请参考下述文献：

Morey, R. D., Rouder, J. N., & Jamil, T. (2015). BayesFactor: Computation of Bayes Factors for Common Designs

[R package version 0.9.12-2]. Retrieved from <https://cran.r-project.org/package=BayesFactor>

Rouder, J. N., & Morey, R. D. (2012). Default bayes factors for model selection in regression. *Multivariate*

Behavioral Research, 47(6), 877–903. <https://doi.org/10.1080/00273171.2012.734737>

回应: 感谢专家肯定。本版修改成稿之前，我们查阅审稿专家提供了文献，另外也增阅了“胡传鹏，孔祥祯，Wagenmakers E. J., Ly, A., 彭凯平, (2018). 贝叶斯因子及其在 JASP 中的实现. *心理科学进展*, 26(6), 951–965”；以及“Rouder, J. N., & Morey, R. D. (2012). Default Bayes factors for model selection in regression. *Multivariate Behavioral Research*, 47(6), 877–903”等文献。参照这些论文对相应结果进行了重组。比如，对实验一中的贝叶斯因子分析的描述，修改成如下描述：

“本实验重点关注语境预测性与词频变量交互作用，基于零假设检验的统计分析发现交

互作用不显著。考虑到通过零假设检验的统计结果(p 值)推断无效应有所不足, 因此参照以往研究在 R 中补做贝叶斯统计分析(Morey et al., 2018), 报告注视时间指标(首次注视时间、凝视时间和总注视时间)的贝叶斯因子(Rouder, & Morey,2012; 胡传鹏,孔祥祯,Wagenmakers, Ly, 彭凯平, 2018)。贝叶斯因子是由“不包含交互作用项模型的后验概率”与“包含交互作用项全模型的后验概率”间比值所得, 这个值代表“假定语境预测性与词频之间的交互作用不显著时, 出现当前数据的可能性”是“在假定交互作用显著时, 出现当前数据可能性”的倍数, 当这个比值大于 1 时被认为倾向于接受虚无假设 H_0 , 即交互作用不存在。默认先验概率(0.5)下, 首次注视时间、凝视时间和总注视时间的贝叶斯因子分别为: 6.29, 4.90, 6.33, 根据相关标准可知当前结果在中等程度上支持零假设(Wagenmakers et al., 2017; 胡传鹏等, 2018); 变换先验概率选择(0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8)多次计算得到的贝叶斯因子值均大于 2.32。可见实验一结果更加倾向于接受虚无假设, 即语境预测性与词频交互作用不存在。”

如果还有其他方面的问题, 还请专家继续批评指正。

意见 4: 作者分别对实验句子的通顺性与合理性进行 5 点评定, 但是通顺性的平均数为 4 以上, 但是合理性的评分是 2 多一点, 请问这样的合理性能满足作者实验的需要吗? 比如, 像作者给出的材料举例, “哥哥要出远门, 妈妈一边帮他收拾岩石一边叮嘱他注意安全。”这样的实验句子的合理性是否会影响到实验结果?

回应: 本版修改稿中, 我们增加的实验材料的通顺性与合理性评定数值。各实验材料通顺性的均值在 4 分(5 点评定, 1 代表非常不通顺, 5 代表非常通顺)以上, 这表明实验材料非常通顺。但合理性的均值为 2 分多一点, 表面上看貌似有些不合要求, 但仔细思考下, 实验材料的评定结果是合理的。理由如下: 首先, 通顺性是 5 点评定, 其中 1 代表非常不通顺, 5 代表非常通顺; 通顺性评定实质上确实是评定了材料的“通顺性”。相反, 合理性的 5 点评定中, 1 代表非常合理, 5 代表非常不合理, 这种评定方式实质上是评定材料的“不合理性”, 这种框架效应会扩大不合理得分。其次, 综合各实验材料的评定结果可见, “有些包含低语境预测性目标词汇框架句子合理性分数”还略优于“包含高语境预测性目标词汇框架句子合理性”, 可见合理性得分并不受目标词语境预测性的影响。最后, 通顺性与合理性是相关的, 没有理由认为不合理的句子会非常通顺。预测性是语境的一种, 合理性是语境的另一种。在我们举例的材料中, 在没有语境的情况下“收拾岩石”和“收拾行李”都是合理的, 但读者根据语境“要出远门”较难预测性前者。综合这些考虑可认为实验材料对结果的具体影响并不起绝对作用, 通过这些实验材料可以探讨阅读中词汇加工的语境效应。

意见 5: 作者新添加的混合线性模型分析结果不符合心理学论文规范。

回应: 本版修改稿已经按照《心理学报》投稿规范修改了统计值, 将小数点保留到 2 位, $p > 0.001$ 时给出具体值, p 值小于这个值时直接写成“ $p < 0.001$ ”。

意见 6: 作者分实验的讨论其实只是实验结果的描述, 太简略。

回应: 本版修改稿已经在每个分实验后面增加了相应的讨论, 请审查。

意见 7: 本研究根据研究结果在讨论中使用大量篇幅对认知模块理论和交互激活理论的争论进行了解释, 但是需要更加谨慎, 一些表述的合理性需要斟酌。“语境预测性与整词词频、首字字频间的交互不显著, 这支持认知模块理论假设; 语境预测性与尾字字频间的交互显著, 则支持交互激活理论假设。”这样的表述是否恰当? 在对两个变量之间是否存在交互进行解释时, 要注意两个变量发生作用的时程, 不在同一时程发生作用的两个变量是无法产生交互作用的, 因此不应据此结果就拒绝交互激活理论的合理性。

回应: 谢谢专家意见。我们的研究结果更加倾向于支持交互激活理论假设。只是根据当前研究和以往研究也并不能推测认知模块理论完全错误, 这个理论也有可取之处。仔细考虑专家的意见, 发现上版修改稿中一些描述并不严谨。本版修改稿做了相应修改, 放弃“语境预测性与整词词频、首字字频间的交互不显著, 这支持认知模块理论假设”这种说法。另外, 本版修改稿将“总讨论”的最后两段进行了合并和整合。修改后的内容如下:

“本研究也检验了交互激活理论与认知模块理论(Fodor, 1983; Forster, 1979; McClelland, & Rumelhart, 1981; Rumelhart, & McClelland, 1982; McClelland, 1987; Morton, 1969)。我们发现, 语境预测性与频率变量交互影响早期词汇加工的某环节(尾字加工), 这符合交互激活理论预期。不过, 至少词汇加工进入通达阶段以后, 两类变量以相对独立的方式影响词汇加工, 这也意味着认知模块理论有一定可取之处。研究发现, 解码词汇较难时(尤其对于阅读能力较低读者而言)会加重对语境信息的依赖(Ashby et al., 2005; Huck et al., 2017), 解码过程达到自动化程度时, 两类加工可相对独立。识别中文词汇需要词切分, 尾字加工与词切分密切相关, 由此可见解码中文词汇的工序较为繁琐复杂(Inhoff & Liu, 1998; Bai et al., 2008; Li et al., 2009, 2014; Shen et al., 2012; Yen et al., 2011; Liang et al., 2015), 这可能是导致成年中文阅读中语境预测性与频率变量(尾字字频)间交互作用显著的重要原因。交互激活加工和认知模块化加工共存于认知系统, 当自下而上编码加工效率不足时, 人类认知系统会加重对背景信息的依赖, 以便产生交互激活的加工策略(Stanovich, 1986)。中文读者也许会根据文本难度和

自身能力实时调整阅读加工策略，进而影响交互激活程度，当然这个问题还需深入探讨。”

如果还有其他方面的问题，请专家继续批评指正。

意见 8：“从整体上看语境预测性与频率变量间的交互作用模式，我们得到了阳性结果(其中发现语境预测性与词内尾字字频交互作用显著)，支持交互激活理论观点。”“中文阅读中，语境预测性与尾字字频变量间的交互作用显著，说明语境预测性对词汇通达的影响，至少起始于字的加工环节。”这些表述仅仅是以实验三的研究结果提出的解释，以此作为整个研究的结论是否恰当？实验二也是以字作为单位进行考察，可是作者并未发现交互作用，那该如何解释？实验二和三研究结果的不一致，作者缺乏相应的讨论。“语境预测性与词内尾字字频交互作用有利于揭示阅读中的切词机制。”如何解释？

回应：本版修改稿放弃了“从整体上看语境预测性与频率变量间的交互作用模式，我们得到了阳性结果(其中发现语境预测性与词内尾字字频交互作用显著)，支持交互激活理论观点。”和“中文阅读中，语境预测性与尾字字频变量间的交互作用显著，说明语境预测性对词汇通达的影响，至少起始于字的加工环节。”这些说法。本版修改稿将“5.3 理论启示”第一段的前两句话修改成如下内容：

“观察语境预测性与词频、字频变量间的交互作用可推测词汇识别的语境效应机制。本研究仅发现语境预测性与尾字字频交互作用显著，而其与词频、首字字频变量间交互作用不显著，由此可以推测语境预测性至少与频率变量交互影响中文词汇识别过程中尾字加工的某环节(比如，视觉分析、字形知觉或单字识别等)。”

将实验三的结果作为整个研究的结论确实不够完整。参照专家意见，本版修改稿将“6 结论”进行修改，修改时参照“概括 3 项实验结果”的原则，修改成如下内容：

“中文阅读中，语境预测性与词汇频率类变量影响词汇加工的模式在不同加工阶段各有特点。语境预测性与尾字字频交互影响前词汇加工(比如，尾字加工)；词汇加工进入通达阶段时，语境预测性与频率类变量的影响作用则相对独立。”

关于总讨论中“语境预测性与词内尾字字频交互作用有利于揭示阅读中的切词机制”这句内容。我们认为，语境加工与尾字加工之间的交互作用，可能是词切分的潜在机制之一。作为总讨论，我们希望能够为后续研究做出探索和扩展，因此并没有以绝对的语气描述。本版修改稿给出了具体理由。相应段落描述如下：

“语境预测性分别与首字字频、尾字字频交互差异有利于揭示阅读中的切词机制，理由

如下：首先，研究发现，目标词汇尾字后添加空格可促进目标词汇识别，而在目标词汇首字前增加空格则达不到相应效果(Liu & Li, 2012)；汉字出现在词尾频率与其出现在词首频率间的比值影响词汇加工与习得(Yen et al., 2011; Liang et al., 2015)，这些表明尾字加工与切词相关。其次，已知语境是阅读切词的重要参照线索(Li et al., 2009, 2011；苏衡等，2017)。最后，词切分发生时程较早(Gu & Li, 2015)，从发生时间上，语境预测性效应、尾字字频效应的发生时程有可能与词切分有所重叠。综合本研究与以往研究结论可以推测：语境加工与尾字加工之间的交互作用，可能是词切分的潜在机制之一。当然，对此还需深入研究。”

如果还有其他方面的问题，请专家继续批评指正。

意见 9：最后一个问题是给作者的建议，一般来讲做研究或者要保持目标词相同，或者要保持语境相同，以便对目标词的差异进行比较，但是作者在这里的目标词和语境均不同，所以建议作者考虑在接下来的研究中采用更好的无关变量操控，以减少无关变量对研究结果的干扰。

回应：感谢专家的建议。我们在编制实验材料的时候也曾经想到过这种情况。研究语境预测性编制实验材料时，要么保持目标词汇相同，要么保持语境相同。这两种思路各有优势和局限。所以仅选择了保持语境相同这种思路。目前，我们正在参照专家的意见按照保持“保持目标词相同”的思路重新编制实验材料，后续研究计划系统考虑语境效应的神经反应基础。

再次感谢专家外审意见，如果还有其他方面的问题，请继续批评指正。

.....

审稿人 2 意见：还有如下建议，供作者考虑：

意见 1：在前言中，作者列举了非常多关于词频和预测性的研究成果。建议作者在总结前期研究的基础上，提出较为明确的研究假设。此外，考察预测性和字频之间的交互作用是本研究的重点，然而在前言中，对预测性和字频的研究成果总结以及理论贡献提及的非常少，建议压缩前一部分内容，在后一部分中增加对预测性和字频研究的文献回顾。

回应：感谢专家的建议。本版修改稿已经参照这个意见增加了对预测性和字频研究的文献回顾。尽管语境预测性和字频交互作用是本研究的重点之一，但我们更加注重从整体上关注语境预测性效应的发生机制特点。虽然语境预测性与词频、首字字频交互作用不显著，提供贝叶斯因子后可得出不存在交互的结论，实验1和实验2也是本研究的重要发现，综合这三项实

验的结果可在整体上得出语境预测性效应的发生时程特点。综上所述，本版修改稿在增加预测性和字频研究文献回顾时，尽量也做到兼顾研究主题的整体结构；总讨论部分的修改也试图从语境预测性的整体角度出发，讨论研究发现，概括提炼研究结论。

意见 2：表 4，8，12 中间的竖线去掉。

回应：相应竖线已经去掉。

总回应：再次，感谢两位审稿专家提出宝贵意见。我们综合参照了所有审稿意见继续对本文进行了修改，本次修改的内容全部以红色字体标出。论文的前言、各实验的分讨论和总讨论进行修改的篇幅较大。如果还有其他方面的问题，还需专家继续批评指正。

第三轮

审稿人 1 意见：

意见 1：建议通顺性和合理性的数值含义统一，以免造成理解偏差，如评定句子通顺性(1 代表非常不通顺，5 代表非常通顺)，评定句子的合理性(1 代表非常不合理，5 代表非常合理)。

回应：感谢专家的意见。本版修改稿件已经统一了通顺性和合理性的数值含义。评定句子通顺性时采用“1 代表非常不通顺，5 代表非常通顺”标准，评定句子的合理性时采用“1 代表非常不合理，5 代表非常合理”标准。评定结果显示实验材料合适，相应内容已经做好了修改。

意见 2：研究结果对已有理论进行阐述时，作者的表述要更明确，避免一些主观评价性的语言，比如“这也意味着认知模块理论有一定可取之处。”，直接说明该理论的合理性的内容即可。

回应：感谢专家的意见。根据专家意见，本版修改稿已经做了相应修改。我们通读全文多编，尽量找出所有不合适的描述，并做出修改。另外，本版修改稿，针对引言进行了精简和重新组织，力图有结构地提出研究主题。本次修改内容以绿色字体标出。论文在一定程度上进行了修改。如果还有其他方面的问题，还请专家继续批评指正。

审稿人 2 意见：建议发表