

《心理学报》审稿意见与作者回应

题目：话题结构对语篇情绪更新的影响

作者：张金璐 吴莹莹 杨晓虹 杨玉芳

第一轮

审稿人 1 意见：该研究将话题结构和语篇情绪结合起来探讨语篇情绪理解是有积极意义的，也做了大量工作，但文章存在着一些漏洞。

意见 1：研究的主题问题。从题目和研究结果的分析来看，作者是想探讨话题结构对语篇情绪更新的影响，但实验设计是将这二者都作为了自变量，是 2×2 两因素被试内设计（因素 1 为话题结构，分为话题延续和话题转换两种水平；因素 2 是情绪状态，分为情绪延续和情绪转换两种水平）。作者在实验 2 和实验 3 发现两因素的交互作用后，在话题结构这个方向上做了简单效应分析，于是认为话题结构对语篇情绪有影响。但是如果分析情绪延续上话题转换和话题延续的简单效应，肯定也有显著差异。这种差异能说又是语篇情绪对话题结构的影响吗？总之，研究的问题不够明确。其实，作者在文中也说到，“在阅读理解的过程中，读者的确既加工了情绪方面的信息又加工了结构方面的信息，并且情绪和结构这两个因素都有可能随着故事的展开而发生变化”。

回应：这个问题要回到本研究的出发点上，前人关于语篇情绪的研究表明，读者能够推断出故事主人公的情绪状态，并且当故事中的情绪发生变化时，读者能够在心理模型上做出相应的情绪更新。接下来的问题是，这个情绪更新的过程是如何发生发展的，其影响因素都有哪些。本研究选取前后两句话的情绪状态是否一致作为自变量之一，目的在于观察对后一句的情绪加工是否会受到前一句情绪的影响，逐句地考察情绪更新的动态过程；同时，选取话题结构作为可能会影响情绪更新的因素，作为另一个自变量，目的在于考察情绪更新的过程是否会受到语篇结构性因素的调节。因此，在两个自变量出现交互作用时，重点在于分析话题结构的延续或转换是如何影响情绪更新的过程，而不是考察情绪的变化是如何影响话题结构的。

意见 2：实验 1 的程序不清楚，而且被试似乎也不好做出评定。读者读完句子后，“屏幕上会呈现一个提示被试做情绪愉悦度判断的图片。图片从左到右依次呈现哭脸、无表情和笑脸的图形，图形下方是一个右指向的长箭头，并标有 1-9 九个阿拉伯数字。当看到这个图片时，被试的任务是需要对刚才读到的语篇做情绪愉悦度的判断。”（1）三个图形依次呈现要表达什么？是从正情绪到负情绪连续体吗？（2）当看到这个图时，是这个，还是这三个？

回应：实验 1 中，在读完每个语篇后，被试的任务是对该语篇的情绪愉悦度做出等级判断，屏幕中出现图片旨在提示被试在此时进行打分，为了确保被试不会弄错打分标准，图片中以三个表情图形和 1-9 九个数字提示被试，1 表示最不愉快（对应的是哭脸的图形），5 表示情绪中性（对应的是无表情的图形），9 表示最愉快（对应的是笑脸的图形）。这是一个整体的图片，表情图形、1-9 九个数字及右指向的长箭头是同时呈现在屏幕上的，并不是依次呈现，看到这个图片是指这个整体的图片。另外，被试在充分理解实验指导语后，会先做一些练习，在练习中有不明白的地方会向主试人员询问，待全部明晰和熟悉后才开始正式实验。据观察，被试基本都没有因实验程序产生疑惑，特别是在练习之后，都能够顺利进行评定。

意见 3：作者认为情绪愉悦度是指“语篇整体愉悦度”，对情绪延续的句子尚可做出情绪的整

体评定，但在情绪转换的句子中被试怎样判断呢？前面一句是积极情绪，后面一句是消极情绪，这样语篇的整体愉悦度是什么，被试恐怕难以做出判断。还需要指出的是，作者要考察的是情绪的转换，既然情绪发生了转换，那么就是第二句话代表的转换后的情绪，而不是整体的。作者又不是研究那种悲喜交加的复杂情绪的，为什么要做整体判断呢？因此这样的评定结果作为因变量是不合适的，是达不到作者的研究目的的。

回应：相较于情绪一致的语篇，情绪上发生转换的语篇，其整体愉悦度的确更难判断一些，但在人们日常所阅读到的语言材料中，前后情绪不一致的情况也是十分常见的，人们在阅读完整个语篇后，依然能够做出情绪推断，众多研究者想探讨的正是读者做出情绪推断的过程和机制。在本研究中，指导语并没有让被试给出具体的情绪（那样比较困难），而是以情绪愉悦度作为因变量，发现在前面一句是积极情绪、后面一句是消极情绪的情况下，从“定性”的角度分析，读者所判断的情绪的确是负性的，但同时我们可以看到，从“定量”的角度分析，最终的情绪愉悦度不单单是由后面一句的情绪决定的，而是会受到第一句情绪的影响（虽然是负性，但负的程度要小于情绪一致的语篇）。这个结果验证了风景模型的假设，也是我们的研究目的之一。

意见 4：作者认为实验 2 和实验 3 是从内隐角度来研究的，这值得商榷。因为在读者自定步速阅读条件下，记录目标句的阅读时间与内隐任务没有什么关系。

回应：我们知道，外显与内隐是一对相对的概念，如外显记忆与内隐记忆、外显学习与内隐学习。概括地说，外显的加工方式与内隐的加工方式，二者最主要的区别在于是否有意识的主动参与。在本文的三个实验中，外显任务是让被试在阅读语篇后对其情绪愉悦度给出等级判断。因此在阅读时，被试是有意识地提取语篇中的情绪信息的，被试的注意是指向情绪的。而相对地，在自然阅读任务中，没有让被试特别注意情绪信息，并不是说被试没有任务，他们的任务是阅读理解，需要判断的是针对语篇内容所做表述的正误，这项探测问题不涉及情绪信息。因此在阅读过程中不需要对情绪信息进行有意识地提取，不受自上而下的影响，对情绪信息的加工是以内隐的方式进行的，因而可以作为内隐任务。

意见 5：格式上有不符合《心理学报》要求的地方，如 p 值，也缺少综合讨论。

回应：在数据的格式方面，包括 p 值在内的各统计指标小数点前的 0 不应省略，现已补全。综合讨论也已经补充完整，见“5 总讨论”。

审稿人 2 意见：本研究通过内隐和外显的加工任务考察了话题结构对语篇情绪更新的影响。选题有一定的价值，但是本研究还存在以下几个问题。

意见 1：问题提出的创新性不够明确。作者认为“不过到目前为止，心理语言学领域中有关语篇情绪理解的研究很少考察语篇结构与情绪理解的关系”，以此作为本研究的主要立足点，还不够充分。语篇结构和情绪理解的关系探讨的较少，可能是因为探讨两者之间关系价值不大，或者说理论意义不强。因此，作者比较结合更多的文献来充分说明本研究的创新性，特别是理论创新方面要重点阐述。

回应：无论是在心理语言学还是自然语言处理领域，语篇结构在语篇理解和情感加工整合中的作用都是一个非常前沿性的课题。我们在本文的引言部分对此进行了介绍，在问题提出部分作了进一步阐述。在前人研究基础上，说明语篇结构对情绪理解可能会产生影响，突出了本研究不同于以往情绪理解相关研究之处，说明了本研究的理论意义和潜在的应用价值。

意见 2：文章前言部分介绍的关于语篇理解的四个理论（建构主义理论、最小化原则、情境模型、风景模型）主要基于大语篇（阅读材料中包含多个句子）的研究结果，而本研究使用

只包含 2 个或 3 个句子的小语篇，不同很好的体现这些理论的观点，并且本研究也没有很好的基于本研究的问题去探讨这些理论的差异。

回应：多位语言学家都提到，语篇的形式丰富多样，并不受限于长度。我们认为，大语篇和小语篇各有优劣之处，关键在于能否与研究目的相适宜。在早期对语篇情绪理解的研究中，研究者使用的材料大多是较长的、由十几个句子组成的语篇。但后来随着研究技术的发展，ERP 和 fMRI 的研究需要考虑到整个实验的时长和被试的承受力，因此语篇材料就变短了很多，由两三句话组成一个语篇十分常见，如 Holt 等人（2008）的实验材料只有两个句子，第一句作为语境，第二句中包含一个关键词。本研究使用只包含 2-3 个句子的小语篇为实验材料，原因有三：一是出于实验目的的考虑，前人通常在一整段语篇后给出目标词和目标句，考察读者对情绪的推断是否正确，而我们的目的不仅仅在于加工的结果，还在于加工的过程，因此必须逐句地测量而不是在语篇末尾才进行测量。二是出于实验设计严密性的考虑，既然要逐句地考察加工过程，就必须控制好无关变量，使条件间除了在自变量上有变化，其余的信息尽量保持一致，在实际的实验操作中，越是长的语篇就越难实现这一点。三是出于经济性的考虑，只要在前两句话中能够提供我们需要的语境，并在条件间充分实现变量的操作，这就足以达到实验目的，并不需要为了语篇长度而繁复赘述。

另外，现已对总讨论加以补充，探讨了本研究与几个语篇理解理论间的关系。

意见 3：文章没有综合讨论。虽然每个实验都有一个小讨论，但作为一个完整的研究，最后必须有一个综合讨论，整合所有实验结果进行更深入的分析 and 探讨。特别是对于实验结果中的一些主要结果要着重说明，比如内隐任务和外显任务的结果为何不同，有何理论价值，再如，被试对情绪转换条件（由正性情绪转为负性情绪）的语篇整体情绪愉悦度判断高于情绪延续的条件（一直保持负性情绪），似乎常识上有正转负，负性情绪会更明显。

回应：根据审稿专家意见，已补充总讨论，对包括内隐任务和外显任务在内的几个问题进行了更深入的分析 and 探讨。关于“由正转负，负性情绪会更明显”这仅是一种直觉上的判断。从实验数据上来看似乎并不如此。读者在对语篇整体愉悦度做判断时，不是单独以第二句话作为标准的，第二句话决定了情绪的最终效价是负性的，但负的程度是受到第一句话的影响的。简单地说，负+负时，整体愉悦度更负，正+负时，情绪是转到负性上来了，但负的程度不那么深。

意见 4：研究方法上也有不妥之处。比如实验一的被试人数为 24 人，实验二、三的被试人数均为 20 人，人数偏少。

回应：关于样本量的确定，我们在实验设计时一方面参考了语篇情绪理解的相关研究，其中行为实验最少的样本量是 Gygas 等（2003）的研究，他们三个实验的被试数分别为 11、19 和 24 人，当然他们的实验设计与本研究的不一樣，不能完全作为参考依据。另一方面我们查找到几篇报告了效应量的文献（有关话题结构的研究和大部分情绪研究的行为实验都没有报告效应量），其中跟本研究比较相近的是一项 ERP 研究（Chen 等人，2010），被试 15 人，实验操纵了语篇前后情绪是否一致和情绪发生改变的位置这两个因素，结果显示主效应和交互作用的 η^2 在 0.18-0.7 的范围内。接下来我们在 G*Power 软件中计算所需样本量，如果取中等大小效应量 0.3， $\alpha=.05$ ，统计检验力取 0.8，经计算，对于一个 2x2 的实验设计来说，最低样本需求量为 17，为了平衡各条件间人数，我们采用 24 个或 20 个被试，此时统计检验力分别能达到 0.94 和 0.88 左右。

意见 5：写作不够规范和认真，有些地方表达不清晰。3 个实验均未写明实验目的，建议明确实验目的。填充材料没有详细说明。因变量指标没有明确说明。

回应：原本三个实验的实验目的都写了在描述实验之前的文字中，由上一段自然引出，但似乎对读者来说不太清晰。现根据审稿人建议，在每个实验的开头一段说明了实验目的，使之更加明确。并且，在实验一的实验材料部分补充了对填充材料的详细描述。因变量的指标是在实验结果中以每个小标题的形式反映出来的，于是没有在前面重复描述。

审稿人 3 意见：《话题结构对语篇情绪更新的影响》一文，研究思路清晰，变量控制较好，方法恰当，有创新性，但还需说明以下几个问题。

意见 1：四种条件下，语篇的连贯性如何？实验 2 和实验 3 中 a 和 b 条件下阅读时间差异显著，c 和 d 条件下阅读时间差异不显著，很可能是前者语篇连贯性差异大，后者连贯性差异小造成的。

回应：的确，语篇的连贯性可能会对实验结果产生潜在影响，“a 条件和 b 条件之间连贯性差异大，c 条件和 d 条件之间连贯性差异小”这个顾虑也是十分合理的。事实上我们在编实验材料时也考虑到这个问题，已经尽量在材料的各条件间进行匹配和调整了。现在，我们补上一个语篇连贯性等级评定的后测，评定结果显示四个条件在语篇连贯性上无显著统计差异（见实验二 3.2 结果部分的最后一段）。

意见 2：作者为什么认为实验任务的不同可能会导致话题结构作用于情绪更新的不同？

回应：首先参考了一些以情绪为考察对象的研究，这类研究提示了在外显的和内隐的加工任务下，被试对情绪加工的结果可能不同。外显任务的优势能够确保被试充分地加工情绪，被试投放注意的地方也正是研究者感兴趣的地方。但外显任务也有缺陷，被试在这种任务下选择性地注意分配给情绪类别或情绪评定，这会使不同刺激（如情绪词和中性词）反应时上的差异，既受到词汇情绪属性本身的影响，又受到选择性注意分配的影响。相对来说，内隐任务不会混淆注意这个变量，但却存在被试由于对情绪的注意不够而导致对情绪的加工深度变浅这个风险（参见 Citron, 2011）。其次，联系到本研究当中，当话题结构信息和情绪信息同时需要被加工时，这两种信息被读者加工的程度也许会因为实验任务的不同而不同：外显的情绪判断任务使得被试把注意资源更多地分配给了情绪信息的加工。此时，语篇中的其他信息就有可能被忽视或无法得到充分加工，话题结构对情绪更新的作用也就不明显。而在自然阅读理解状态下，被试没有特别把注意资源分配给情绪信息，此时包括话题结构信息在内的语篇中其他信息就能够获得更深的加工，那么话题结构对情绪更新的作用就相对突显出来。基于以上两点，我们认为，不同实验任务可能会导致话题结构作用于情绪更新的模式不同。实验结果表明，的确如此。

意见 3：研究总的假设是什么？

回应：1) 语篇理解的心理语言学研究中，人们逐渐注意到，在语篇阅读理解过程中，语篇结构信息可能会对理解过程产生影响，包括语义理解和情感加工。本研究选取话题结构作为语篇结构的代表，检验话题结构对语篇情绪更新是否产生影响。2) 另外，基于实验任务不同对情绪加工带来的影响，本研究提出在外显和内隐加工任务下，话题结构对情绪更新的作用模式可能不同的假设。

意见 4：情绪更新应该指的是读者建立的前后情绪情境模型的变化，如从积极转为消极，从消极转为积极等，在话题延续、情绪延续和在话题转换、情绪延续的情况下，其实进行的是情绪效应的累积，根据实验 2 和 3 的结果，在话题延续的情况下，读者对情绪进行累积而不是更新，在话题转换的情况下，没有进行更新，因此，可以考虑修改论文题目。

回应：对于情绪更新和情绪转换，我们也细致考虑过。这两个概念不同而又存在密切联系。

在本研究中，我们把前后两句情绪一致的条件，视为情绪累加；前后两句情绪不一致的条件，视为情绪转换。同时，我们认为情绪更新这个概念的范畴比较大，所谓更新（updating），是指随着故事情节的展开，读者对语篇中的各种信息在心理模型上进行连续表征的动态过程，只要有变化就会有更新。情绪累加时，情绪的唤醒度发生了变化了；情绪转换时，情绪的效价发生了变化了。因此不论是情绪累加还是情绪转换，读者在心理模型上都要做出相应的更新。基于这些分析，我们最终采用了话题结构对语篇情绪更新的影响这个题目。

意见 5：实验 1 在情绪的判断时间上，话题转换条件的情绪判断时间要短于话题延续条件，为什么？怎样才能证实猜测？

回应：我们发现，在实验 1 的所有结果中，只有这一个指标与其他不同，也就是语篇情绪判断的时间上：情绪转换的语篇需要更长的判断时间。这个结果可能是合理的，因为被试需要额外的时间将第一句话的正性情绪与第二句话的负性情绪做情绪信息的整合，这种带有情绪转换的整合要比情绪延续的情况困难一些。然而另一方面，被试对话题转换的语篇做出情绪判断的时间则要短于话题延续的语篇，我们推测，这有可能是因为当话题转换了，人们并不需要将当前句与前一句的情绪性建立连结，因此省去了额外的整合时间。这种解释是基于前人对话题结构对歧义消解的作用的研究，被试在话题转换条件下对歧义词的注视时间较短，说明话题的转换促进了词汇歧义的消解。他们认为这是因为当话题转换后，一个新的结构被建立了，这使得与旧的结构相关的信息消退了或是很少可用了（见 Clifton & Ferreira, 1987; Lesgold 等, 1979; O'Brien 等, 1986）。在本研究中，要想验证这种解释，可以在话题结构信息得到充分加工时，比较话题转换和话题延续的情况下，读者在情绪理解上的加工表现。在实验 2 和实验 3 中，被试以内隐的方式加工情绪信息，此时话题结构信息得到充分加工，结果发现话题延续时，情绪转换句子比情绪延续句子的阅读时间长。而话题转换时，情绪转换句子与情绪延续句子阅读时间无差异，这说明当话题转换了，人们重新建立关于新结构的表征，不把当前句的情绪信息与前一句做整合，从另一个角度验证了对实验 1 情绪判断时间上的解释。

意见 6：缺少总讨论。

回应：已补充，见“5 总讨论”。

意见 7：需补充关于建构主义理论的相关文献。

回应：已补充，见引言部分第三段蓝色括号。

意见 8：需补充统计检验效力和效果大小值。

回应：已在每个指标的 F 值和 p 值后补充报告了对应的 η^2 值。

第二轮

审稿人 1 意见：作者针对审稿意见进行了很好的回应，也进行了相应的修改。建议作者在接下来的修改中，把回应中的一些观点更好的体现在正文中，特别是文章的创新性和理论价值。修改后可以直接上交编委会。

回应：总讨论增加了小标题，最后一点里增加了理论贡献和实践价值的讨论。

审稿人 2 意见：作者根据审稿老师的意见对论文进行了认真修改，对修改的内容进行了较为详尽的说明，可以看出作者有较为深厚的研究功底。我支持该论文发表。

回应：感谢审稿专家对本研究的悉心指导与肯定鼓励。