

## 《心理学报》审稿意见与作者回应

题目：不同认知方式个体句法成分整合歧义消解的眼动研究

作者：何文广，赵晓静，沈兰玉

---

### 第一轮

#### 审稿人 1 意见：

文章在两个实验中，以母语为汉语的大学生为被试，分别采用离线的评定任务（实验 1）和实时的、在线眼动阅读范式（实验 2），考察了人们在阅读包含低/高依附歧义的汉语句子中，歧义加工的偏好以及语言学（句子结构）和非语言学（场依存或场独立等认知方式）两方面的影响因素。论文研究的问题有理论意义。然而，论文研究方法和数据分析存在明显不足，这些不足使得实验结果并不可靠或者结果的解释并不清楚。这样，论文的结论并未得到可靠证据的支持。此外，论文的立论依据/问题的提出也比较薄弱。论文具体存在以下几个主要问题：

#### 意见 1：

p. 4 “本研究认为其主要原因来自于他们的实验设计。Payne (2014) 等人主要基于相关技术考查了心理冲动性和从句依附偏好间的关系，而没有具体操纵心理冲动性变量，因此，他们的实验结果主要反映的是被试离线加工结果，而不能揭示被试的在线认知具体过程”。这段评论有问题。相关研究还是操纵变量研究与实验结果究竟是反映离线还是在线加工无关。

#### 回应：

谢谢审稿人的意见。造成这样的理解，是我们在表述方面的问题所导致。我们的本意是想表达 Payne (2014) 等人主要是通过问卷调查手段，而后采用相关分析数据处理方式对心理冲动性和从句依附偏好间的关系，同时在确保被试在该变量上没有显著差异后，采用自定步速阅读技术考查年龄、工作记忆、阅读经验和从句依附偏好间的关系。也就是说，他们在研究中没有直接操纵心理冲动性和从句依附间的关系，仅是把冲动性作为控制条件。根据您的意见，我们对文中相关内容做了重新表述。见文中蓝色字体。

#### 意见 2：

p. 4 “与 Payne (2014) 等人的研究不同，本研究采用“认知方式”这一心理变量作为考查

心理控制能力的指标。因为有研究表明，不同场依存类型的被试在句子加工或文本阅读中对外来干扰信息的抑制能力不同，一般说来，场独立型个体抑制干扰信息的能力强于场依存型个体（李寿欣，徐增杰，陈慧媛，2010；李寿欣，陈慧媛，陈建鹏，2013）”。将“认知方式”作为考查心理控制能力的指标，并不妥当。尽管文章引用了两篇将认知方式与抑制能力关联起来的文献，但仅凭数量有限、且来自同一个实验室的两篇文献并不能提供强的证据支持这种关联。

**回应：**

谢谢审稿人的意见。在 Payne (2014) 的研究中，发现心理冲动性和从句依附歧义消解存在显著关系，说明句法成分依附歧义消解受到个体心理控制能力的影响。但是心理控制力是如何影响从句依附歧义消解的，至今还没有研究涉及到。由于依附歧义成分嵌套于句子内部，因此，该成分的加工必定受到临近句子成分的影响，因此我们推测，心理控制能力低的被试，很容易将依附歧义成分挂靠到最先碰到的名词上，而控制力强的被试则有肯能充分考虑该名词的特性而不急于完成挂靠。而这种心理表现，其实就是“场独立”和“场依存”认知方式的体现，所以，本文也就将主要研究目的定位在“认知方式”对句法成分依附的影响方面。

诚然，如审稿人所言，通过“认知方式”考查心理控制能力，可能有些不够全面，因为“认知方式”仅反映心理控制力的一个方面，主要反映个体在认知方面的一种稳定风格。但鉴于已有文献表明，“认知方式”和个体心理控制能力确实有着高相关性，为此，本研究中我们也将“认知方式”看做是考查心理控制能力的一个窗口。

鉴于文献问题，文中又增添了来自不同实验室的相关文献。

**意见 3：**

p. 5 “对汉语这一具有独特语言结构、且拥有众多使用人口的语言来说，人们习惯于何种依附偏好呢？面对句法成分整合歧义，我们又是如何完成歧义消解的呢？依附偏好会因成分句法属性(VP & PP)不同而发生变化吗？”此处需要具体论述“独特语言结构”指的是什么，尤其是需要说明所谓的独特语言结构为什么会影响句法歧义加工。这些方面，如果不能清楚阐述，那么，文章的立论依据就不充分。

**回应：**

谢谢审稿人的意见。根据审稿人的意见，我们补充了有关“汉语独特语言句法结构”内容的表述，见文中蓝色字体。

**意见 4：**

实验 1 数据分析。论文只报告了以被试为随机变量的检验结果，还应该报告以项目为随

机变量的检验结果。另外，“在动宾结构下，句法成分高、低依附差异边缘显著， $t(39) = 1.495$ ， $p = .072$ ”（p. 6），边缘显著意味着作者没有获得可靠的证据。

**回应：**

谢谢审稿人的建议。按照审稿人的建议，我们补充了项目分析结果。在动宾结构条件下，无论是被试分析，还是项目分析，都没有达到显著水平，但都处于边缘显著水平，说明被试具有高依附倾向性。造成该结果的原因可能在于被试量不足，今后的研究中我们会增大被试量进一步验证该结果。

**意见 5：**

实验 2 的句子材料实际上包含永久歧义，这种情形不同于句法歧义领域中研究者实际上考察的几乎都是暂时歧义，所建立的理论也都基于来自暂时歧义加工的实验证据。这意味着，文章所获得的结果，实际上并不能用来讨论迄今为止心理语言学领域已经建立的句法歧义加工或句子加工理论。

**回应：**

谢谢审稿人的意见。诚如审稿人所言，有关句法歧义消解的大部分研究都来自暂时句法歧义，如著名的“花园路径句”，这样可以更方便考查人们面对歧义时是如何消解歧义的。但从上世纪 90 年代开始，该领域中不少研究开始关注永久性歧义句，因为这类句子在生活中大量存在，是人们必须面对的语言现象，如从句依附、介词短语依附、时间副词依附、地点状语依附等。所以，关注这类句子的认知过程更具有现实意义。而且，已有研究表明，对这类句子的研究依然能够很好的讨论已有的句子加工理论，如花园路径理论、约束满意限制理论、竞争激活加工理论、语言分布频率理论等几个较有影响的理论，都可以基于永久性歧义材料的研究得到很好的说明。具体研究可参见 Payne et al., (2014)、Lee, et al., (2004)、Jun et al., (2003)、Traxler et al., (2008)、Frazier et al., (1997)、Swets et al., (2008)、Traxler et al., (1998) 等文献。

**意见 6：**

实验 2 将 NP2（如例句 2a 和 2b 中的“秘书”）作为眼动数据分析的主要区段。问题是，VP 和 PP 两个条件之间，该区段之前的句子内容并不一致，该区段的下一个词也不一致（在上述例句中，分别为介词“在”和动词“佩戴”）。由于可能存在的加工困难溢出效应，或者旁视野预览效应（parafoveal preview effects），上述差异都会影响 NP2 这个区段上的眼动数据。然而，作者并未认真分析和讨论上述可能性。这样，实验 2 结果的解释实际上并不清楚。

**回应：**

谢谢审稿人的意见。兴趣区上、下语境的影响的确是眼动研究中必须考虑的匹配因素。在本实验中我们是这样考虑的：

首先，由于句中第二个名词 NP2 是成分依附歧义产生的地方，所以该区域是本研究关注的核心兴趣区。VP 和 PP 两个条件下，该兴趣区的前四个字都控制在同一水平下“...的经理的 秘书”，由于兴趣区前 4 个成分都是一样的，我们认为前面成分的加工难度溢出效应（旁视预览效应）基本上可以忽略不计。因为，Rayner .K (2008) 等人在研究中指出 word n+2 的预视效应非常小，而且，如果中央窝处的单词词长超过 4 个字母的话，其对 word n+1 的预视效应也是有限的。

其次，对 VP 和 PP 条件下兴趣区前句子成分笔画数 T 检验表明，两种条件下笔画数不存在显著差异， $t(23)=-.932, p=.37$ ，可以消除笔画数带来的影响。

再次，本研究的重心不在于比较 VP 和 PP 两种句式条件下句子成分依附歧义消解机制差异，而是重点考察场独立和场依存个体是如何完成句子成分依附歧义消解的。对于不同的被试来说，他们在实验中阅读的材料是完全一样的。

#### 意见 7：

实验 2 场依存型和场独立型两组被试之间，作者没有进行必要的匹配，如阅读广度等言语工作记忆能力上的匹配，尽管类似这样的匹配是基本的、必要的。

#### 回应：

研究中被试皆来自在校大一学生，参加正式实验前都需要填写“被试资料调查表”，从而保证实验中被试的一致性。实验中所选被试皆为右利手、裸眼或矫正视力正常、自报无阅读障碍、无重大智力疾病史、参加实验当天无身体不适感、教育水平无差异、母语皆为汉语、年龄差异不显著。由于考虑到是眼动实验，被试可以自由回视，且句子长度控制在 20 个汉字以内，对工作记忆载荷要求不高，同时参照国内同行类似研究，所以没测量被试的工作记忆广度。这是我们的不足。现接受审稿人的建议，重新召回参加正式眼动实验的被试，采用 Denman (1980) 言语工作记忆广度测量方法，对被试进行工作记忆测量。结果见正文。具体程序如下：

测试材料最少为 2 个句子一组，最多为 6 个句子一组。每组有 5 套汉语句子，句子长度为 14-22 个汉字。每套句子后均有一个判断句，以考察被试对前面呈现句子的理解情况。使用 E-prime1.1 编制测验程序。共有 100 个句子，分为 25 组，这 25 组句子以随机方式呈现在计算机屏幕上。句子呈现后，要求被试认真阅读，理解每个句子的意思并记住每个句子的尾词。每套句子呈现完毕后，计算机屏幕上会首先呈现一个判断句，要求被试判断这个句

子的意思是否与刚才呈现的那组句子中的次要问题：某个句子的意思一致。随后电脑屏幕上会呈现提示符号，要求被试尽可能按顺序回忆出刚才那组句子的尾词。当被试完成尾词回忆任务后，紧接着呈现另外一组句子。被试工作记忆容量的得分为整个测验中能够正确回忆出来的单词总数。

**意见 8:**

p. 4 “结果发现，从句依附歧义消解除却受到词汇亚范畴信息 (DeDe, 2012)、语境 (Pan & Felser, 2011)、句法复杂性 (Kim & Christianson, 2013) 等语言自身层面的影响外，还受到诸如工作记忆容量、认知控制能力、语言经验、加工速度等非语言层面的影响”。这句话中，“却”应该改为“除”。另外，语言经验不能视作非语言层面的因素。

**回应**

谢谢审稿人的意见。这是我们在语言表述方面不严谨造成的，我们已对此做了修改，并认真通读全文，尽量消除容易产生误解或歧义的语言表达。

**意见 9:**

p. 5 “24 套含有句法成分整合歧义的汉语复杂句式”。这个句子不通顺。另外，论文中不少句子不够通顺和简练。论文中还存在一些错别字，如“对句子语义合理性进行主管评定” (p. 6)。

**回应**

谢谢审稿人的意见。首先为我们的疏忽表示歉意，我们对此已做了相应修改。为尽可能减少类似错误，我们对全文进行了认真通读，并对不当地方做了修改。

**意见 10:**

p. 8 “首次通过阅读时间”。如果作者指的是“first-pass reading time”，那么，中文建议译作“第一遍阅读时间”。

**回应**

谢谢审稿人的建议。按照您的建议，我们对此做了重新表述。

.....

**审稿人 2 意见:**

文章采用离线主观评定和实时眼动技术，探讨了认知方式（场依存和场独立）对汉语句法成分依附歧义消解机制的影响。研究思路清晰，研究问题对于了解汉语句法加工特点具有重要的理论意义。研究在以下几个方面还需要作者进一步思考。

### 意见 1:

两个实验都比较了动宾结构和介宾结构句子下句法成分整合的特点,但为什么对这两种情况进行比较,有什么预期的理论假设?而且结果出现了差异,又说明了什么问题。文章在引言和讨论中还需要进一步解释清楚。

### 回应:

谢谢审稿人的意见。之所以选择动宾和介宾这两种结构,是因为它们是汉语领域比较常见的两类修饰成分,关注这两类句法结构的整合机制,更具有现实意义。同时,由于动宾结构属于“verb+noun”构式,其间涉及到论元建构,而介宾结构属于“prep+noun”构式,其间不涉及论元建构。所以,我们预期动宾结构依附歧义消解可能需要消耗更多的认知资源。结果证实了这种设想,说明论元建构是导致句子加工难度的一个重要因素,对于解释不通句子加工难度差异具有一定的理论意义。

接受审稿人的建议,我们在文中相应部分增添了有关内容。

### 意见 2:

实验设计仅包含存在整合歧义的句子,从一般的实验设计来说,通常把不存在整合歧义的句子作为控制变量,这样可以比较,在不同条件下被试是否都倾向于高依附偏好。本研究似乎缺少这一重要的控制变量。

### 回应:

谢谢审稿人的意见。

实验设计遵从于研究目的。的确,如审稿人所言,在以探究句子具体加工过程的实验中,尤其是考查句法歧义消解的研究中,在材料选取上往往设置歧义句式和正常句式,以便于能更清晰的揭示歧义句式的认知过程。我们在另一项旨在揭示依附歧义消解认知机制的研究中,便采取了这种实验设计。由于本研究目的不在于探究修饰成分依附歧义的具体消解机制,而在于考查不同认知方式的个体在面临含有修饰成分依附歧义的句子时,是如何加工这类句子的,重点在于考查认知方式带来的影响,所以没有加入正常句子作为额外控制变量。因为,我们认为,通过单独考查含有歧义的句式的认知过程,更能直接比较两类被试的具体认知差异性。

当然,您的建议可能更有助于揭示认知方式不同的个体在面临歧义句法结构时,是如何完成歧义消解的,而这也是我们接下来的研究要做的工作。在此,再次向您表示衷心的感谢。

### 意见 3:

实验二的结果在任何指标上都没有发现显著的交互作用,作者如何得出“不同认知方式

的被试在句法整合机制上存在差异，场依存型被试更容易受句中其他成分的影响”？从数据上看，只能看出，场依存被试只是需要更长的时间来阅读，而这很可能与场依存被试的工作记忆容量有关（如引言和讨论中提到）。而且，既然工作记忆是一个潜在的重要的混淆变量，为什么作者并没有对被试工作记忆容量进行测量，或者说已经测量，但文章中没有提供这方面的信息？

回应：

谢谢审稿人的意见。

(1) 由于本研究旨在考查不同认知方式的个体加工含有句法依附歧义的句子差异性，因而两类被试在实验中阅读同样的实验材料，主要通过考查他们阅读同样实验材料时间的差异性来推测其认知过程的差异性。结果中没有发现材料和被试类型间的交互作用，都表现为动宾加工困难、介宾加工容易，而且在兴趣区位置，都表现为场独立型被试用时较短。如果两类被试在加工含有依附歧义的句子时，采用了类似的认知机制，则在关键区位置（NP2），阅读时间应不存在显著差异性。由于结果中发现了差异性，所以结合实验假设，我们推测两类被试在加工机制上存在差异性。

(2) 工作记忆问题：由于考虑到眼动实验，被试可以自由回视，且句子长度控制在20个汉字以内，对工作记忆载荷要求不高，同时参照国内同行类似研究，所以没测量被试的工作记忆广度。这是我们的不足。现接受审稿人的建议，重新召回参加正式眼动实验的被试，采用 Denman (1980) 言语工作记忆广度测量方法，对被试进行工作记忆测量。具体程序如下：

测试材料最少为2个句子一组，最多为6个句子一组。每组有5套汉语句，句子长度为14-22个汉字。每套句子后均有一个判断句，以考察被试对前面呈现句子的理解情况。使用 E-prime1.1 编制测验程序。共有100个句子，分为25组，这25组句子以随机方式呈现在计算机屏幕上。句子呈现后，要求被试认真阅读，理解每个句子的意思并记住每个句子的尾词。每套句子呈现完毕后，计算机屏幕上会首先呈现一个判断句，要求被试判断这个句子的意思是否与刚才呈现的那组句子中的次要问题：某个句子的意思一致。随后电脑屏幕上会呈现提示符号，要求被试尽可能按顺序回忆出刚才那组句子的尾词。当被试完成尾词回忆任务后，紧接着呈现另外一组句子。被试工作记忆容量的得分为整个测验中能够正确回忆出来的单词总数。

结果显示两类被试在言语工作记忆广度上不存在显著性差异： $M_{\text{场独立}} = 78.11(7.92)$ ， $M_{\text{场依存}} = 75.08(6.85)$ ， $t(70) = 1.74$ ， $p = .087$ ，结果已在文中补上。

#### 意见 4:

实验一的结果表明,“汉语句法成分整合歧义具有稳定性,表现为高依附现象”。但根据实验结果,在动宾结构下,高低依附的差异并没有达到可靠的显著性差异,所以,得出这样的结论要小心。

#### 回应:

谢谢审稿人的意见。根据您的意见,我们对结论做了重新表述。

#### 意见 5:

实验采用 Eyelink1000 进行测试,被试与屏幕之间的距离为 55cm-60cm,不知道为什么这个距离还在一定范围之内,而不是一个固定数值。请解释。

#### 回应:

谢谢审稿人的意见。实验中下巴托到刺激显示屏中心直线距离是 60cm,但是在具体实验过程中,考虑到被试有的戴眼镜,有的下巴放置的靠前,或者由于被试座椅位置的差异,造成其眼睛到刺激屏的真实距离可能有浮动,但浮动不会太大,不会超过 5cm,且只能变小,不能增加,因为下巴托固定在那儿,被试的眼睛只能往前靠近。

#### 意见 6:

作者将注视点小于 60ms 和长于 4000ms 的数据删除,请提供删除的依据

#### 回应:

谢谢审稿人的意见。有关眼动原始数据删除标准,就最小注视时间标准来说,有些研究将该标准界定在 70-100ms,当前国内外一般采用 80ms,但有关中文阅读中,短注视点较英文更多,故将该标准定为 60ms。(闫国利等,2013,“阅读研究中的主要眼动指标评述”《心理科学进展》, )。Rayner 教授认为一次阅读中,注视点很难超过 1000ms,结合我们的研究数据及 3 个标准差的标准,我们发现研究中极个别被试的极个别数据超过 4000ms,说明被试期间可能在思考别的事情或短暂放松,故作为极端值处理。由于这些数据的存在影响到 3 个标准差的数值,故参照类似研究,先剔除极端值,而后再按照 3 个标准差的原则筛选数据。另外,国内有关研究也采用了该删除标准(汉语主、宾关系从句加工难度及其核心名词生命性效应,心理科学,2016(1))

#### 意见 7:

对实验二中被试回答问题的倾向性需要做进一步统计分析



回应:

谢谢审稿人的意见。根据你的建议我们对这部分做了进一步统计分析。

意见 8:

文章中 MSE 的数据异常大, 请进一步仔细检查结果, 确保结果的正确性。

回应:

谢谢审稿人的意见。我们对文中数据进行了认真校对, 对其中因打字、小数点忽略造成的错误性数字做了相应矫正。再次向您的严谨和认真表示感谢。

意见 9:

句尾区域的内容不一致, 没有进行任何控制, 注视时间的这些结果很难单纯地说明要研究的问题。

回应:

谢谢审稿人的意见。虽然句尾区域不是关键兴趣区, 但对于揭示 NP2 加工的溢出效应和句子的 wrap-up 效应有帮助, 所以在研究中作为考查区域进行分析。考虑到句尾两部分内容不一, 为尽可能消除材料方面的差异性, 我们对这部分汉字的笔画数做了相应匹配处理, 结果显示不存在显著差异性,  $t(23)=-.932, p=.37$ 。同时为减少材料的干扰, 我们在制作实验材料时, 尽量保持 vp 和 pp 类型材料在汉字水平完全一致, 如

(2a) 警察发现佩戴领带的经理的秘书在会议室开会。(VP)

(2b) 警察发现在会议室开会的经理的秘书佩戴领带。(PP)。

---

## 第二轮

审稿人 1 意见

这是一篇修改稿。在第一轮审稿意见中, 我提出了七个主要问题。作者对这些问题做出了认真回应和修改, 并且较好地解决了其中五个问题。然而, 余下两个非常实质性的问题, 文章修改稿并未解决。这两个问题分别涉及被试和实验材料。具体说明如下:

意见 1:

被试。第一轮审稿意见中, 我提出, “实验 2 场依存型和场独立型两组被试之间, 作者没有进行必要的匹配, 如阅读广度等言语工作记忆能力上的匹配, 尽管类似这样的匹配是基本的、必要的”(见第一轮审稿主要问题#7)。

修改稿补充了两组被试之间言语工作记忆广度方面的数据: “Daneman (1980)言语阅读工作记忆广度测试显示, 两类被试差异不显著, M 场独立 = 78.11(7.92), M 场依存 = 75.08(6.85),  $t(70) = 1.74, p = .087$ ”(p. 18)。问题是, t 检验结果发现了两组被试之间工作记

忆广度存在边缘显著的差异 ( $0.05 < p < 0.10$ )。这样的统计检验结果不能作为两组被试工作记忆广度没有差别的证据。这意味着,场依存型和场独立型两组被试之间不仅认知方式上有差异,工作记忆上也可能有差异,因此,文章所使用的两组被试,实际上存在被试变量上的混淆。这种混淆也可能是认知方式与修饰成分两个因素之间,在各个眼动指标上交互作用基本上都不显著的原因之一。在存在上述问题的情况下,文章实际上无法得出任何确切的结论。

#### 回应:

谢谢审稿人的意见。非常感谢审稿人对该问题的重视,您的意见对提升本文质量,进一步厘清认知方式对句法成分依附歧义消解有重要意义。如部分研究所言,工作记忆和认知方式有着密切关联,一般而言高工作记忆被试倾向于场独立认知方式,低工作记忆被试倾向于场依存认知方式。很难筛选出在工作记忆方面无显著差异、而在认知方式上又截然不同的两类被试,这可能是本研究中两类被试在工作记忆方面存在边缘差异的原因。结合审稿人的意见和本研究实验设计,我们以工作记忆为协变量,以认知方式和修饰成分为考查变量,重新对实验数据做了分析。以便尽可能分离出工作记忆带来的干扰。具体见文中结果绿色字体部分。分析结果显示,当工作记忆作为协变量进入数据分析时,对认知方式的影响不大,但对修饰成分的影响相对较大。在某种程度上证实了审稿人的预期。在此,再次向审稿人表示感谢!

#### 意见 2:

第一轮审稿意见中,我提出,“实验 2 将 NP2 (如例句 2a 和 2b 中的“秘书”)作为眼动数据分析的主要区段。问题是,VP 和 PP 两个条件之间,该区段之前的句子内容并不一致,该区段的下一个词也不一致(在上述例句中,分别为介词“在”和动词“佩戴”)。由于可能存在的加工困难溢出效应,或者旁视野预览效应 (parafoveal preview effects),上述差异都会影响 NP2 这个区段上的眼动数据。然而,作者并未认真分析和讨论上述可能性。这样,实验 2 结果的解释实际上并不清楚”(见第一轮审稿主要问题#6)。

(2a) 警察发现佩戴领带[的经理的]【秘书】[在会议室]开会。(动宾结构,VP)

(2b) 警察发现在会议室开会[的经理的]【秘书】[佩戴]领带。(介宾结构,PP)

修改稿在回应上述问题时声称:“VP 和 PP 两个条件下,该兴趣区的前四个字都控制在同一水平下“...的经理的 秘书”,由于兴趣区前 4 个成分都是一样的,我们认为前面成分的加工难度溢出效应(旁视野预览效应)基本上可以忽略不计。因为,Rayner .K (2008) 等人在研究中指出 word n+2 的预视效应非常小,而且,如果中央窝处的单词词长超过 4 个字母

的话，其对 word n+1 的预视效应也是有限的。”

问题是，只考虑关键区的前四个字是不够的。尽管修改稿补充报告说 VP 和 PP 两种条件之间兴趣区前句子成分笔画数不存在显著差异，但笔画数只是混淆变量之一。总之，为了排除加工困难溢出的可能性，恰当的方法是对关键区之前区段的眼动数据进行统计检验，而不是简单地说“前四个字”相同。

此外，作者错误地理解了“旁视野预览效应”的含义，该效应指的是“秘书”之后的字或词造成的效应，而不是“秘书”之前的语言内容。在 2a 和 2b 中，关键区“秘书”之后的第一个字分别为“在”和“佩”，二者之间笔画数、字频等因素非常可能造成两个条件之间“秘书”注视时间的差异——这种差异是我所指的“旁视野预览效应”。

在回应质疑时，作者说“本研究的重心不在于比较 VP 和 PP 两种句式条件下句子成分依附歧义消解机制差异，而是重点考察场独立和场依存个体是如何完成句子成分依附歧义消解的。对于不同的被试来说，他们在实验中阅读的材料是完全一样的”。这是相当让人惊讶的说法。文章采用的是 2（认知方式：场独立、场依存） $\times$ 2（修饰成分：VP、PP）两因素设计，除了认知方式这个被试变量之外，修饰成分是唯一的一个语言学变量。难道文章不是应该比较 VP 与 PP 之间的差异，然后再看这个差异是否受被试认知方式的调整吗？

#### 回应：

谢谢审稿人的意见。在第一稿审稿意见中，由于我们的注意焦点主要集中于加工溢出效应，受传统思维习惯的影响，所以主要针对兴趣区之前的句子成分做了讨论，而且在讨论时主要基于 Rayner 的观点，因此忽视了兴趣区前所有成分和兴趣区后部分区域的影响。在此，我们接受审稿人的建议，对兴趣区（NP2）前的句子成分进行统计分析。由于在兴趣区（NP2）前，两种句式仅是在某一区域（如“佩戴领带”和“在会议室开会”）存在差异，其余成分完全相同，因此我们基于两种句式（VP 和 PP）对上述差异区域的总阅读时间进行了单因素方差分析，结果显示，两区域不存在显著性差异，具体见文中实验 2 眼动指标分析部分。另外，我们接受审稿人的建议，将工作记忆容量作为协变量，重新对眼动数据进行分析，结果发现句尾区域（NP2 后的所有部分）句子类型主效应不显著，即“佩戴领带”和“在会议室开会”在阅读时间上不存在显著差异。另外，该发现也证实了审稿人的推测，工作记忆确实对两类句式的加工存在影响，因为，在论文的上一稿，在没考虑工作记忆的条件下，句尾区域的加工是存在显著的句式类型主效应的，而当控制工作记忆变量后，句式类型主效应消失。由于在控制工作记忆变量后，两类句式（VP 和 PP）不存在类型差异，因此，兴趣区 NP2 加工

中来自句尾区域旁视野效应，就没再做分析。

在此，再次向审稿人表示感谢，同时也为初稿中我们的疏忽表示歉意。

.....

#### 审稿人 2 意见

##### 意见 1:

既然“论元建构”是比较动宾结构和介宾结构句子下句法整合特点的关键点和理论推动，为什么不在引言中引入该论点，并对此进行详细定义和介绍？这样增强了文章的理论基础，明确了预期假设，讨论起来还将更顺理成章。

##### 回应:

谢谢审稿人的意见。基于您的建议，我在文中增加了相应内容。并基于该内容对研究结果做了些许讨论。

##### 意见 2:

作者在不同眼动指标上发现了认知方式的主效应，说明场依存被试在 NP1 区域已经完成了句法成分整合，.....。这种推理有点牵强。两组被试之间的差异，不能说一组被试完成了句法整合，另外一组没有完成。这种差异很可能简单地反映了两组被试阅读能力的差异，而这种能力受工作记忆容量的调节。

##### 回应:

谢谢审稿人的意见。针对您的意见，我们对文中相应部分内容做了重新表述。而且我们将工作记忆作为协变量，对已有数据重新进行了分析，结果发现，工作记忆的确对句法成分依附有影响，但这种影响主要体现在句式类型方面。在前一稿中，在没考虑工作记忆的条件下，在多项眼动指标上，我们发现 VP 结构的加工要难于 PP 结构的加工，但是当把工作记忆作为协变量进入数据分析时，结果发现 VP 结构和 PP 结构在上述眼动指标上并不存在差异性。该发现说明，VP 和 PP 在加工难度上差异跟多是基于工作记忆容量的原因，认知方式对其影响不具有差异性。

##### 意见 3:

作者对被试进行了工作记忆测试，结果显示场依存被试的工作记忆小于场独立被试，这种差异在.087 水平上显著。两组被试在工作记忆上的差异似乎无法忽视，不能简单地说不存在显著差异。作者应该充分利用搜集的这些数据，进一步挖掘到底什么因素贡献了最后的结果。

##### 回应:

谢谢审稿人的建议。基于您的建议，我们在修改稿中将工作记忆作为协变量对数据重新进行了处理，结果见意见 2 中的答复。

**意见 4:**

通常被试眼睛与屏幕之间的距离是一个固定值，作者应该避免那些可变的因素，特别是在阅读的眼动研究中。

**回应:**

谢谢审稿人的意见。我们已经数值调回屏幕与眼睛之间的固定距离：60cm.今后的研究中，我们会尽量控制这方面的误差。

**意见 5:**

MSE 数据仍然非常大，请作者再次核实结果。

**回应:**

谢谢审稿人的建议。我们对文中数据做了重新处理，上述问题已得到解决。

**意见 6:**

为什么实验一的结果分析采用 t 检验，而实验二回答倾向性分析采用方差分析？图 1 显示的是不同条件下的百分比，图 2 显示的是频次统计，但分析时依据百分比进行分析？

**回应:**

谢谢审稿人的建议。实验 1 和实验 2 是在不同的时段完成和成文的。最初是基于频次采用卡方完成的统计，后接受国、内外同行专家的建议，同时考虑到时单因素实验设计，主要是针对两种不同句式类型的依附偏好进行比较，所以改用了配对 t 检验，更能有效体现实验结果。实验 2 中，由于完成实验后，对选择倾向是基于频次统计，成文时没再做改动。统计中由于涉及到三个因素，不适合做 t 检验，所以采用方差分析，考虑到计数数据不适合做方差分析，所以基于频次将数据转换为百分比。接受您的建议，我们将实验 2 中的频次图改为百分比图，以便和实验 1 统一。

**意见 7:**

作者没有对眼动指标进行定义，而且似乎同一指标在不同地方还有不同的名称？

**回应:**

谢谢审稿人的建议。基于您的建议，相关眼动指标定义做了补充，并在文中做了统一。

**意见 8:**

作者采用 Daneman (1980)言语阅读工作记忆广度任务对被试进行测试，作者真正采用

Daneman 的版本来测量被试？还是采用人家的方法来测量自己的被试？

回应：

谢谢审稿人的意见。工作记忆的测量是基于 Daneman 的方法，结合汉语被试，自行编制材料完成的，而不是采用他们的版本。

---

### 第三轮

审稿人 2 意见

意见 1：

在二审中，建议作者在引言中对“论元建构”进行详细定义和介绍。在修改稿中，作者只在引言中增加了一句话，让读者很难明白到底什么是“论元建构”，这与本研究的操纵有什么关系，完全不清楚。

回应：

谢谢审稿人的意见。根据您的建议，我们再次对论文相应部分内容做了完善和补充，具体内容见文中相应部分绿色字体。

意见 2：

作者提出“动宾结构涉及到论元建构，需要消耗认知资源，而介宾结构，不涉及到的论元建构，资源消耗相对较少”。但“实验 2 的结果没有证实有关动宾结构（VP）的加工难于介宾结构（PP）的预期，在控制工作记忆容量的条件下，VP 和 PP 在考查眼动指标上没有显著差异。但该结果并不能完全否认两种句式结构再加工难度上不存在差异性。因为，在没有排除工作记忆的条件下，VP 结构的阅读时间显著长于 PP 结构。该发现在一定程度上说明 VP 结构和 PP 结构的加工受工作记忆的调节，而较少受认知方式的影响”。这些逻辑很难让人理解。

回应：

谢谢审稿人的意见。基于另一位审稿人的建议，我们对论文材料和结构做了很大调整，仅在预实验中对动宾结构（VP）和介宾结构（PP）做了研究，正式实验仅选取动宾结构作为实验材料，新的实验中不涉及认知方式和材料属性间的交互关系。

意见 3：

作者“对两种句式下词汇水平存在差异的区域（如“佩戴领带”\_&“在会议室开会”）的总阅读时间进行单因素方差分析，结果显示两个区域在总阅读时间上不存在显著性差异”，作

者没有提供相关的平均数，根据 F 检验的结果， $F_2 = 2.41$ ，接近显著了。

**回应：**

谢谢审稿人的意见。基于您的问题和另一位审稿人的建议，我们重新设计了实验材料，并用新材料进行了新的实验。新材料的设计克服了上述不足。

**意见 4：**

文章存在一些低级拼写打印错误，请认真修改。

**回应：**

谢谢审稿人的意见。根据新的实验完成文稿后，我们多次对论文进行了通读，尽量克服因拼写和打印而造成的低级错误。

.....

**审稿人 3 意见**

**意见 1：**

根据实验二结果并不能得出可靠的结论，因为所有实验材料都是歧义句，缺乏正常句子对照组。“认知方式”是一个组间因素，两组被试存在系统性的差异（认知方式，以及可能存在的工作记忆差异等），当两组被试的阅读时间出现显著差异时，并不清楚被试只是在歧义句上如此，还是对所有句子都是如此。如果是前者，则可认为认知方式对解歧过程存在特定的影响，文章的结论是合适的；如果是后者，无论对歧义句还是正常句，场独立型的被试整体上都比场依存型的被试阅读得更快，那么，文章将不能得出任何结论，也就失去了意义。而区分是前者还是后者的关键，是设置一个正常句子对照组。

**回应：**

谢谢审稿人的意见。基于您的建议，我们重新设计了实验材料。新材料包含三类句式：依附歧义句、高依附句式、低依附句式。并使用新材料重新做了实验。希望合乎您的要求。

**意见 2：**

研究旨在考察不同认知方式个体的句法解歧过程是否存在差别，而实验一只考察两种不同句式的解歧偏好，并不涉及“不同认知方式个体”，显然不能回答该问题，似乎没有必要单独作为一个正式实验列出，考虑到文中提到实验一“旨在考查汉语领域中成分整合歧义偏好表现，为实验二提供实验假设依据”，似乎实验一作为实验二的预实验较为妥当。

**回应：**

谢谢审稿人的建议。基于您的建议，我们对论文结构进行了重新调整。将原文的实验 1 改为预实验部分，原文实验 2 删除。基于新材料重新做了正式实验。

### 意见 3:

实验一发现读者对两种句式的歧义整合偏好是一致的。那么，在实验一的结果“为实验二提供实验假设依据”的前提下，实验二似乎没有必要再坚持把句式作为一个变量，可以把两种不同句式的材料合二为一。——况且实验二结果也证实了在眼动实验中，句式依然没有产生影响。作者说明了坚持把句式作为一个变量的原因，这个理由说得通，但是在实验一的结果不支持的情况下，可能需要相关研究文献的支持。

### 回应:

谢谢审稿人的建议。基于您的建议我们对论文材料和结构做了很大调整和改变。仅在预实验中对两种句式做了比较和分析。正式眼动实验中仅选取更具有代表性的动宾结构做了认知方式方面的研究。

### 意见 4:

另外，在讨论中，作者指出“认知方式不同的个体在依附倾向选择上都倾向于高依附”，说明两种不同认知方式个体是以相同方式来进行歧义消解的，然后又指出实验结果中两组的差异是因为“对于场依存型被试来说，由于心理控制能力相对较弱，当阅读到 NP1 区域时，很容易将前面的修饰成分挂靠到 NP1。但是，当阅读到 NP2 区域时，由于受语言习惯和经验的影响，认识到前面的句法分析可能存在错误，于是启动句法再分析过程，从而导致阅读时间的延长。”这个解释是没有说服力的，因为场独立型的被试也可能是这样做的，只是他们加工得更快。——由于没有对照组的数据，所以没办法排除这种可能性。这跟第 1 点其实是同一个问题。

### 回应:

谢谢审稿人的意见。基于您的建议我们对论文材料和实验设计做了重新调整。并就不同认知方式个体在 NP2 上出现的差异性做了新的理论分析。具体见文中相应部分绿色字体。

### 意见 5:

前言第二段中写到：“句法整合歧义是人类语言领域内一种非常普遍的语言现象”。似乎不够严谨。句法整合歧义是语言研究领域常见的研究对象，但其实是一种特殊的语言现象，研究中使用的实验材料在现实生活中并不常见，人们在日常口语中会尽量避免此类歧义。另外，同一段中，“俗称高依附偏好”，“俗称”也不合适。文中还有多处表述和词汇欠妥，请作者检查修改。

### 回应:

谢谢审稿人的意见。基于新的实验论文完稿后，我们对论文遣词造句做了认真、多次重



复性阅读和修改，希望能合乎您的标准。

**意见 6:**

前言中，“本研究采用“认知方式”这一心理变量作为考查心理控制能力的指标”，“认知方式”在这里第一次出现，需要下定义。对于不了解此概念的读者来说，此处逻辑上是不连贯的。

**回应:**

谢谢审稿人的意见。我们在文中相应部分补充了有关认知方式的含义，见文中相应部分绿色字体。

**意见 7:**

前言中，“另外，动宾结构（VP）的加工涉及到论元建构，而介宾结构（PP）则不涉及到论元建构，由于论元建构需要消耗更多的认知资源（Gibson, 1998），因此，本文预期认知方式对句法成分依附歧义消解的影响可能因为句法属性的不同而有所不同。”此处论元建构也需要定义和必要的解释，以方便读者理解。另外，这几个句子之间的逻辑跳跃比较大。如，为什么动宾结构涉及论元建构，介宾结构不涉及？为什么消耗更多认知资源会对认知方式的作用产生影响？又是如何产生影响的？需要展开说明。

**回应:**

谢谢审稿人的意见。我们在文中相应部分就论元和论元建构的含义做了补充说明，并就动宾和介宾结构加工中为什么消耗的认知资源不同，也做了说明。见文中相应部分绿色字体。

**意见 8:**

“本研究的主要目的是：（1）汉语领域中，句法成分整合偏好的具体表现形式？”不是完整的句子。

**回应:**

谢谢审稿人的意见。我们对该问题做了重新表述，希望合乎语言标准和习惯。

**意见 9:**

“场独立型个体由于心理控制能力较强，不容易受外界因素的影响，因此更有可能基于实际句法结构建构句子表征。”这里的问题是，“基于实际句法结构建构句子表征”指的是被试阅读 NP2 之前还是之后的事？根据上下文，应该指的是之前，但如果是之前，这是不合理的，在阅读到 NP2 之前，被试并不知道“实际句法结构”是什么（其实考虑到实验采用的

是永久歧义句，实验中并不存在一个“实际句法结构”，而只存在“被试心目中的句法结构”，且该结构对于每个被试每个句子都是变化的）；而如果是在阅读之后，那么场独立和场依存型就没什么区别了，阅读完 NP2 之后，场依存型的被试也是根据他们心中认为的“实际句法结构”来重新构建句法结构的。

**回应：**

谢谢审稿人的意见。我们基于新的实验，对不同认知方式个体为什么在歧义消解区表现出认知上的差异重新做了理论分析，见问题 4 和文中相应部分绿色字体。

**意见 10：**

“屏幕刷新率为 1024×768 像素。”这不是屏幕刷新率。。。这是分辨率。

**回应：**

谢谢审稿人的建议。我们对此做了改正。

**意见 11：**

“从认知过程看，不同认知方式的个体在在上述两个区域的各项眼动指标上都表现出差异性，说明不同认知方式的个体在句法成分依附歧义消解方面有着不同的心理机制。”如前所述，应有正常句作为对照，证明在正常句的阅读上不出现此种差异，才能下此结论。

**回应：**

谢谢审稿人的建议。我们基于您的建议重新设计了实验材料，增加了对照材料。

**意见 12：**

“从离线结果看，认知方式对句法成分依附偏好没有影响，”如前所述，离线研究并没有涉及认知方式的操控，不能下此结论。

**回应：**

谢谢审稿人的建议。

我们对论文结果和讨论部分重新做了更具针对性的表述，新的表述中避免了这类问题的出现。

---

## 第四轮

**审稿人 2 意见**

文章重新设计了实验材料，重新收集了数据，解决了以往难以说明的问题。还有几个小问题请作者认真检查和修改：

**意见 1:**

新材料中是否包括原有的材料？实验中被试没有变化，也就是说，之前的被试读了新的材料？

**回应:**

谢谢审稿人的意见。根据专家们的建议，我们使用新材料重新做了眼动实验。新材料中部分句子是根据原有材料改编的，但仅是在部分字段或人物角色上和原有材料相同，其余不同。新实验中依然使用原来的被试，主要基于如下两点考虑：第一，不需要对被试再次进行工作记忆容量的测试，可以在编辑部规定的时间内，尽可能完成新实验；第二，该研究中的原初实验是在 2016 年 3-4 月间完成的，到开展新实验已经间隔了 1 年多的时间，且实验材料发生很大改变，所以，因材料而带来的熟悉效应应当得到了有效控制。

**意见 2:**

摘要中缺少明确的结论。

**回应:**

谢谢审稿人的意见。根据您的意见，我们对论文摘要做了相应修改和补充，见文中对应部分蓝色字体。

**意见 3:**

文章中存在很多种标准差的表述方式，一会用  $M\pm$  表示，一会用  $SD$ ，一会直接放括号，什么也不解释。请采用专业的表述方式进行描述。

**回应:**

谢谢审稿人的意见。根据您的建议我们对文中标准差的呈现方式做了修改。

**意见 4:**

请弄清楚哪个是采样率，哪一个是刷新率？

**回应:**

谢谢审稿人的意见。根据您的意见，我们进一步参阅了相关资料，并在此基础上对文中相应内容做了修改。

**意见 5:**

英文摘要存在语法和表述问题，请认真修改。

**回应:**

谢谢审稿人的建议。根据您的建议我们对英文摘要做了新的表述，并请在国外有 10 多年生活经历的心理学专业人士做了润色处理。

.....

### 审稿人 3 意见

这是一篇修改稿。作者报告了新的实验，修改了文稿。新的实验设计和实验材料有很大改进，实验数据更丰富，实验结果较好地解决了之前的问题。作者对审稿意见逐一做了回复，说明了所做的修改，整体改善很大。论文还存在一些不足，建议小修后发表。

#### 意见 1:

针对上一轮审稿意见第 10 点，作者答复已修改，但正文中实际并没有修改。请作者核实。

#### 回应:

谢谢审稿人的意见。为我们上次的疏忽向审稿专家表示衷心的感谢！

#### 意见 2:

有一些较隐蔽的错别字，如 3.1.2 小节的“参杂”，应为“掺杂”，List 最好还是用中文表述，等等

#### 回应:

谢谢审稿人的建议。向您的严谨和认真表示致意！基于您的建议，我们对论文做了认真研读，并对类似错误和表达做了修改和矫正。

#### 意见 3:

在前言中，作者提到“来自印欧语系的研究发现，人们在从句依附歧义消解方面表现出鲜明的跨语言差异性。英语、罗马尼亚语、瑞典语、挪威语、巴斯克语、葡萄牙语、阿拉伯语等语种倾向于低依附偏好，而西班牙语、荷兰语、希腊语、德语、法语、俄语则倾向于高依附偏好”，需有参考文献支持。

#### 回应:

谢谢审稿人的建议。根据您的建议，我们补充了相应文献。

#### 意见 4:

前言中，“心理控制能力”首次出现应介绍。场依存和场独立型均应给出定义。

#### 回应:

谢谢审稿人的建议。根据您的建议，我们补充了相应内容。

#### 意见 5:

“认知方式图形镶嵌测验”应有一两句简介。

#### 回应:

谢谢审稿人的建议。根据您的建议，我们对相关内容做了补充。

#### 意见 6:

实验结果中, NP1 上高依附与另外两种条件的加工差异实际上从另一个角度支持了本研究的假设, 应展开讨论。

#### 回应:

谢谢审稿人的建议。根据您的建议, 我们对该现象展开了进一步讨论。

---

### 第五轮

#### 审稿人 3 意见

作者对文章进行了多次修改, 论文水平有了很大提高, 已经达到心理学报的发表标准, 建议修改下面问题后发表:

#### 意见 1:

去掉“尤其对于易受语境影响的场依存型被试来说, 该倾向更为明显。”这句话 (p. 40)。从数据上来看, 场依存型被试和场独立型被试在 NP1 上的加工并无差异, 不能说场依存型被试的倾向更明显。

#### 回应:

谢谢审稿人的建议, 根据您的建议我们删除了上述内容。

---

### 主编终审

#### 意见 1:

文中还存在一些细节的问题有待修改, 如 P30, “工作记忆容量与从句依附偏好存在显著负相关 ( $r = -.27, p < .05$ )”中,  $r = -.27$  中的负号写作不规范;  $r = -.27, p < .05$  应为  $r = -0.27, p < 0.05$  请补充全文中小数点前面的 0; 还有, 斜体的部分, 如 P33 中,  $MVP = 4.68 \pm 1.04$ ;  $MPP = 4.59 \pm 1.22$ , 斜体的仅为字母部分, 数字不需斜体, 建议根据《学报》的要求斟酌进行全文修改。

#### 回应:

谢谢主编的建议。根据您的建议, 我们对论文相应部分做了修改和完善。

#### 意见 2:

文中 F、t 以及 M 值后面数据的小数点有些保留 2 位, 有些 3 位, 显得比较凌乱, 规范性不够, 建议统一修改为小数点后 2 位。像文中表述  $p < .0001$ , 建议 p 值之后最多保留 3 位小数, 如  $p < 0.001$ 。

**回应：**

谢谢主编的建议。根据您的建议，我们对文中统计数值的表示方式进行了规范和统一。