

《心理学报》审稿意见与作者回应

题目：优势语序还是优势解读？利用 ERP 考察汉语双论元歧义句的解歧过程

作者：王路明

第一轮

非常感谢两位专家给出的评审意见，我们据此对原稿进行了大幅度的修改和补充。文中主要修改部分用黄色高亮显示。

审稿人一意见

《优势语顺还是优势解读？利用 ERP 考察汉语双论元歧义句的解歧过程》一文利用 ERP 考察了语顺和生命性对论元解歧的影响。该研究探讨了中文句子理解中重要的认知加工过程，具有一定的理论意义。但在问题提出、数据分析和讨论之中，皆有多处值得商榷。建议返回作者大修后再考虑是否可以接受发表于《心理学报》。

1. 前言对于问题的提出过于简略而快速地过渡到了假设。例如对于 Wang 等人 2009 年的研究本是该研究立论基础之一，相关例句、结果及机制却未作深入介绍。进一步，在提出“然而这仅限于 NP1 是歧义的情况下”时，应对词汇的生命性及其可能对论元指派所起作用做更为详细介绍。
2. 在假设部分又相对冗长不够简洁。事实上，假设部分应作为论元指派中的不同机制进行介绍，从而简化研究假设。
3. 前言中对于为什么要研究双论元句子也欠缺交代。
4. 正如作者在前言 (p4) 所说，OSV 和 SOV 在信息结构上有差异。如此说来，语顺本身和信息结构在该研究中是否相互混淆？
5. 假设二最后提出语言加工机制是遵循优势语顺还是优势解读有待研究，这类表述似乎无法合乎“假设”这一主题。建议调整说法。

回应：由于问题 1~5 都是针对原稿的前言部分，在此一并做出回答。在修改稿中，我们重新组织了前言和假设。

意见 1 和 3：我们补充了对 Wang 2009 实验例句和结果的介绍，并结合前人研究，引出研究双论元歧义句的动因，机制部分放在第二段。

意见 2 和 5：原先假设中的一部分内容作为两个语言加工原则介绍放在第三段；按生命性和语序可能发挥的作用重新整理了假设部分。

意见 4：非常中肯的点评。汉语的语序通常被认为是语用驱动的(pragmatic-driven, Li & Thompson 1978, p687)，语序和信息结构很难分离。我们在前言部分解释了 OSV 和 SOV 信息结构的区别，简化结构也包括简化信息结构，并且在讨论部分也对语境影响做出了说明(p11)。

6. 2.3 部分最后一句是讨论内容，不应该放在材料部分

回应：已修正。

7. 2.4 步骤部分对于任务的阐述和数据分析中的数据分析描述未很好对应起来。比如首先要求被试判断所

听到的问答组合是否自然，对应句子可接受度判断吗？而随后的判断对应的是句子理解判断？

回应：是的，先是句子可接受度判断，后是句子内容理解判断，已按数据分析顺序对文字部分进行了调整。

8. 2.5 脑电分析部分，为什么作者选用 0.3-20Hz 进行滤波？特别是 0.3 的选取，因为数值相对较高，可能使得部分慢波（如延续 300ms 左右的波）被剔除。

9. 脑电分析方法交代得过于简单。例如，基线是如何设置的都未有交代。伪迹排除标准也未交代。无法有效判断其做法是否合适。

回应：由于意见 8 和 9 都是针对本文的方法，这里统一作答。我们不采用基线纠正而采取 0.3-20Hz 进行滤波，是基于以下两点考虑：

这个方法得到了国际同行认可，相对成熟。采用这种数据处理方法并非我们首创，很多已经在国际刊物上发表的跨语言研究都采用了这种方法，汉语研究如 Philipp et al., 2008 *Brain and Language*; Bai et al., 2008 *Neuroreport*; 日语研究如 Wolff et al., 2007 *Brain and Language*; 德语研究如 Haupt et al., 2008, *Journal of Memory and Language*。为保证数据的可比性，我们沿袭了此法。其次，这个方法具有科学性，适合听觉实验。研究表明，语言加工的 ERP 效应主要和约在 0.5-4Hz 或 5Hz 的脑波频率有关 (Roehm et al., 2002, *Journal of Cognitive neuroscience supplement*)，选择这个波段不会造成有效信息的损失。事实上，在排除慢漂移方面，使用这个滤波和使用基线纠正的效果十分相似，但它又规避了使用基线纠正带来的关键区间 ERP 效应的改变（譬如有时基线纠正会导致 ERP 波幅增加造成“显著效应”）。另外，听觉实验的特质也决定了刺激的识别不可能在其呈现之前完全相同 (Friederici, Wang, Herrmann, Maess, & Oertel, 2000, *Human brain mapping*)。因此我们认为这种方法比较合理，在修改稿中进行了补充说明。

10. 2.6 部分为什么选择交互作用 p 值为 0.08 才会进一步分析而不是经典的 0.05？

回应：已改为 0.05。

11. p6 第二段最后“F1/F2<1”描述不清晰不规范。这里是指 F1 和 F2 的比值小于 1？还是两个 F 值都小于 1？多处“=”前后未留空格，需要作者通篇检查。

回应：是指 F1 小于 1 并且 F2 也小于 1，修改稿已将其分开显示。已通篇查阅等号前后空格。

12. 结果描述交互作用时，都采用了“X”，这是一个字母，而非乘号“×”，请使用数字符号插入功能。

回应：已修正。

13. 多处只报告交互作用而不报告主效应，理由是什么？

回应：修改稿中增加了主效应的报告。

14. 在描述 ERP 结果时，如 p7，“在 450-700ms 内，语顺主效应显著”，此时建议描述得更清楚，究竟是谁强谁弱，正波还是负波。

回应：数据分析部分（行为和 ERP）凡是涉及到这种表达方式的地方都按要求进行了修改，具体指出比较的方向和效应类型。

15. p7 第二段第一个 all ps，只有一个两两比较，无法用此说法；第二个 all ps，建议去掉 all。

回应：这个 all ps 原想表达“左前区的 p 值 大于 0.16，右前区的 p 值也大于 0.16”所以虽然只是一个两两比较，但写了 ps，不料给读者造成了误解。修改稿中采用了评审的解读，将 p 表达为一个两两比较的结果，ps 表达为两个两两比较的结果。

16. 讨论相对简略。例如，（1）对于为什么 700-850ms 的前额负波反映了加工成本（p8 第二段）所述过于简单，从而不容易评价这种解释是否合理；（2）在谈到假设二时（p9），作者指出“由此可见生命性在各个语言中的重要程度不尽相同，相对其他语言，它对汉语句子的影响比较大”。为什么中文需要依赖生命性？

回应：（1）修改稿中前言假设部分澄清了生命性和语序的竞争关系，此结果是和其呼应的。700~850ms 呈现的生命性和语序的交互作用为：在无生命 NP1 条件中 $S_{(in)}O_{(an)}V$ 比 OSV 诱发了前区负波，而在有生命条件中 $O_{(an)}S_{(in)}V$ 比 SOV 诱发了前区负波。产生前区负波的这两个条件 $S_{(in)}O_{(an)}V$ 和 $O_{(an)}S_{(in)}V$ 都违反了有生命性施事者这一原则，可见在加工晚期生命性最终强于语序对语言加工机制起主导作用。

（2）汉语依赖生命性可能是汉语自身特征造成的。修改稿讨论了汉语相关特征(Miao, Chen, & Ying, 1986; Li, Bates, & MacWhinney, 1993)，另外也补充讨论了在同一语言中，生命性在不同句型中的重要程度也不尽相同(详见 p11)。

17. 作者提出，当生命性和语顺不一致时，语顺的优势消失，部分项目遵从语顺，部分遵从优势解读。如果是这样的话，是否有依据表明，被试是如何在某些项目采取某一策略而在另一些项目采取某一策略。这里是否能够提供清晰的标准区分是两类项目，各会被采取不同策略进行加工？

回应：因为原文中对“生命性和语序不一致时”的假设部分表达不全，给读者留下了印象是 450~700ms 的数据主要通过被试在不同项目上采取不同策略来解释，事实上这不是本文的侧重点。在修改稿中，我们重新整理了假设 ii (p4)，并对此结果进行了基于假设的讨论 (p9)。我们认为 450~700ms 的有生命 NP1 条件没有产生加工差异，目前有两种可能：首先，回到设计这个实验的出发点，即观测生命性和语序在论元指派上的方向不一致的时候，也就是两者处在竞争状态的时候，语言加工机制是如何选择的。正如假设的那样，OSV 和有生命施事者在加工动词的早期阶段势均力敌，导致语言加工机制无法做出特定的论元指派。其次，也有可能语言加工机制做出特定的论元指派，但这一指派随着测试的变化而变化，并不稳定。这个可以通过 item split 分析进行验证。但是需要指出的是，即使证明了同一条件下不同项目的句子，语言加工机制会做出不同的指派，平均后的 ERP 仍然是两条件总体持平，不影响当前结论。因此，我们认为无论是语言加工机制在加工早期缺少明显的论元指派导向，还是导向随项目而变得不稳定，结果都说明 OSV 语序和生命性共同参与 NP1-NP2-Verb 的在线加工，它们影响语言加工机制的力量总体持平(p9~10)。所以我们倾向于保留两种解释。

18. 在图 2，似乎看到非生命性语境下，两个条件之间在基线附近仍然存在显著差异。这似乎说明，在非生命语境下，还存在更为早期的效应。作者应做数据检查、讨论。

回应：的确在图 2A 非生命语境下，两条件在基线附近 200~400ms 处看似有早期效应，之前我们做过数据检查，但检查结果显示这个语序差异并不显著(侧区： $F < 1$ ；中线：F > 1 但 $p > 0.08$)。如果做基线纠正，这两个条件在基线附近重合，导致关键区间前区负波“更加显著”。

实际上，使用基线和不使用基线都不会改变负波产生的方向(在有生命条件下，仍旧是 OSV vs SOV；在非生命条件下，仍旧是 SOV vs OSV)，因此我们认为本实验结果是稳定的。在回应意见 9 中我们已经阐述了不使用基线的理由，所以选择检查早期效应是否显著排除其他可能。

19. 进一步地，非生命与生命语境下，效应出现的早晚似乎有重要意义。建议作者进行讨论。

回应：谢谢建议。修改稿就早期加工和晚期加工的数据对语言加工模型有何启示进行了探讨（详见 p11 最后一段）。

20. 大部分情况下似乎都用“语序”比“语顺”更常见。

回应：已改为“语序”。

审稿人二意见

本研究通过运用行为实验与 ERP 实验考察了汉语双论元歧义句的解歧过程，并提出了具有一定新意的解释，具有一定的理论意义。以下问题与作者讨论。

1、“结构简单性”是该研究的非常重要的概念，作者认为结构简单性决定了是主语优势还是宾语优势。但审稿人认为作者没有把这一核心概念讲清楚，比如为什么 OSV 比 SOV 的结构简单等？

回应：谢谢建议。请参见对评审一意见 4 的回应。在修改稿的前言部分，我们补充解释了 OSV 和 SOV 信息结构的区别(p4)，简化结构也包括简化信息结构，并且在讨论部分也对语境影响做出了说明(p11)。

2、前言最后一段中“根据假设，我们预测 OSV 语顺与生命性会产生交互作用：当句子解歧为 $O_{(无生命)}S_{(有生命)}V$ 时，它同时满足简单结构和有生命施事者的要求，产生加工优势；当句子解歧为 $S_{(无生命)}O_{(有生命)}V$ 时，同时违反这两个要求，产生加工成本；而当句子解歧为 $O_{(有生命)}S_{(无生命)}V$ 或 $S_{(有生命)}O_{(无生命)}V$ 时，或满足简单结构或满足有生命施事者要求，语言加工机制是遵循优势语顺还是优势解读，有待本文实验揭晓。”所谓的加工优势或加工成本，在行为结果与 ERP 结果表现的具体数据模式是什么样的，可以在此先做出预测。后面的统计结果不是非常清楚地呈现 $O_{(无生命)}S_{(有生命)}V$ 与 $S_{(无生命)}O_{(有生命)}V$ 两种条件下的结果，读者需要付很大的努力才能去把握。另外，何谓优势解读？应该有所交待。该条件下究竟是哪种优先，作者是否应该有预测？

回应：谢谢建议。在修改稿前言部分，我们重新组织了假设的表达方式，明确了实验可能出现的数据模式(p4 最后一段)。我们对后面的统计结果的表述也进行了更正，具体指出比较的方向和效应类型，前后统一表达方法。另外，我们在前言开头部分，紧随优势语序，对优势解读进行了前人研究的说明和解释(p3)。假设部分对生命性和语序指派论元方向不一致时进行了补充说明(p4)。

3、2.3 的材料介绍中，在音强上，NP2 和 Verb 的差异从平均数来看，似乎大于 1dB，不像作者所说的小于 1dB。

回应：由于原文表述不清，给读者造成了误解。声学分析的目的是想控制各个句子成分的声学属性在四个条件(cross conditions)没有明显的差异。因此不是横向比较 NP2 和 verb，而是纵向比较 NP2 在四个条件的差异(即比较 62.1, 62.3, 61.9 和 62.7)，和 Verb 在四个条件的差异。已修改表达方式。

4、ERP 实验中，没有说明兴趣区选择的依据。

回应：已修改，是依据前期汉语 ERP 研究的分法。

5、统计结果没有提供 effect size。

回应：已补充。

6、讨论部分未能就该研究的结果的深层意义进行分析，比如对心理语言学模型的修正或建立的理论意义进行讨论等。

回应：在修改稿中补充了这方面的讨论(详见 p11 最后一段)。

7、前言部分假设一中的“跨语言模型 the eADM”的表述方式正确吗？全称是什么？

回应：已修改，全称是“扩展的论元依存模型(The extended Argument Dependency Model)”。

第二轮

感谢两位评审的意见。文中主要修改部分用黄色高亮显示。

审稿人一意见

1 摘要：去掉“和其他语言相比，”，研究中没有直接对比，这一表述不大合适

回应：已删去这句。

2 摘要：语言层次不清晰。“通过听觉 ERP 实验揭示双论元解歧过程中受到两个因素的影响：”这一表述似乎是结论/结果性的，而以这样的句子为第一句，不妥。“被试听到 NP1_(无生命)NP2_(有生命)V 和 NP1_(有生命)NP2_(无生命)V 两种序列。”与前后的衔接也不够紧凑。请考虑修改。

回应：已修改整个摘要。

3 前言：“然而，这仅限于只有一个歧义论元(NP1)出现的情况。”请指明是何种语言的研究结果？

回应：已注明语言和出处。

4 前言：“Left-detached position 和代表焦点的 pre-core slot”请提供中文翻译。

回应：已添加中文翻译。

5 回复第一次审稿意见 18 条时，作者提出“如果做基线纠正，这两个条件在基线附近重合，导致关键区间前区负波“更加显著”。”建议作者在文章中报告前面时间窗对后面时间窗效应的可能影响，并解释是否影响当前结论。

回应：谢谢建议，请参见脚注 1。

审稿人二意见：

作者对审稿人提出的意见进行了逐条答复，审稿人基本认可。但在一修稿件中还存在以下一些写作方面的问题，请作者再认真通读全文。

1、前言的第二句话不通，“问题诸如句子理解—从声音、文字或手势等简单的语言形式到复杂的句子意义—这一过程是否在所有语言中都相同，还是因语言的不同而不同。”

回应：已修改。

2、前言部分最后一句话遗漏了最重要的一个词“高或大”，“若遵从优势语序，则违发该语序的 S_(有生命)O_(无生命)V (优势语序-, 优势解读+)应比 O_(有生命)S_(无生命)V (优势语序+, 优势解读-)产生加工成本（高或大）(p10)

回应：已修改。类似表达之处也做了修改。

3、讨论部分 4.2 中第二段最后部分“也有观点认为，实验中的问句总是主题化 NP1，OSV 语序优势或许来自语境需要。因为汉语是一个主题性显著的语言，一个主题化 NP1 的语境可能对整个 NP1-NP2-Verb 的加工产生影响，譬如它可能使语言加工机制从一开始就放弃了 NP1[-dep]的分析。这种观点倾向于将语境因素视为影响句子早期加工的因素之一(如同句子结构模板)。然而本实验句子包含两个歧义论元，NP1 上的语境影响不太可能在句末动词处还能观测到，另外 NP2 的出现也可能改变语言加工机制对 NP1 的解读，因此是结构还是语境产生 OSV 优势还需今后通过操作不同的语境才能进一步分离”，审稿人不清楚这种观点是在哪里提出的，因此需要文献出处？

回应：谢谢指出。这一观点是根据多重因素平行加工模型提出的。该模型与层级加工模型相左，认为非结构因素（包括语义因素和语境因素等）可以干预语言的早期加工，而无需等在结构因素之后。修改稿中添

加了文献出处。

第三轮

经过两轮审议，作者对审稿人的意见作了较全面答复，稿子经过修改有较大的改进。但现仍有以下两个问题需要作者补充进行回答：

1、作者谈到实验任务中，一个任务是句子接受度判断，要求判断所听到的问答组合是否自然(是，选择笑脸图标；否，选择哭脸图标)，请问何为问句组合是否自然？这一任务很难理解，与传统常用的接受度任务很不一致，请作者补充使用这一任务的原因。此外，句子内容理解任务中，请举例说明让被试回答的问题是什么？

回应：这确实是一个值得说明的问题。通常，我们通过句子接受度任务控制被试能否探测到违反（如句法或语义违反），甚至精确到违反程度。但在本实验中，我们只设计“是”和“否”两种选择，是基于两个考虑：第一，这四个条件句（SOV和OSV）都是汉语合法的句子，也不存在明显的语义违反（见回应2），它们之间的差别十分细微。能够根据语境体会出句子的细微差别做出选择，即使这种选择只是二选一，就已经对被试的语感提出了很高的要求。第二，同一事件可以有不同的表达方式，比如，同样是表达“侦探保存了子弹”这一事件，可以有两种说法，子弹侦探保存了，或侦探子弹保存了；同样是表达“子弹击中了侦探”这一事件，也可以有两种说法，子弹侦探击中了，或侦探子弹击中了。实验说明中，我们要求被试在听完问答组合之后，根据语境尽可能地区分自己认为自然的和不自然的回答。因此，被试的选择看似绝对，却是种相对选择，即在特定语境下，哪种说法更加自然。

从接受度任务的执行结果来看，被试能够完成这项颇具挑战性的任务，甚至比我们预想的更加出色，四种条件句呈现出四种完全不同的接受度。同见Haupt et al. 2008, *Journal of Memory and Language* 在德语中，SOV和OSV都为合法语序，也不存在明显的语义违反，但被试对OSV的接受度显然没有SOV高。而且，他们对S，O的语义特点十分敏感，能够据此做出不同的区分。因此，我们认为本实验的接受度任务数据是比较成功的。

下面是句子理解任务的例子。比如，针对 OSV 条件句(a)的内容，设计提问 (b) 或 (c)。根据句子内容，(b) 或 (c) 的回答都为“是”，回答为“否”的提问也使用主动或被动句式，只是这些问句包含不同的名词或动词，如使用“保姆”而不是“侦探”，使用“丢弃”而不是“保存”。回答“是”和“否”的提问在每个条件中都数量相等。

a. OSV 条件的例句

子弹怎么了？子弹侦探保存了。

b. 或使用主动句对其提问

侦探保存了子弹吗？

c. 或使用被动句对其提问

子弹被侦探保存了吗？

这些我们在 2.4 步骤处进行了补充说明，用黄色高亮显示。

2、在实验的例子中，子弹侦探击中了如果完全按句法分析 OSV 来理解其实与常识的符合度不如其他三种条件，这种效应可能会增强生命性的效应而减弱语序的效应，而使结果出现偏差，不知这是否为实验材料选择的特例还是在大部分材料中普遍存在这种现象。由于作者在本文中主要关心语序效应与生命性

效应，理应控制两种效应下意义的合理性程度保持均衡。作者需要补充一个语义评定任务，要求评定者按他们的理解用正常语序写出句子，并同时给出句子的语义合理性的评定。如果评定者对句子的理解有误，或他们对某一条件下材料的语义合理性评定得分低，这可能影响整体实验结果的可靠性。作者需要对此因素可能造成的影响进行全面评估。

回应： 我们赞同评审，这是个重要的问题。评审认为，某些测试（**trials**）若按照OSV策略会形成不合理的解读，因此会削弱对语序的依赖，所观测到的效应不能排除这种不合理因素的影响。然而我们认为，这四个条件句的加工不太可能受到这种影响（参见**implausibility processing**）：

首先，我们通过句子内容理解任务对不合理的解读进行了控制。如果被试对任何一个条件产生不合理的解读，会导致句子内容理解任务的正确率低下。如果被试按OSV语序将“子弹侦探击中了”错误解读成侦探击中了子弹，那么面临“子弹击中了侦探吗？”这样的提问时，他们会误按“否”。然而，在本实验中，四个条件的句子内容理解正确率都在90%以上，这说明，被试能够正确理解题元关系。另外，针对句子内容的提问，有的采用主动语态，有的采用被动语态，被试只有正确理解题元关系，才能获得如此高的正确率。

其次，我们只对句子理解正确的测试进行ERP分析，即使被试对个别句子产生了不合理的解读，这些句子已经被作为理解错误的测试排除在ERP数据分析之外，因此所见ERP效应已经剔除不合理的解读的测试。

最后，有关合理性加工的前期ERP研究表明，合理性加工或诱发广泛分布但主要是中央顶部位的N400（van de Meerendonk et al., 2010 *Journal of Cognitive Neuroscience*, Bornkessel-Schlesewsky et al., 2011 *Brain and Language*）或诱发语义P600（见综述 Kuperburg et al., 2007 *Brain and Language*）。而本实验中观察到的是一种负波，且分布于头皮前区，因此我们认为该波与非合理性效应有本质的不同。

对此，我们在讨论中进行了补充说明,用黄色高亮显示。