

《心理学报》审稿意见与作者回应

题目：汉语确定性无界否定句模拟加工的时间进程

作者：崔如霞 高志华 唐艺琳 何皓璠 鲁忠义

第一轮

审稿人 1 意见：该文在对否定句的类型进行新的分类基础上，对其中一种类型即确定性无界否定句的加工进程进行了探讨，发现了不同于以往的结果，并提出了是否定句类型不同导致不同结果的观点。研究具有一定的价值，以下问题供作者参考：

意见 1：关于否定句的分类问题，作者分为有界否定与无界否定，无界否定又分为确定性无界否定与不确定性无界否定，从文章的介绍来看，有一定的合理性。以往研究（陈广耀等，2014）把否定分为状态否定与动作否定，状态否定又分为状态确定与状态不确定否定两类，似乎也有其合理性。审稿人认为，该文章提出的分类似乎只针对了状态否定，请问作者该文章的分类方法与以往分类方法的区别在哪？

回应：谢谢审稿专家的意见和建议。关于否定句的分类，从我们查阅的大量文献看，目前尚无统一的标准。最近陈广耀等（2014）将否定句分为状态否定句与动作否定句；根据状态的连续性特性又将状态否定句区分为类别型状态否定句和连续型状态否定句；进一步又根据能否推断出事件的实际状态把类别型状态否定句分成类别型状态确定否定句和类别型状态不确定否定句。这种分类较为全面、合理，因此这次修改我们参考了他们的分类。我们的分类是按照有界性对否定结果的分类，是陈广耀等所说的连续型否定句。本文主要是对否定结果的状态的分析，没有涉及动作否定等其他情况。

意见 2：能否把 250ms、750ms 与 1500ms 作为一个变量的三个水平，即把三个实验结果放在一起分析，看看结果如何？可能对结论的可靠性具有帮助。

回应：将三个独立的实验做综合检验是可以的，但它毕竟是三个独立的实验，而且我们认为之所以独立开来做研究，就是想清楚地了解否定句的加工阶段，既然已经到达了我们的研究目的，所以就没有综合起来分析其交互作用。

意见 3：实验 1、2 的句子类型与匹配性只有在被试检验显著，项目检验不显著，作者在讨

论中需要解释为什么项目检验不显著，这对结论是否可靠具有重要影响。

回应：心理语言学的研究需要做以被试为随机变量的统计检验，也要做以项目为随机变量的检验。本研究的实验 1 和实验 2 中，被试检验显著，而项目检验不显著，从国内外的文献看，这种情况也有。被试检验显著而项目检验不显著的情况下，虽然可以下结论，但也暗示着，我们的材料取样可能还有局限性，或者说实验材料的样本还不够大。

意见 4：为什么实验 1 的数据删除比例高达 9.5%？

回应：实验 1 的数据删除比例高可能是因为简单随机取样的原因，也可能是实验过程中刺激呈现的时间较短（250ms）。虽然数据删除的比例高了一些，但还是可以接受的。

意见 5：建议在前言部分把实验逻辑与预测清楚写出来，便于读者理解。

回应：感谢审稿专家的建议，我们在前言部分增加了三个实验的实验逻辑。为了便于专家审稿，复制如下：

实验 1 考察汉语确定性无界否定句理解早期的心理模拟。由于确定性无界否定句既有有界的一面，又有无界的一面，因此实验的逻辑应该是，它的加工可能不是一步完成的，而可能是依据否定句加工的两步模拟假设，先模拟被否定状态，再模拟实际状态。这样，可以预测的实验结果是，句子理解的早期对与被否定状态匹配的图片的判断的反应时会显著短于与否定句实际状态匹配的图片的反应时。实验 2 和实验 3 分别考察汉语确定性无界否定句理解中期和晚期的心理模拟。根据上述实验 1 的分析，汉语确定性无界否定句的加工如果是符合两步模拟假设的，那么实验 2 和实验 3 的实验逻辑是，如果在中期已经完成心理模拟，那么，否定句和实际状态相匹配条件下图片判断的反应时会显著快于与实际状态不匹配条件下图片的反应时，否定句加工完成，实验 3 只是对它的确认。如果实验 2 还没有出现实际状态加工的优势效应，实验 3 就是要考察汉语确定性无界否定句的加工是不是在晚期才完成。

意见 6：以往的一个文献中作者是“陈广耀”而不是“陈光耀”

回应：是我们笔误所致，向作者致歉！已经改过来了。

审稿人 2 意见：本研究以汉语确定性无界否定句为研究对象，采用句-图匹配范式探讨汉语确定性无界否定句模拟加工的时间进程。研究发现，确定性无界否定句理解的心理模拟是分两步进行的，首先模拟的是否定句中的被否定状态，随着加工时间的推进，完成对实际状态的模拟，但模拟过程相比于其他类型的否定句所用的时间更短。本研究采用汉语确定性无界否定句作为研究对象，和前人已有的研究结果形成对比，发现了不同的研究结果。研究结果有意义，研究方法恰当，数据分析合理，研究结论恰当。以下意见供修改参考：

意见 1：建议对题目进行适当的修改。现题目为“汉语确定性无界否定句的加工为什么在中期完成：否定句类型的作用”。由于本研究并没有直接操纵否定句的类型，题目容易引起误解，切题目偏长。建议修改为“汉语确定性无界否定句模拟加工的时间进程”或其它更为合适的题目。

回应：接受审稿人的意见，作了修改。

意见 2：给出一些填充句子的例子，以帮助读者更好的理解实验材料。

回应：我们作了修改，增加了例子。肯定的填充句如“凉席铺好了”，否定的填充句如“和平鸽没有飞走”。

意见 3：剔除“每个被试在各个条件下 3 个标准差以外的数据”，应该修改为“每个被试在各个条件下平均数加减 3 个标准差以外的数据”

回应：谢谢审稿专家的意见，作了修改。

意见 4：“句子类型和匹配状态的交互作用材料的方差分析差异不显著 $F_2(1,23) = 0.046$, $p > 0.05$ 。”应修改为“句子类型和匹配状态的交互作用，以项目为变量的方差分析差异不显著 $F_2(1,23) = 0.046$, $p > 0.05$ 。”

回应：谢谢审稿专家的意见，作了修改。

第二轮

审稿人意见：文章作者针对审稿人提出的意见与建议，对全文进行了相应的修改，并对审稿意见进行了逐条答复，审稿人较为满意。另外提出一个建议，否定句的加工是非常复杂的一个过程，建议做结论时，采用更加谨慎的表述方式更加合适。同意修后采用。

回应：接受专家的意见，我们对全文做了进一步的修改，特别是讨论和综合讨论部分用了 11 个“可能”来阐述。以此来表明，否定句的理解是复杂的，我们的结论是审慎而下的。（三审意见后我们又做了进一步的修改，见“三审意见及修改”）

第三轮

编委意见：作者多次将该研究和陈广耀等（2014a；2014b）的研究结果进行比较，作者的研究是发现在中期就完成了对实际状态的模拟，而陈广耀的研究则发现要到晚期才出现实际状态的模拟，作者认为这个差异主要来自于否定句类型的差异。虽然并不排除这一可能性，但由于作者在实验中并没有直接对照两类否定句，两类否定句本身所使用的词汇在词汇特性等方面也必然存在差异，在未精细控制两类材料差异的基础上直接对比不同的研究结果，尤其在时间进程上直接比较并做出推论可能过于草率。作者立论的关键点应该还是在否定句的解决是一阶段还是两阶段完成，而不应该是时间进程上的差异。建议作者在推论时间进程的差异时应更为谨慎，并且考虑其他的可能性，如被试或句子材料难度等其他可能的影响变量，并在讨论中加以论述。

回应：接受审稿专家的意见，由于我们没有在此次研究中直接控制否定句的类型，这样在得出否定句类型直接影响实验结果的说法就显得草率。对此我们做了如下的修改：

（1）在讨论的行文中不再强调否定句类型的作用，提及的地方也只是说是一种可能。

（2）在与陈广耀等（2014a；2014b）的比较中，虽然说与其研究结果不同，但原因正像审稿专家所说那样这只是一种可能性。因此我们改为“这与本研究得出的在中期就完成了对确定性无界否定句的加工是不同的。我们推测可能与两项研究使用的材料的单维性和多维性以及否定结果的确定和不确定的差异有关。不过这需要做进一步的研究来证实。”

（3）将摘要和结论中的“否定句的类型直接影响了否定句理解的心理模拟过程”这句过度推论的话删除了，主要强调本研究得出的确定性无界否定句理解的心理模拟是在中期完成的。

第四轮

主编意见：1.请就如下两点进行进一步说明和修改完善。 1、论文实验并未得到统计上可靠

的结果。具体地说，实验 1 和 2 中，图片判断反应时数据统计分析发现，句子类型和匹配状态的交互作用，仅仅在以被试为随机变量的检验中达到显著，而在以项目为随机变量的检验中并不显著（两个实验中， F 均小于 1）。这说明，文章所报告的结果以及在这些结果上推论的结论，可能仅限于实验中研究者所采用的特定的语言材料，而并不能推广到其它实验材料上。换句话说，如果研究者采用不同的语言材料样本，研究发现可能不能重复。对于以语言材料为刺激的实验而言，以被试为随机变量和以项目为随机变量的统计检验都要达到统计学上的显著性，是基本（甚至是最低）要求。更严格的要求是在 F_1 和 F_2 基础上，计算 $\min F'$ 。

回应：谢谢主编的意见！

本研究的数据确实是以被试为随机变量的统计检验达到了显著水平，而在以项目为随机变量的检验中并不显著。对主编提出的这一问题我们的解释和对文章的修改如下：

（1）我们承认本研究中确实存在着以项目为随机变量的统计检验不显著的问题，像主编所说，这可能不能推广到这类材料的总体上。分析其原因可能是汉语确定性无界否定句本身的复杂性所致。比如“门不宽”，对于理解者来说“不宽”到底是多宽可能不易确定。另外，否定句的分类至今尚没有统一的看法，Paradis et al（2006）与 Bianchi et al.（2011）的否定分类研究中也承认否定分类的复杂性和争议。这说明理解确定性无界否定句这种语言现象是一个复杂的过程，今后研究中需要在实验材料的编制以及实验设计上做出更大的努力。

（2）已有的心理语言学文献中确实也有以被试为随机变量和以项目为随机变量的统计检验不一致的地方，这种情况下就采信了以被试为随机变量的统计检验结果。不过这提示读者这样的研究结果是有一定局限性的。这方面的文献我们列举几篇如下：

Kaup B., Lüdtke J., Zwaan R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics*, 38, 1033–1050.

Autry, K. S., & Levine, W. H. (2012). Activation of negated and non-negated entities. *Journal of Pragmatics*, 44, 1474–1485.

张积家, 陈穗清, 张广岩戴东红. (2012) 聋大学生的词汇习得年龄效应. *心理学报*, 44 (11), 1421-1433.

麦穗妍, 陈俊. (2014) .非熟练潮-粤双言者的语义通达：来自听觉词加工的证据. *心理学报*, 46 (2), 227-237.

（3）根据主编的意见我们对综述做了修改（见综合讨论最后一段），为了便于主编审阅，

抄录如下：

需要说明的是，本研究还存在一定的局限性，即在否定句加工的 250ms 时被否定状态的优势效应和 750ms 时的实际状态的优势效应都只出现在被试检验中，而没有出现在项目分析中，因此，对本研究的结论应持更为审慎的观点。我们认为这一结论可以推广到人的总体中，而不是材料。另一方面，项目检验并没有出现实验预期的显著效应，也恰恰反映出汉语确定性无界否定句心理加工的复杂性，不同的否定词以及不同的否定辖域在加工过程中都可能存在某些差异，从而影响项目分析的结果，这值得进一步的探讨。

意见 2：三个实验之间的主要差别在于句子与图片呈现的时间间隔（分别为 250ms、750ms 与 1500ms）。这样，将“时间间隔”（或“实验”）作为一个因素，进行三因素（另外两个因素分别为句子类型和句图是否匹配）全方差分析，考察句子类型和句图是否匹配的作用是否受间隔时间调整，非常必要。文章没有进行这样的分析，尽管一位审稿人已经提出了这个问题。

回应：接受主编意见，文中已经加上了全方差分析，并发现了句子类型、匹配状态与时间间隔三者的交互作用，即不同类型的句子的匹配效应确实受到间隔时间的调整作用，即在 750ms 时，图片再认时间最短。这与独立分析三个实验的结果是一致的，不过全方差分析的结果更能说明问题。谢谢主编！