

句子阅读理解过程中句意的建构时间*

钟伟芳 莫 雷 金 花 徐贵平

(华南师范大学心理应用研究中心, 广州 510631)

摘 要 一般语言学理论认为, 在句子阅读理解过程中, 句子意义建构是基于词汇语义整合的, 其反映在 N400 之上, 开始于词语呈现后约 250 ms。然而, 近年关于语义 P600 效应的研究却提示句意建构可能存在多通道或方式, 读者甚至可能在 N400 反映的加工出现之前就已建立了初步句意。为探明在句子阅读理解过程中是否存在比 N400 反映的加工更早的句意建构, 以及如果存在这样的句意建构, 其是从何时开始的, 本研究开展了四个实验。实验 1 发现被试在句末双字词呈现 200 ms 后已将其联系到上文语义表征而建立了句意, 提示句意建构开始于 N400 反映的加工出现之前。实验 2 与 3 进一步发现, 读者开始建立句意的时间约在句末词呈现了 150 ms 之时。实验 4 排除了实验 1~3 效应的其他一些可能解释, 提升了实验 1~3 结果的可靠性。总体上, 本研究提示: 在句子阅读理解过程中, 读者在句末双字词呈现了约 150 ms 之时就已建立了句意; 可能存在比 N400 反映的加工更早的句意建构。

关键词 句子理解; 意义建构; 时间特征; N400

分类号 B842

1 问题提出

在句子理解过程中, 读者是何时将新进入的词语联系到上文语义表征而建构句子意义的呢? 为探明此问题, 大量研究探讨了句子理解过程中各种信息加工的时程、方式等问题。

经典语言加工模型认为, 句子理解过程中新进入词语的加工需经过词汇通达、词汇选择与词汇整合等三个阶段(Kintsch & Mross, 1985; Till, Mross, & Kintsch, 1988; Hinojosa, Mart ín-Loeches, & Rubia, 2001), 而加工需到达词汇整合阶段才能建立句子的统一意义。那么, 词语呈现后何时会得到整合呢? 事件相关脑电位(Event-related Potential, ERP)成分 N400 的发现与应用很好地揭示了这一问题。

N400 由 Kutas 和 Hillyard (1980)发现。他们给被试逐词呈现句子, 发现不能整合的句末词引起了 N400 效应, 即相对于正确句, 语义违背句的句末词在 250~550 ms 诱发了波幅更大的负成分 N400。基于大量研究, 许多研究者认为 N400 是主要与语

义信息整合相关的成分, 反映着句子理解的语义整合过程。然而, N400 时间窗有数百毫秒, 读者是何时开始建构句子的统一意义呢? 许多研究者认为读者在语义整合一开始就建立了句意, 也就是说 N400 始点就是建立句意的始点。如 Hagoort 等探讨句意及其真实性建构的时间关系时就把问题转为探讨语义知识与世界知识整合的时间关系。他们让荷兰语被试阅读正确、世界知识违背和语义违背三种版本的句子, 如“The Dutch trains are yellow/white/sour and very crowded.”(英语翻译句。下划线标出的为关键词)。结果表明, 两类违背句都诱发了显著的 N400 效应, 且两者在潜伏时、波幅及脑地形图上都相似。据此, 他们认为在句子理解过程中, 语义和世界知识是同时整合的, 相应地, 句意及其真实性的建构也是同时的(Hagoort, Hald, Bastiaansen, & Petersson, 2004)。

随着研究的深入, N400 反映语义整合的观点为越来越多研究者所接受, N400 也被广泛地用于探讨句子意义加工的实质。有赖于此, 句子意义加

收稿日期: 2010-08-29

* 国家自然科学基金项目(31170997)资助。

通讯作者: 莫雷, E-mail: molei@scnu.edu.cn

工的实质,尤其是时间特征得到了越来越清晰的揭露。然而,近年关于语义 P600 效应的一系列研究却指示句子意义建构可能存在语义整合以外的其他通道或方式。也就是说,这些研究提示,基于 N400 的研究可能仅部分地揭示了句子意义加工的实质。这使得句子意义加工的实质又重新模糊了起来。

例如, Kolk 等让被试阅读如“De kat die voor de muizen vluchtte.....”(猫从老鼠身边逃走.....)的荷兰语句子(下划线标出的为关键词)。这些句子描述了与事实相反的事件。数据表明,这一类语义违背未诱发 N400 效应而是诱发了 P600 效应。这是意外的发现,因为一般认为 N400 反映语义加工,而 P600 则反映句法加工。为什么会出现这一现象呢?他们认为,在句子理解到达语义整合之前,大脑会通过概念倾向(conceptual bias)和句法结构(syntactic structure)两条通道各自建立对句子的语义解释。由于两条通道建立的语义解释存在冲突,对于例句是概念倾向把句子解释为“老鼠从猫身边逃走.....”,而句法结构分析则解释为“猫从老鼠身边逃走.....”,大脑因怀疑是否正确地加工了句子而重新注意这一意外事件并检查其真实性。正是这一检查操作(checking operation)阻碍了大脑将意外事件整合到当前语义表征中的尝试,引致了 N400 效应的缺失,并表现为 P600 效应(Kolk, Chwilla, van Herten, & Oor, 2003)。进一步,研究者认为 P600 反映的并非纯句法再分析加工,而是更广泛意义上的语言理解过程中的认知监控(Kolk et al., 2003; van Herten, Kolk, & Chwilla, 2005; Kolk & Chwilla, 2007)。

对于语义 P600 效应,有其他学者提出了不同的见解。如 Kuperberg 认为句子加工至少存在两条并行的通道:基于语义记忆的通道(semantic memory-based stream)和联合通道(combinatorial streams)。前者负责处理句子实词之间的语义关系,并将其与语义记忆中储存的信息相比较。当句子语义违背时出现 N400 效应。后者不仅对词素句法信息敏感,而且对生命性等语义——题元限制敏感。当联合通道中题元加工的分析结果与句法加工的分析结果相冲突,或是联合通道的分析结果与基于语义记忆通道的分析结果相冲突,联合通道就会进行持续的分析,从而表现为 P600 效应。在理解上述句子时的情况是基于语义记忆和题元分析把句子解释为“老鼠从猫身边逃走.....”,而句法分析则

解释为“猫从老鼠身边逃走.....”,两种不同解释的冲突引起了联合通道的持续分析,从而关闭了或者掩盖了 N400 所反映的整合加工(Kuperberg, 2007; 王瑞乐, 李妮, 陈宝国, 2010)。

尽管学者对语义 P600 效应有不同的解释,但这些研究给了我们两个方面的提示:一是在句子理解过程中,句子意义建构可能存在语义整合以外的其他通道或方式;二是在 N400 反映的加工出现之前大脑可能已建立了句子的初步语义解释。事实是否如此呢?如果回答是肯定的,那么以往基于 N400 的研究可能就仅或主要地揭示了语义整合这一通道或方式的加工,而其他通道或方式的加工,以及意义建构的时间特征问题则未能很好地揭示。因此,为进一步揭示句子意义建构的实质,很有必要探明是否存在多通道或方式的语义加工,以及是否存在比 N400 反映的加工出现得更早的句子意义建构。本研究对后一个问题进行了初步的探讨。

具体来说,本研究关注两个问题:1、读者是否在 N400 出现之前已把新进入的词语联系到句子上文语义表征而建立早期的句子意义?2、如果对问题1的回答是肯定的,那么读者具体是在何时开始建立句子意义?

2 实验 1

实验 1 旨在探讨第一个问题,即读者是否在 N400 出现之前就已把新进入的词语联系到句子上文语义表征而建立了句子意义。

依上所述, N400 可能并不能完全地反映句意加工过程,因此找到另一种指标或方法探测读者开始建立句意的时间是研究的首要问题。那么,应该给予被试什么实验任务以即时探测出其是否已建立了句意呢?另一方面,语义 P600 效应的相关研究也提示,建立句意可能是分步进行的。如依 Kuperberg 的研究,在词语呈现后大脑有可能先通过基于语义记忆通道及联合通道建立句子的初步语义解释,在评定不同通道的各种解释之间无冲突后,再启用语义记忆系统中的各种语义相关信息对词语进行整合(Kuperberg, 2007)。如果事实如此,那么大脑可能并未将早期建构的句子语义解释与语义记忆系统中存贮的语义相关知识相对照而评定其合理性。因此,要探测出早期建立的句意,还需要避免让被试对建立的句意作任何合理性评估。要设计如此实验任务是困难的,幸运的是知觉符号理论相关的实验范式为研究提供了可能。

关于人类如何对信息进行编码和贮存的问题,经典的命题符号理论(Propositional Symbol Theory)认为信息在头脑中是在一个命题符号系统中编码和贮存的(Kintsch, 1998)。近年, Barsalou 等提出了一个新的观点——知觉符号理论(Perceptual Symbol Theory), 认为命题表征和知觉表征同处于一个神经网络, 符号与其指示物存在类比的关系(Barsalou, 1999a, 1999b)。Zwaan 等人的研究支持了知觉符号理论(Stanfield & Zwaan, 2001; Zwaan, Stanfield, & Yaxley, 2002)。他们让被试阅读一系列句子, 如“*There is an eagle in the sky*”(天空中有一只老鹰), 然后判断其后呈现的图片中的物体是否在句子中被描述过。判断的标准是不管图片中的物体与句子描述的在状态上是否一样, 只要是同一物体就判断为“是”。图 1 是与例句对应的一对图片, 按照标准两幅图片都应判断为“是”。

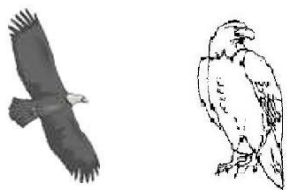


图 1 Zwaan 等(2002)实验中的一对图片

研究者认为, 如果符号与其指示物间存在类比关系, 那么不匹配效应(mismatch effect)将会出现。也就是, 如被试阅读上述例句后, 其判断图 1 中左图的速度会快于右图, 因为前者描绘了与句子描述的状态相匹配的“飞鹰”, 而后者则描绘了不匹配的“站鹰”。数据支持了他们的假设。

两种理论之争是很有意思的问题, 但本研究更关注的是上述的研究范式。我们认为, 在 Zwaan 等的研究中, 如果不匹配效应出现了, 则可以推断出读者已经建立了句意。因为如读者未将句子的关键词, 如“sky”联系到上文语义表征中, 其不会发现两幅图片的差别, 两幅图片的判断反应时应无差异。因此, 不匹配效应的出现意味着读者已经把句末词联系到了上文表征, 建立了句意。而且, 被试要顺利完成句图匹配判断任务只需建立句子的初步意义, 而不必对其合理性作出判断。这一实验范式很好地满足了上述的限制, 可以运用于探讨本研究的问题。

实验 1 将以上述范式给被试呈现一系列句子与图片。句子逐词呈现, 句末词为实验关键词。句末

词前的每个词语均呈现 800 ms, 以保证到达句末词时被试有足够的时间理解句子上文。依照 N400 的时间窗设置句末词的呈现时间。句子呈现完毕后立即呈现图片。据前人研究, 句末词呈现 550 ms 后, 即 N400 结束之时, 读者已建立了句意, 因此可以预期当句末词呈现 550 ms 时被试的图片判断会出现不匹配效应。如果当句末词呈现时间短于 N400 的潜伏时间时也出现了与呈现时间为 550 ms 时相似的不匹配效应, 即句图一致条件下的反应时短于不一致条件下的, 且相同一致性条件下的反应时无差异(如果两种呈现时间下都出现了不匹配效应, 但反应时存在差异, 很可能是短呈现时间的长于 550 ms 的, 不能认为被试在此短呈现时间内建立了句意。因为短呈现时间条件下消耗更长的反应时, 正是被试对句子的加工未达到与 550 ms 条件下相同的可以顺利完成实验任务的程度, 需要在句末词消失后继续加工句子所致的。因此, 要推断被试是否建立了句意, 考虑相关呈现时间下是否与 550 ms 条件下一样出现了不匹配效应的同时, 还须考虑两种条件下的反应时是否存在显著差异。), 说明 N400 出现之前大脑已建立了句意。如果在句末词呈现时间等于 N400 的潜伏时间时才出现与 550 ms 时相似的不匹配效应, 说明读者开始建立句意的时间大体上与 N400 潜伏时间相对应, 不存在 N400 之前的句意建构。

由于句末词的呈现时间需依 N400 的时间窗设定, 因此必须先明确此时间窗。Kutas 等最早的研究发现, N400 大约出现在语义违背事件后 250~550 ms。其后基于视觉语言刺激的研究尽管有部分研究发现其出现在稍早的 200~250 ms (van Berkum, Hagoort, & Brown, 1999), 但大部分都报告 N400 出现在语义事件后 250~550 ms。另外, 本实验的关键词将以汉语双字词形式呈现, 因此还需要了解汉语双字词诱发的 N400 的时间窗。查阅文献可知, 基于视觉汉语双字词的研究基本上都报告 N400 出现在事件后 250~550 ms (李荣宝, 彭聃龄, 郭桃梅, 2003; 金花等, 2009)。因此, 在视觉呈现刺激的情况下, N400 的时间窗更可能在事件后 250~550 ms。据此, 实验将关键词的呈现时间设置为 200 ms、250 ms 和 550 ms, 分别对应短于、等于和长于 N400 的潜伏时间。

2.1 方法

实验方法与 Zwaan (2002) 等的基本相同。不同的是句子逐词呈现, 句子与随后的图片间无时间间隔。

2.1.1 被试 44名在校大学生,男17人,女27人,年龄在19~26岁之间,均视力正常或矫正后正常。所有被试均获得适量报酬。

2.1.2 材料 部分材料直接把 Zwaan 等的材料翻译为中文后使用,其余材料依与其相同的方法编写。

实验共采用216幅黑白图片,其中156幅图片作为实验图片,其余60幅图片为填充图片。实验图片两两配对,配对的图片描绘了同一事物的不同状态(如图1)。216个句子被用于实验,包括156个实验句和60个填充句,分别与实验图片和填充图片相对应。实验句子亦两两配对,每对句子描述了同一物体的不同状态。如与图1的两幅图片相对应的一对句子是“天空中有一只老鹰”和“巢穴中有一只老鹰”。其中前者与图1左右两幅图片分别组成句图匹配与不匹配两种条件,后者则相反。填充句均描述了与随后图片不同的物体,相关的图片判断任务中被试均需作“否”的回答。

正式实验前,按 Zwaan 等的方法对实验材料的有效性进行了评定。结果表明实验材料有效。

2.1.3 设计与程序 实验采用2(句图匹配关系:匹配与不匹配)×3(关键词呈现时间:200 ms、250 ms与550 ms)两因素被试内设计。在句图匹配关系因素上,为平衡实验项目和条件间的顺序误差设计了2(句子)×2(图片)共4个实验序列。在关键词呈现时间因素上,同一呈现时间的句子在同一组块呈现,每个被试共阅读三个组块。为平衡不同组块和每个组块的句子组接受不同关键词呈现时间的顺序误差,以不完全抵消平衡方法设计了3个实验序列。如此,实验1设计了4×3共12个实验序列,每个被试阅读其中一个实验序列。

采用计算机动窗技术呈现实验刺激。被试要阅读一系列句子。句子均由四词组成,并以词语为单位逐词在屏幕中央呈现,如“天空中/有/一只/老鹰/”。每个句子出现之前屏幕中央会呈现注视点2000 ms。句末词(关键词)前的所有词语均呈现800 ms,句末词在每一实验序列的不同组块的呈现时间分别为200 ms、250 ms或550 ms。句子消失后立即呈现图片,被试此时要又快又准地判断图片中事物是否在刚才的句子中提到过,判断标准与 Zwaan 等研究中的相同。判断为“是”按“J”键,判断为“否”则按“F”键。按键反应后图片消失。计算机记录被试图片判断的反应时和正确率。正式实验前被试有操作练习,实验中间有两次休息时间,实验约需时15分钟。

2.2 结果

用 SPSS 11.5 分析实验数据。删除判断正确率低于85%的被试,无被试被删除。分析反应时数据时,删除三个标准差以外的极端数据。被试在不同句图匹配条件、不同关键词呈现时间下图片判断的反应时的平均值及标准差见表1。

表1 实验1中不同句图匹配条件下图片判断的反应时间(单位:ms)

关键词呈现时间	匹配		不匹配	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
200	705.68	173.82	729.44	180.38
250	710.21	199.64	735.07	210.52
550	694.23	197.06	730.78	244.11

运用两因素重复测量方差分析对反应时数据进行基于被试的分析。分析表明:句图匹配条件主效应显著, $F(1,43)=18.86$, $p<0.001$; 关键词呈现时间主效应不显著, $F(2,86)=0.11$, $p>0.05$; 两因素交互作用不显著, $F(2,86)=0.27$, $p>0.05$ 。为更清晰地比较不同呈现时间下的反应时模式,进行了简单效应分析。分析表明:呈现时间为550 ms时,不匹配条件下的反应时显著长于匹配条件下的, $F(1,43)=7.22$, $p<0.01$; 呈现时间为250 ms时,两种条件下的反应时差异边缘显著, $F(1,43)=3.01$, $p=0.09$; 呈现时间为200 ms时,不匹配条件下的反应时显著长于匹配条件下的反应时, $F(1,43)=4.94$, $p<0.05$ 。

数据分析结果显示,关键词呈现时间为200 ms时出现了与550 ms时相似的不匹配效应,表明关键词呈现200 ms后读者已将其联系到了上文语义表征而建立了句意,提示可能存在早于N400出现的句意建构。

3 实验2

实验2关注第二个问题,即读者何时开始将句末词联系到句子上文而建立句意。

本实验沿用实验1的方法,只是将关键词的呈现时间分别缩短为150 ms、180 ms和200 ms。根据实验1,关键词呈现时间为200 ms时出现的不匹配效应与550 ms时的相似,为缩短实验时间,本实验将以200 ms条件下的不匹配效应数据模式作为参照,以与实验1相同的思路,通过分析150 ms和180 ms条件下是否出现了与200 ms条件下相似的不匹配效应,确定被试在这两种条件下是否建构了句意,进而揭示被试开始建立句意的具体时间。

3.1 方法

除关键词呈现时间外, 实验方法与实验 1 相同。46 名与实验 1 同质的被试参加了本实验, 男 20 人, 女 26 人, 均获得适量报酬。

3.2 结果

被试在不同句图匹配条件、不同关键词呈现时间下图片判断的反应时的平均值及标准差见表 2。

表 2 实验 2 中不同句图匹配条件下图片判断的反应时间(单位: ms)

关键词呈现时间	匹配		不匹配	
	M	SD	M	SD
150	713.47	197.08	753.90	245.77
180	718.60	203.99	769.98	280.34
200	710.53	220.52	763.39	296.95

对反应时数据分析表明: 匹配条件主效应显著, $F(1,45)=14.85$, $p<0.001$; 关键词呈现时间主效应不显著, $F(2,90)=0.19$, $p>0.05$; 两因素交互作用不显著, $F(2,90)=0.17$, $p>0.05$ 。简单效应分析表明: 呈现时间分别为 200 ms、180 ms 和 150 ms 时不匹配条件下的反应时间均显著长于匹配条件下的, 检验值分别为 $F(1,45)=8.64$, $p<0.01$; $F(1,45)=7.89$, $p<0.01$; $F(1,45)=4.59$, $p<0.05$ 。因此实验 2 表明, 读者在句末词呈现了 150 ms 时或之前已建立了句意。

4 实验 3

本实验旨在进一步探讨读者开始建立句意的更精确时间。实验思路与实验 2 相同。实验中, 关键词呈现时间将被缩短至 80 ms、120 ms 和 200 ms。66 名与实验 1 同质的被试参加了本实验, 男 29 人, 女 37 人, 均获得适量报酬。结果见表 3。

表 3 实验 3 中不同句图匹配条件下图片判断的反应时间(单位: ms)

关键词呈现时间	匹配		不匹配	
	M	SD	M	SD
80	771.36	185.14	772.33	199.32
120	736.56	174.13	771.32	178.20
200	707.17	170.53	738.06	192.35

对反应时数据分析表明: 匹配条件主效应显著, $F(1,65)=10.79$, $p<0.01$; 关键词呈现时间主效应显著, $F(2,130)=4.87$, $p<0.01$; 两因素交互作用边缘显著, $F(2,130)=2.94$, $p=0.057$ 。对三种呈现时间进行两

两配对比较表明: 呈现时间为 200 ms 时的反应时显著短于 120 及 80 ms 时的($p<0.05$), 后两者的反应时无显著差异($p>0.05$)。简单效应分析表明: 呈现时间为 200 ms 和 120 ms 时不匹配条件下的反应时间均显著长于匹配条件下的, 检验值分别为 $F(1,65)=7.42$, $p<0.01$, $F(1,65)=10.26$, $p<0.01$; 呈现时间为 80 ms 两种条件的反应时间无显著差异, $F(1,65)=0.01$, $p>0.05$ 。这些结果表明, 当关键词呈现时间短于及等于 120 ms 时, 图片判断反应时的模式与呈现时间为 200 ms 时的有显著差异, 即 80 ms 条件下无匹配效应, 120 ms 条件下有不匹配效应但反应时显著长于 200 ms 条件下的。因此, 根据前述的推断被试是否建立了句意的标准, 实验 3 提示被试在句末词呈现 120 ms 或 80 ms 后均未建立起句意。

综合上述的三个实验, 本研究提示在句子阅读理解过程中, 句意建构开始于句末词呈现后 120~150 ms 之间。进一步, 分析实验 3 的结果可知, 句末词呈现时间为 120 ms 时的不同句图匹配条件下的反应时均比呈现时间为 200 ms 时的约长 30 ms。结合实验 2, 可以认为这些反应时也均比呈现时间为 150 ms 时的约长 30 ms。由此, 可以推测出被试开始建立起句意的更精确时间约就在句末词呈现了 150 ms 之时。因此, 总体上, 实验 1~3 提示在句子阅读理解过程中, 读者在句末词呈现了约 150 ms 之时就开始建立起了句意。

由于实验 1、2 与 3 是实验间设计, 为排除实验间被试差异对实验效应产生影响的可能性, 考察了三个实验关键词呈现时间为 200 ms 条件下不同句图匹配关系的平均值。两因素方差分析(其中句图匹配关系为被试内因素, 实验为被试间因素)表明, 句图匹配关系主效显著, $F(1,153)=20.63$, $p<0.001$; 实验主效应不显著, $F(2,153)=0.12$, $p>0.05$; 两因素交互作用不显著, $F(2,153)=1.14$, $p>0.05$ 。进一步的分析表明, 不同实验句图匹配及不匹配条件下的反应时差异均不显著(所有 $p>0.05$)。这一结果说明, 三个实验的关键词呈现时间为 200 ms 条件下的反应时间模式无显著差异, 支持了实验的结果。

5 实验 4

本实验旨在排除一些对实验 1~3 效应的其他可能解释, 以提高实验结果的可靠性。

5.1 实验 4a

对实验 1~3 的结果, 一种可能的质疑是认为不

匹配效应来源于词汇启动效应而非建构了句意。诚然, 词汇启动效应可能来自于两个方面。以句子“天空中有一只老鹰”为例。一方面是关键词“老鹰”可能被首词“天空”启动, 使得被试主要表征其为“飞鹰”, 从而出现了不匹配效应; 另一方面是图片被与之匹配的首词启动, 如“天空”可能启动了“飞鹰”的图片, 使得其比“站鹰”判断得更快。真实的情况是否如此呢? 实验 4a 将对此进行探析。

实验 4a 将让被试完成与实验 1~3 相似的任务, 不同的是将句子变为以其首词与关键词组成的词对, 如“天空-老鹰”。句子变为词对后, 词对与图片的匹配关系与其对应的句子与图片的匹配关系相同。如此, 如果在图片判断时不匹配效应仍然出现, 说明句子首词对关键词或/和图片起了启动作用, 词汇启动可能是上述实验效应的主要来源; 如果不匹配效应不出现, 则说明实验 1~3 的结果是有意义的。

5.1.1 方法

(1) 被试

24 名与实验 1 同质的被试参加了本实验, 男 11 人, 女 13 人, 均获得适量报酬。

(2) 材料

从实验 1 的材料中随机抽取 40 对实验句子, 提取每一个句子的首词与句末词分别组成 40 对实验词对和 40 个填充词对。120 幅与抽取的句子相对应的图片同时被选入本实验。

(3) 设计与程序

单因素被试内设计。自变量为词对与图片的匹配关系。设计了 4 个实验序列以控制项目与条件间的顺序效应。每一序列包含每一物体的四种可能版本(2 词对 $\times$ 2 图片)之一。每一被试阅读其中一个序列。

程序与实验 1~3 相同。为与实验 1~3 相一致且保证被试完成词对语义加工, 所有词对的首词将呈现 800 ms, 关键词呈现 550 ms。

5.1.2 结果 以实验 1~3 相同方法处理数据。

被试在不同词对-图匹配条件下图片判断的反应时的平均值及标准差见表 4。

表 4 实验 4a 中图片判断的反应时间(单位: ms)

词对-图匹配条件	<i>M</i>	<i>SD</i>
匹配	722.76	153.42
不匹配	730.70	166.63

对反应时数据进行单因素重复测量方差分析。分析表明, 不同匹配条件的反应时差异不显著,

$F(1,23)=0.35, p>0.05$ 。这一结果表明, 词汇启动效应并非实验 1~3 结果的主要原因, 实验 1~3 的结果是有意义的。

5.2 实验 4b

对实验 1~3 的结果另一种质疑可能是认为在短句末词呈现时间, 如 200 ms 和 150 ms 的条件下, 被试可能并未建立起句意, 加工仅处于句末词的词汇通达或选择阶段, 之所以会出现不匹配效应是因为句末词与语境相一致的知觉符号的通达或选择得到了促进。事实是否如此? 实验 4b 将对此作初步的探析。

实验 4b 中, 实验句子与实验 1~3 不同的是将状语放于句后, 如“天空中有一只老鹰”将变为“一只老鹰在天空”。逐词呈现句子, 如“一只/老鹰/在/天空/”。关键词为句末词, 如“天空”, 呈现时间设定为 200 ms 或 150 ms。如例句, 如果呈现了 200 ms 或 150 ms, 关键词的加工仅处于词汇通达或选择阶段, 被试未能将其联系到上文语境而建立句意, 那么语境中的“老鹰”不同的知觉符号将可能处于相同的激活水平, 因而将不会出现不匹配效应。如果出现此情况, 说明实验 1~3 的效应并非来源于句意建构; 相反则在一定程度上支持了实验 1~3 的结果。

5.2.1 方法 23 名与实验 1 同质的被试参加了本实验, 男 14 人, 女 9 人, 均获得适量报酬。

实验 4b 与实验 1~3 相似。不同的是句子按上述方式作了改变, 把状语放在句后。句末词呈现时间设置 150 ms 和 200 ms 两种条件。

采用 2 (句图匹配关系: 匹配与不匹配) $\times$ 2 (关键词的呈现时间: 150 ms 与 200 ms)两因素被试内设计。在句图匹配关系因素上, 为平衡实验项目和条件间的顺序误差设计了 4 个实验序列。在关键词呈现时间因素上, 同一呈现时间的句子在同一组块呈现, 每个被试共阅读两个组块。4 个实验序列中的项目均随机分成两个组块。为平衡不同组块接受不同关键词呈现时间的顺序误差, 设计了 2 个实验序列。如此, 实验 4b 设计了 4 $\times$ 2 共 8 个实验序列, 每个被试阅读其中一个实验序列。

5.2.2 结果 被试在不同句图匹配条件、不同关键词呈现时间下图片判断的反应时的平均值及标准差见表 5。

数据分析表明: 匹配条件主效应显著, $F(1,22)=7.74, p<0.05$; 关键词呈现时间主效应不显著, $F(1,22)=0.03, p>0.05$; 两因素交互作用不显著, $F(1,22)=0.05, p>0.05$ 。简单效应分析表明: 呈现时

间为 150 ms 和 200 ms 时不匹配条件下的反应时均显著长于匹配条件下的, 检验值分别为 $F(1,22)=4.81, p<0.05, F(1,22)=5.45, p<0.05$ 。这些结果表明, 当关键词呈现时间为 150 ms 和 200 ms 时均出现了不匹配效应, 且两者无显著差异。因此, 实验 4b 可以在一定程度上排除了实验 1~3 的实验效应来源于语境对关键词通达或选择的影响的可能性。

表 5 实验 4b 中不同句图匹配条件下图片判断的反应时间(单位: ms)

关键词呈现时间	匹配		不匹配	
	M	SD	M	SD
150	718.04	135.62	769.19	199.44
200	716.76	123.24	762.29	152.35

6 讨论

句子意义建构的实质, 尤其是其时间特征, 一直是心理语言学研究关注的热点问题。当前学术界一般认为, 在句子理解过程中, 句意建构是基于语义整合的, 主要反映在 N400 之上, 约开始于词语呈现后 250 ms。然而, 近年关于语义 P600 效应的一些研究却提示在语义整合之外, 大脑还可能通过其他加工通道或方式建立句子的语义解释, 而且意义建构可能开始于 N400 反映的加工启动之前。究竟句意建构开始于何时, 句意建构是否会早于 N400 反映的加工出现呢? 本研究设计了四个实验对此进行了初步探讨。实验结果表明, 句子阅读理解过程中读者在句末双字词呈现了约 150 ms 之时, 就将其联系到上文语义表征而建立了句意, 支持了语义 P600 效应相关研究的预期或启示。

除支持了存在早期句意建构的预期或启示之外, 本研究也在一定程度上支持了语义 P600 效应相关的句意建构存在多通道或多方式的假设。如 Kolk 等认为, 大脑可以通过概念倾向与句法分析两条通道建构句子语义解释, 当两种解释一致时, 其会整合正在加工的事件, 而当两种解释相冲突时, 就会启动监控机制检查最初的分析而放弃了整合(Kolk et al., 2003)。本研究发现在句末词呈现了约 150 ms 就出现了句意建构, 支持了在整合之前存在通过概念倾向或/和句法分析建立对句子的最初语义解释的可能性。Kuperberg 则认为, 句意建构有基于语义记忆和联合通道两条通道, 当两条通道形成的各种句子解释发生冲突就会引起联合通道的持续分析, 从而关闭了或者掩盖了 N400 所反映的

整合加工。本研究的结果也支持了这一假设。进一步, Kuperberg 认为基于语义记忆通道的加工发现句子语义违背时会引发 N400 效应, 这一通道的加工至少应部分反映在 N400 之上。他提出, 基于语义记忆通道的运作方式有二种可能。一是在理解过程中大脑会持续地比较不断进入的实词之间的语义关系与贮存于语义记忆之中的相应关系, 以检查两者是否一致; 二是大脑先通过语用或推论性的启发式机制(heuristic mechanisms)把动词、论元(argument)和其他实词捆绑成一个尝试性的命题, 然后再对照存贮于语义记忆系统中的真实世界知识来评估这一命题的合理性(Kuperberg, 2007)。本研究的结果在一定程度上支持了第二种运作方式。如果按第一种方式运作, 各种加工是平行进行的, 那么大脑应在句意加工的一开始就能发现语义违背, 从而引起 N400 效应, 而本研究发现大脑建立句意的时间早于 N400 的始点, 因此第二种运作方式更为可能。

句意建构存在多通道或多方式的假设看似是合理的, 但如上述的不同理论阐述孰是孰非呢? 显然, 本研究并不能给出明确的答案。更准确地讲, 本研究仅是在一定程度上支持了句意建构存在多通道或多方式的假设, 并未提供直接的证据, 因此句意建构是否存在多通道或多方式仍是一个待解决的问题。事实上, 本研究探测到的早期句意仍可能是基于语义整合而非所谓的其他加工通道而建立的。当然, 这一可能性要成立, 必须回答一个问题: 为什么基于 N400 的研究未能发现早期的语义整合? 一种可能的解释是基于 N400 的研究多是以正常句与语义信息违背句的 N400 开始出现差异始点, 即 N400 效应的始点来反映语义整合开始的时间, 如前述 Hagoort 等(2004)的研究, 但最早期的语义整合对于正常句和语义相关违背句可能是相同的, 如此分析 N400 效应自然也就不能发现这一早期的语义整合。事实究竟是如何的呢? 要准确回答需待进一步的探讨。

总体来说, 本研究发现在句子阅读理解过程中, 被试在句末词呈现了约 150 ms 之时就将其联系到句子上文语义表征而建立了句意, 提示句意建构可能早于 N400 反映的加工出现, 句子意义建构可能存在不同的通道或方式, 支持了语义 P600 效应相关研究的预期或启示。但是, 句子意义加工的过程究竟是如何的呢? 是否如语义 P600 效应相关研究提出的那样呢? 要解答这些问题, 需待未来更多的

研究。

另一方面,依上述讨论,本研究的发现对进一步揭示句子意义建构的实质似乎有一定的意义,但其毕竟是与关于句子意义加工机制的主流看法相冲突的,因此其是否合理、可靠是一个值得关注的问题。诚然,本研究可能会受到很多质疑,除实验处理以外,最可能的质疑有以下两个。第一,本研究发现被试在句末词呈现了约 150 ms 就建立了句意,而一般认为 N400 出现在语义刺激呈现后 250 ms,由此就认为句意建构的开始时间要早于 N400 效应,这样的推论是否合适呢?如果本研究的实验处理加快了句意建构,那么探测到的早期句意建构就可能仍是在 N400 上反映的。第二,本研究在一定程度上是以反应时数据对许多 ERP 实验结果提出了挑战,如此以低精确度的研究挑战高精度的研究是否合适?客观地讲,本研究并不能很好地回答这两个质疑。对于第一个质疑,尽管大量关于句子加工的研究都较一致地报告 N400 开始出现于语义刺激呈现后约 250 ms 时,其潜伏时间较少受到实验处理的影响,而且在一些实验任务与本研究的相似的研究中也未发现(Hoeks, Stowe, & Doedens, 2004; Chwilla & Kolk, 2003),但毕竟本研究未同时收集 N400 相关数据,因此不能直接证明探测到的早期语义建构就一定是早于 N400 出现的。第二个质疑,本研究同样难以很好地回答,要让人信服本研究的结果的确需要来自 ERP 实验的证据。可喜的是,我们的一个后续研究已利用 ERP 技术进一步发现了早期的句意加工,相信这一项研究将会为本研究提供有力的证据支持(相关的研究报告正在审稿中)。

尽管本研究存在着一些不足,也可能会受到更多的质疑,但相信其同时也能为今后的研究带来一些有意义的启示。首先,语义 P600 效应发现以后,许多学者重新讨论了 P600 的功能性意义,也对句子意义加工的实质进行了许多重新的审视,但鲜有研究直接去验证了语义 P600 效应相关研究提出的一些推测或假设,也未就其对关于句子意义加工的主流观点的挑战作出实质性的辨析。而要进一步揭示句子意义建构的实质,直接验证这些推测或假设是很有必要的,本研究的这次初步的尝试相信可以为往后的相关研究提供一些思路上的启示。其次,不论本研究的合理性之处或不足有几何,其都进一步给了我们一个重要的启示,就是句子意义建构的实质可能并非人们先前所了解的那样的,可能尚有

许多未被发现的特性,因此还需要拓宽思路,进一步对其进行探讨与揭示。基于这两方面可能带来的启示,本研究应该是具有一定的意义的,期待其能为今后的研究带来一些积极的影响。

7 总结

本研究探讨了句子阅读理解过程中是否存在比 N400 反映的加工出现得更早的句意建构,以及如果存在这样的句意建构,其开始出现的时间,得到以下结论:(1)读者在句末双字词呈现了约 150 ms 之时就将其联系到上文语义表征而建立了句意;(2)可能存在比 N400 反映的加工更早的句意建构,读者可能在 N400 出现之前就已建立了初步句意。

参 考 文 献

- Barsalou, L. W. (1999a). Language comprehension: Archival memory or preparation for situated action? *Discourse Processes*, 28, 61–80.
- Barsalou, L. W. (1999b). Perceptual symbol systems. *Behavioral and Brain Science*, 22, 577–660.
- Chwilla, D. J., & Kolk, H. H. J. (2003). Event-related potential and reaction time evidence for inhibition between alternative meanings of ambiguous words. *Brain and Language*, 86, 167–192.
- Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304(5669), 438–441.
- Hinojosa, J. A., Mart ín-Loeches, M., & Rubia, F. J. (2001). Event-related potentials and semantics: An overview and an integrative proposal. *Brain and Language*, 78, 128–139.
- Hoeks, J. C. J., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, 19, 59–73.
- Jin, H., Zhong, W. F., Xu, G. P., Cai, M. X., Yang, Y. F., & Mo, L. (2009). The time course of world knowledge integration in sentence comprehension. *Acta Psychologica Sinica*, 41(7), 565–571.
- [金花, 钟伟芳, 徐贵平, 蔡梦娴, 杨玉芳, 莫雷. (2009). 世界知识在句子理解中的整合时程. *心理学报*, 41(7), 565–571.]
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, MA: Cambridge University Press.
- Kintsch, W., & Mross, E. F. (1985). Context effects in word identification. *Journal of Memory and Language*, 24, 336–349.
- Kolk, H., & Chwilla, D. (2007). Late positivities in unusual situations. *Brain and Language*, 100, 257–262.
- Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, 85, 1–36.
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146, 23–49.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless

- sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203–205.
- Li, R. B., Peng, D. L., & Guo, T. M. (2003). A study on Chinese and English semantic access with ERP technology. *Acta Psychologica Sinica*, 35(3), 309–316.
- [李荣宝, 彭聘龄, 郭桃梅. (2003). 汉英语义通达过程的事件相关电位研究. *心理学报*, 35(3), 309–316.]
- Stanfield, L. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, 12, 153–156.
- Till, R. E., Moss, E. F., & Kintsch, W. (1988). Time course of priming for associate and inference words in a discourse context. *Memory and Cognition*, 16(4), 283–298.
- van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, 11(6), 657–671.
- van Herten, M., Kolk, H. H. J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22, 241–255.
- Wang, R. L., Li, N., & Chen, B. G. (2010). The semantic P600 effect during sentence processing. *Advances in Psychological Science*, 18(4), 545–552.
- [王瑞乐, 李妮, 陈宝国. (2010). 句子加工中的语义 P600 效应. *心理科学进展*, 18(4), 545–552.]
- Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *American Psychological Society*, 13, 168–171.

The Timing of Sentence Meaning Establishment During Sentence Comprehension

ZHONG Wei-Fang; MO Lei; JIN Hua; XU Gui-Ping

(Center for Studies of Psychological Application, South China Normal University, Guangzhou, 510631, China)

Abstract

Many studies suggested the establishment of sentence meaning in sentence comprehension was based on lexical integration, which reflected on N400 starting around 250ms after the onset of an incoming word. However, research related to the semantic P600 effect proposed that there might be other route (s) for meaning establishment besides lexical integration, which might start earlier than the onset of the processing reflected on N400. Here four experiments were conducted to investigate whether there are other route (s) earlier than the processing indexed by N400 and when they begin.

In Experiment 1, a paradigm related to Perceptual Symbol Theory (PST) was used for study. Participants were asked to read Chinese sentences like 天空中有一只老鹰 (In the sky there is an eagle) word by word, and then to judge whether the pictured objects that followed without interval had been mentioned in the preceding sentences ignoring their shape and state. The reaction times for judgment were recorded. The last word of each sentence which was a two-character word was the critical word. All critical words were presented in three different time durations: 550ms, 250ms and 200ms. According to previous studies and PST, when the critical word was presented at 550ms, participants could have established the sentence meaning and a mismatch effect would appear. The mismatch effect is an effect that the reaction time is shorter when the sentence matches the following picture. The result found that mismatch effect also appeared in both 250ms and 200ms durations, and their reaction times were similar to those of 550ms duration. The result suggested that participants could establish sentence meaning even when the critical word was presented at 200ms, which indicated that sentence meaning establishment started before the onset of N400.

Experiments 2 and 3 aimed at revealing a more precise time point at which readers began to relate the last word of a sentence to the preceding linguistic input and establish sentence meaning. The experimental methods were the same to that of Experiment 1 except for shortening the duration of the critical word. The results suggested that participants began to establish sentence meaning after the critical word was presented at 150ms.

Experiment 4 was conducted to counter some possible explanations for the experimental effects of Experiment 1~3. Doubters might think that the mismatch effects in Experiments 1~3 came from prime effect but

not sentence meaning establishment. That is, the first word might prime the processing of the critical word or/and the following picture in the match condition. Experiment 4a was conducted to deal with this challenge. Participants were presented with Chinese word pairs like 天空—老鹰 (sky-eagle), and then asked to complete a judgment task as in Experiments 1~3. All word pairs were constructed based on Experiments 1~3 by combining the first word and critical word of each sentence. The results showed that the mismatch effect did not appear in Experiment 2, which supported the findings of Experiments 1~3. One might also doubt whether readers could establish sentence meaning in the short critical word present time conditions, like 150ms and 200ms, and think that the reason for the appearance of the mismatch effects was that the sentential context facilitated the access or selection of the critical word's perceptual symbol which was consistent with it. In Experiment 4b, participants were asked to complete experimental tasks similar to Experiments 1~3. Unlike Experiments 1~3, sentences were reconstructed, the critical words were moved away from the sentence end, and the last words were present at 150ms or 200ms. The results showed that in both the duration conditions, mismatch effect was present, which also supported Experiments 1~3.

In conclusion, the results suggested that during sentence comprehension, readers could relate an incoming two-character Chinese sentence-end word to the preceding linguistic input about 150ms after its presentation and establish the sentence meaning, and sentence meaning establishment started before the onset of the processing reflected on N400.

Key words sentence comprehension; sentence meaning establishment; temporal feature; N400