

· 综述和编译 ·

语言理解的容量理论及其研究方法

王亚同

中国科学院心理研究所

关键词：语言理解，工作记忆，个别差异，容量理论，激活。

〔提要〕 本文介绍了美国心理学家Just和Carpenter最新提出的语言理解容量理论及其研究方法。容量理论认为工作记忆限制语言理解，激活是加工和贮存的媒介，工作记忆的激活量因人而异。容量理论还可以从量和质两方面说明语言理解的个别差异，这一点体现在两个方面，其一是句法调节性，其二是句法模糊性。本文还介绍了有关的研究方法。

在各种复杂思维活动如推理，问题解决和语言理解中工作记忆起着重要作用，这种作用在需要加工一系列符号的语言理解中比较明显。人们阅读课文或听课时由连续的字词形成思想，这一过程中工作记忆起着贮存作用即有利于贮存计算的中介产品和最终产品，工作记忆还可以视为操作性资源库用以完成符号计算，生成中介产品和最终产品。过去，工作记忆主要涉及短暂间隔后再检索的信息贮存，为信息进入长时记忆奠定了基础，记忆信息通过反复练习或精细加工从工作记忆进入长时记忆，因此不论短时贮存还是长时贮存都涉及工作记忆。当前的工作记忆理论认为贮存项目不仅为了随后的检索，而且在复杂的序列计算如语言理解中可以贮存部分产品。理解词汇时贮存的作用比较明显，理解者能够很快检索呈现在句子前后关系中的字词和短语，在语言理解加工的其他水平，贮存的作用也会体现出来，理解者贮存课文的题目，贮存课文情境的表征，贮存句子的主要命题，贮存连续的多水平句子表征。因此，语言理解需要贮存加工的部分产品和最终产品用于更为复杂的信息加工。贮存就是实际的计算，计算就是人类思维核心的符号操作如比较操作和检索操作，逻辑操作和数字操作，语言理解的过程与这些符号操作有关，并且与贮存的资源相结合构成了语言理解的工作记忆。Baddeley(1974)认为工作记忆属于计算范畴，他用实验研究了理解的贮存与加工及其相互关系，结果发现，当被试编码几个数字再回忆时迅速理解个别句子的准确性下降了，这说明贮存和加工采用了一个共同的资源库，工作记忆在一个系统中具有双重作用。本文涉及的工作记忆是指语言理解的一些过程及需要的资源。

1 容量理论概述

本文的目的是从理论上介绍工作记忆在语言理解中的贮存加工作用，涉及的计算理论是指通过激活影响贮存和加工，所谓容量就是指工作记忆的最大激活量，可以说明贮存或激活

的程度。容量理论认为,每一个表征成分对应于一种相关的激活水平,这个成分可以是一个字词一个短语一个命题,也可以是文法结构或课文结构等。人在理解书面课文或口头话语时编码和激活信息,并且通过计算进行加工生成,由长时记忆进行检索。成分的激活达到最小阈值,该成分就是工作记忆的一部分,参与理解操作的各种过程。系统的总有效激活量小于应完成理解的作业量时有的激活就会重新分配已经贮存的一些成分从而产生遗忘,因此,初始形成的句子表征往往在后来需要时却遗忘了。激活用于信息保持也用于产生式系统结构中进行的计算,通过计算改变激活水平,通过产生式操作大量符号。产生式系统有许多动作成分,其中产生式增加一个动作成分的激活水平时就会出现共同的操作过程,同时适当改变激活水平也可增加或消除一些成分。

在语言理解的联结模型中计算就是激活操作,因此产生式规则将激活从一个成分扩散到另一个成分,并且将源成分作为条件,将目标成分作为动作,反复地增加或改变目标成分的激活水平直至达到一定的阈限,例如,看见句子的主语便希望谓语随其后而出现,因为文法主语命题是激活源。符号系统的特点是根据规则进行加工,联结系统的特点是分层次激活,两者相互配合发生作用。理解的过程是并行出现的,因此理解者希望动词出现时还可以计算句子的其他句法特征,语义特征和语用特征。容量理论认为能够进行的所有过程都可以同时进行,而且可以同时生成部分产品。如果过程或产生式的数目较大或者说扩散的激活量超过了已有的容量,那么扩散就在最大范围徘徊以保持激活的总体水平不变。贮存和加工的交换关系受分配图式的控制,就是说,产生式数目超过了激活最大值,分配图式便起作用即扩散的激活和用于保持的激活都按比例缩小,使增加的成分接近阈限要求的数目从而有效地减缓计算。这样,在作业要求较高时,分配图式使加工速度放慢从而导致遗忘一部分信息。总之在这个系统中,语言加工的内容和时间依赖于贮存的容量和计算,作业要求(即贮存要求或计算要求)超出有效资源时贮存与计算的作用就会降低。这种语言理解称之为“容量有限理解”。

课文是由许多句子组成的,贮存这些句子耗费有限的容量,因此系统具有减少贮存要求的抵消机制,并以激活的形式有选择地保持最近最主要的句子表征和最相关的世界性知识,同时降低初始阅读课文时激活的其他知识。当新词或新短语首次出现时,贮存要求进行直接加工,不说明该词或短语的语义。合适的高层次语言结构形成之后就可以激活低层次的语言表征,例如,阅读课文时可以获得少量已经看过的从句词汇信息或句法信息。有限工作记忆容量限制加工句子的贮存要求,而且贮存的语境也有助于加工句子,先激活与理解有关的概念,关系和图式,然后加工相邻的下一个句子,因此加工扩展的课文时,工作记忆可以促进加工并使贮存要求达到最低。语言理解依赖于工作记忆容量。在满足语言加工计算要求和贮存要求上,个体获得的激活量是不同的,个体理解语言的速度和精确性有量的差异,不同的个体之间又有质的差异。研究表明,理解作业容易,被试者的工作记忆容量操作差异比较小,这一点往往可以忽略,如果理解作业难度较大,则操作差异大而且系统化地表现出来,这说明作业资源供不应求,有限容量就会影响操作。

2 容量理论的研究方法

Daneman和Carpenter(1980)的“阅读广度作业”可用于测量语言工作记忆容量的个

别差异,该作业同时利用了工作记忆的加工贮存资源。作业要求被试者读一组无关的句子如“他睁开眼睛时,没有了胜利的希望,也没有一点怨恨”,“出租汽车在密式根大街调头转向,在这个大街他们清楚地看见了湖泊”。读完这两句之后,被试者回忆每句的最后一个词或尾词,如该例中的“怨恨”和“湖泊”。这样,通过被试者回忆每组句子的所有尾词可以发现每组最多能包含几个句子,如果被试者在5组句子中最少可以成功地回忆其中3组句子的所有尾词,这个最大值称为被试者的“阅读广度”,如果被试者能够回忆其中2组句子的所有尾词,这种情形属于中等广度。对于这种类型的句子,大学生的阅读广度在2—5.5词之间,对于大多数学生来说,“高广度”为4个以上的词,“中等广度”为3—3.5个词,“低广度”为不足3个词。低广度的标志与取样有关,在容量理论的研究中,低广度被试者的水平位于标准化测验(如学术能力测试)中语言能力分布的上半部分。高广度读者阅读句子采用的理解过程耗费较少的工作记忆资源,他们有更多的容量保持许多句子的尾词,测量这一个别差异的方法是根据Baddeley和Hitch(1974,1976)的研究形成的,结果表明,语言理解和同时进行的数字回忆可以利用一个共同的资源,因此同时完成两个作业时其操作过程具有交换关系,说明容量是有限的。

当总的加工资源不足以进行不同过程的直接交互作用时,容量限制能够在各种不同过程之间产生许多界线,这些过程的交互作用如各种形式的计算都需要一些操作资源,如贮存部分产品及不同过程的交流,如果缺乏足以支持交互作用的资源,具有必然联系的两个过程就不能交互作用,就是说这两个过程不会影响彼此之间的连续计算,容量限制产生的界线是资源限制的结果,不是由于语言结构障碍引起的。因此,容量限制可以从新的角度解释语言加工的组块,认知组块就是进行速度快范围具体的一些过程,具有限制和信息压缩的性质,Fodor(1983)认为,认知组块完成了句法水平的语言加工。容量理论认为句法加工的组块就是偶尔进行信息压缩的容量限制。信息压缩是认知组块最明显的一种特征,是指在系统任何位置存在的不受各种信息影响的组块的活动与输出。容量理论研究句法加工的个别差异。Ferreira和Clifton(1986)设计的作业通过直接使用非句法信息让读者避免进入果园路(garden path),但读者还是进入了果园路,尽管已经呈现了歧义句正确答案的明确信息,这些明确信息是没有句法的信息,并且对句法加工缺乏影响,这是由于句法组块的压缩作用所致。Just和Carpenter(1992)重复了这个实验,他们将具有不同阅读广度的被试者分开测试,结果发现,低广度被试者的成绩与Ferreira的结果一致,但是高广度被试者在开始解释歧义句时就注意到了非句法信息,他们的句法加工不是组块式的而是交互作用的进行。Ferreira和Clifton(1986)检验了句子的阅读时间,如句子1“The evidence examined by the lawyer shocked the jury(律师发现的证据震惊了陪审员)”,句子2“The defendant examined by the lawyer shocked the jury(律师审查的被告震惊了陪审员)”。这两个句子都省略了关系从句的关系代词和动词(“who was”或“that was”),这种句子叫做“省略的关系从句”。句子2的开始部分“the defendant examined(被告审查)”暂时有两种可能的解释,²一是将“examined”解释为主要动词如“被告审视了法庭”,一是解释为最终正确的一个从句。这个研究的特点就是语用线索即主要名词的生物性是否体现了句子开头模糊部分的正确解释。句子1开头的名词具有非生物性,不可能成为相临动词的代理者,如果非生物性的语用信息影响句法分析的决策,读者很可能将动词解释为省略的关系动词,不会认为是主要动词,而且被试者希望很快具体说明句中的代理者,当代理者短语如“by the lawyer”出

现时不会感到莫名其妙,不会影响具体的加工。相反,句子2的“the examine”很可能是动词“examined”的代理者,读者可能将动词解释为主要动词,开始进入果园路。随后的代理者短语不符合解释为主要动词,遇见“by”短语就会感到加工困难。通过比较句子1和句子2中“by the lawyer”的阅读时间可以测量果园路效应以发现名词“evidence”的非生物性是否减少了遇见“by”短语的惊奇程度。结果发现,句子的主要名词不具有生物性时,读者仍然在“by”短语上用了较长的时间,这说明读者进入了果园路,主要名词如“evidence”的非生物性没有进入动词“examined”的句法分析。Just和Carpenter(1992)的实验几乎完全类乎于Ferreira和Clifton(1986)的实验,唯一的区别是将不同阅读广度被试者分成两组,高广度(4个以上的词)的读者40名,低广度(2.5以下的词)的读者40名,并对原实验材料作了改进,去掉了文法主语解释为工具的一些句子如“The car towed...”,每组20个句子,有间隔地插入各种填充句,随机呈现。实验材料有省略的关系从句,也有未省略的关系从句如句法明确的句子:句子3“The evidence that was examined by the lawyer shocked the jury”和句子4“the defendant who was examined by the lawyer shocked the jury”。这4种句子形式呈现给4组被试者,每一个句子代表一种形式。被试者阅读自动记录监视器上呈现的句子,实验者用ISCAN型RK-426 Pupil/Corneal跟踪系统记录被试者的眼睛注视,每隔16.7毫秒用VAX3200工作站计算注视点。每次屏幕先呈现具有介绍性的不带任何暗示的句子,然后呈现目标句,再呈现真假理解问题。有两个键,读者按其中一个键作出反应。

由工作记忆容量度量的容量限制在加工的时间过程和精确性上不但会产生量的个别差异,而且当句子或作业特别需要容量时,这些量的差异体现得最明显。最近几个研究表明,高广度读者理解难句更快更精确。最需要工作记忆容量的句法结构的经典例子就是位于中间的宾语从句如句5“the reporter that the senator attacked admitted the error”,句首的名词是关系从句的宾语。被试者先听这样一个句子然后试图进行重述,犯错误的次数为15%。将两种加工过程结合起来使这样的句子难以理解,简单说来有两难,一难是中间嵌入的从句干扰主句,二难是句法成分之一(如上例中的“reporter”)不仅是主句的主语而且是嵌入从句的宾语,将一个概念与两个不同角色同时联系起来增加了语言理解的难度,相反,理解主语从句显得更容易。King和Just(1991)让被试者阅读宾语从句或主语从句时测量了一字一字的阅读时间,然后回答问题估计自己解释的精确性。阅读时间的估计采用了自控速度的移动窗口模式,即一开始用代替字母的短线呈现句子,每当被试者压小型开关时当前移动点之前相邻字母代替相应的线段,然后线段代替已经呈现的词,这样就得到了课文每个单词的阅读时间。结果表明,阅读时间有较大的个别差异,这些差异主要是宾语从句引起的,而且,一字一字的阅读时间用于加工较难的宾语从句,它们需要关键的句法信息。

另外,歧义句也需要较多的资源,尤其是在缺乏选择确切解释的语境时。在句子模糊的部分理解者可以作出几种解释,这些就需要较多的资源。但是,现有的数据和相应的理论不同意这种加工歧义句的观点,认为理解者遇见歧义句会选择一种解释;或者随后提供了明确的信息就会选择一种解释。实验证明,这两种观点可以相互协调,因为保持歧义句的多重表征具有个别差异。理解者一开始遇见歧义句就会形成多重表征,其中每一个表征都具有一种激活水平,表征的出现次数及其句法复杂程度和实用可行性都与激活水平有关。容量理论认为,理解者的工作记忆容量影响多重句法表征所保持的时间。低广度读者的容量不足以同时

保持两种表征，而且很快剔除不喜欢的解释，相反，高广度读者能够同时保持两种解释一段时间。

在实验中经常通过采用外部记忆负载来获得工作记忆资源，这种外部负担通过工作记忆保持或竞争资源的复述记录过程耗费资源。随着外部负载的增加，一个或几个阶段的成绩如阅读速度或回忆负载项目 (load item) 下降。保持外部负载干扰句子理解，因为二者竞争相同的资源。实验时，给外部负载组被试者呈现包括主语从句和宾语从句的句子如 “the reporter that the senator attacked admitted the error”，阅读这个句子时要求被试者记住一两个无关字词，在这种条件下他们回答随后理解问题的能力比不上非外部负载组被试者。

语言理解的实质就是人能够将不同成分如从句或复合句的信息相互联系起来。因此，需要保持信息一段时间，保持插入的计算。工作记忆提供贮存居前成分信息的资源，同时提供加工随后成分的计算资源，两个成分相关的距离越大，犯错误的概率越大，整合的过程越长。这是所谓的“距离效应”。另外，工作记忆容量与年龄有关。上述理解的5个方面包括了读者工作记忆容量的质与量差异，有了内部记忆负载（如，保持课文的连续句子）或外部负载，理解成绩一般下降，低广度读者的成绩下降幅度较大，对于要求计算的句子，低广度被试者比高广度被试者的阅读速度要慢许多，但是保持歧义句的两个表征，后者不如前者。每个人有限的容量制约着理解的随意性，因此不会生成一切可能超前的推理，不会一丝不漏地解释每个歧义句，不会想到所有潜在相关的解释线索。

3 容量理论机器模拟的特点

容量理论机器模拟采用CAPS结构即产生式系统与联结系统结合的产品，象在联结系统中那样，激活由源（即条件）成分扩散到目标（即动作）成分，产生式完成这种扩散而且可以相互之间并行操作，可以在几个操作过程重复地扩散激活直至目标成分达到一定阈限，容量限制就是控制该系统具有的激活总量以便在加工过程中保持工作记忆的成分并将激活扩散到其它一些成分。CAPS结构的特点是：（1）与每个工作记忆成分联系的数为实数，称作激活水平，表示该成分的强度，只要一个成分的激活水平处于由产生式规则具体说明的一定阈限之上，该成分就满足产生式的一个条件。（2）大多数工作记忆成分是命题形式：（概念：关系概念）或（概念〔蕴涵的：是一个〕概念）。这些成分可以形成网络。（3）产生式的启动将一个工作记忆成分（叫做“来源”）乘以因素（叫做“份量”）得到的激活流对准于另一个工作记忆成分（叫做“目标”）。（4）所谓的一个加工过程就是将工作记忆的全部产生式与随后平行启动满足“if”条件的全部产生式进行匹配。（5）长时记忆的知识就是脱离了工作记忆的陈述性数据库。

4 容量理论的意义

容量理论主要与思惟的资源有关，象结构理论一样，容量理论假设的基本结构由工作记忆，程序性知识和陈述性知识组成。容量理论可以采用语言进行检验，也可以采用问题解决，决策和高级视觉加工检验。容量理论的意义在于（1）可以说明个体的作业成绩差异。

当作业难度大需要很多容量时个体工作记忆容量小就不能很快完成计算或贮存。(2)个别差异可以作为测量语言理解的一种指标。语言理解的作业要求耗尽被试者的全部资源,语言结构较为复杂时容量限制更加明显。(3)在具体作业操作过程中思维的强度是不一样的,因此,被试者完成不同的作业耗费不同的资源量。作业越难需要的努力越大。(4)容量理论的激活概念来自Kahneman (1973)的注意容量理论,利用了注意资源及其分配的特点,并且根据信息保持和计算的有效激活解释容量,这样能够更好地说明注意理论关注的各种问题。(5)在包括语言加工的各种认知范畴,只要复杂作业的成分加工不受结构限制,只要有合适的资源支持这些成分加工过程的交互作用,这些交互作用就会出现。语言研究经常假设一些基本成分,这样就可以将复杂作业的操作过程分成许多子过程从而有效地研究这些子过程的交互作用。

5 参考文献:

- [1] Just, M.A., & Carpenter, P.A., A capacity theory of comprehension: Individual differences in working memory, *Psychological Review*, 99, 122-149, 1992
- [2] Daneman, M., & Carpenter, P.A., Individual difference in working memory and reading, *Journal of Verbal Learning and Verbal Behavior*, 19, 450-466, 1980
- [3] Kahneman, D., *Attention and effort*, Englewood Cliffs, NJ: Prenticehall, 1973
- [4] King, J., & Just, M.A., individual difference in syntactic processing: The role of working memory, *Journal of Memory and Language*, 30, 580-602, 1991

(上接64页)

术界进行沟通交流无疑是一份重要的参考资料。

本书目次:

1. Empirical Inquiry and Decision Criteria
2. Nature and Limitations of Control
3. Data Collection and Measurement Procedures

4. Basic Issues in the Implementation of Research
5. Fundamentals
6. Beyond One-Way Analysis of Variance
7. Power and Focused Analyses
8. Additional Topics in Data Analysis
9. Appendixes