

第一章 汉语多词语块在线加工的 频次与语义效应^①

第一节 引言

长期以来,语言一直被视为由词汇和语法这两个不同模块构成的一个系统。单个的词是由语素、音素等更小的单位构成的。借助语法规则,单个的词可以进一步组合为短语、句子等更大的语言单位。当前的心理语言学研究对词汇和句子的加工都比较关注,相比之下,处于词汇和句子这两个水平之间的多词单位(multi-words)受到的关注则显然不足(Shaoul & Westbury, 2011)。词作为语言加工的基本单位具有心理现实性,这已经得到了比较充分的证实(李兴珊,刘萍萍,马国杰,2011),那么,心理词典中是否存在大于词而小于句子的多词单位呢?介于词与句子之间的多词单位是否也具有心理现实性呢?

多词单位一般被称为“多词短语”(multi-word phrase)、“多词序列”(multi-word sequence)或“组块”(chunk)。语言学界——尤其是在语言习得领域,人们对多词单位的关注历史较

^① 该章部分内容曾以《语块的心理现实性》(易维,鹿士义)和《Frequency, contingency and online processing of multiword sequences: An eye-tracking study》(Weiyi, Shliyi Lu, Guojie Ma)分别刊发于《心理科学进展》2013年第12期和《Second Language Research》2017年第4期。

长、称谓繁多。Wray (2002) 曾经罗列了 50 多种不同的术语，它们之间互有交叉，内涵也不尽相同。为了研究的方便，Wray (2002) 提出使用“formulaic language”作为一个统一的术语，并给出了具体的操作定义。本章采用的是“语块”这个术语，其内涵和 Wray (2002) 提出的“formulaic language”基本对应，包罗了多词单位的各个子类，如习语、惯用语、高频搭配等。总的来说，语块是指语言使用者在长期、大量的语言接触和语言经验基础之上形成的形式和意义的组合，其基本特征至少包括：第一，语块是一个大于词而小于句子的多词单位，一般由两个或两个以上的词组成；第二，语块具有一定的频次性，语块的分布受到频次等因素的影响 (Ellis, 2002; Wray, 2002)；第三，语块具有一定的可预测性，语块的成分词 (component words) 不是任意组合的，它们之间的内部联系较为紧密，语言使用者能够较为容易地根据语块的前词推断出后词 (Siyanova, 2010)；第四，一般认为，语块的在线生成不受语法规则的作用，语块可能是整体预制的 (Wray, 2002)。

语料库语言学研究 (Biber Johnsson, Leech, Conrad & Finegan, 1999; Erman & Warren, 2000) 发现，自然口语和书面语中都存在着大量重复出现的语块。尽管不同研究的统计数据有所不同，但语言使用中存在着大量语块是显而易见的。Sinclair (1991) 认为，人们使用语言的过程中遵循着“开放性原则” (the open-choice principle) 和“习语性原则” (the idiom principle)。所谓“开放性原则”，指的是人们选择词汇、利用语法规则来生成话语；所谓“习语性原则”，指的是人们依赖存储于长时记忆中的预制语块来完成特定的表达任务。在语言使用过程中，“习语性原则”起主导作用，人们可以根据情境和表达的

需要在“开放性原则”和“习语性原则”之间自由转换。越来越多的语言学研究表明，除了使用语法规则在线生成新的话语之外，人们更倾向于使用预制语块进行表达和交际。

语言学研究的另一个共识是，语块对第一语言发展和第二语言习得都有重要的作用。母语习得研究（Wray & Perkins, 2000; Arnon, 2009）表明，儿童早期语言发展过程中存在着大量各种类型的语块，它们能促进儿童语言能力和交际能力的发展。此外，语块在二语习得方面也有重要意义。Ellis 和 Sinclair（1996）认为，二语习得的一个重要部分就是语块的习得。Pawley 和 Syder（1983）指出，二语习得能否接近母语者水平的关键就在于学习者是否懂得从表达特定意义或功能的一系列可能形式中选择出母语者最常用、最准确的那些类型。事实上，在第二语言习得领域，人们普遍认为，即便是达到熟练水平的二语者，他们在如何选择语块进行表达时仍然存在相当的困难。语块同语言水平的发展、语言表达的流利度有着密切的关联。Nattinger 和 Decarrico（1992）认为，人们使用语言的流利性并不取决于学习者在脑中存储了多少的语法规则，而取决于他们的大脑中存储了多少预制语块。

心理语言学对习语（idiom）的研究起步较早，并对习语加工提出了一系列假说（Bobrow & Bell, 1973; Swinney & Culter, 1979），相比之下，非习语的多词单位的研究一定程度上是在词汇加工研究中引发的。研究表明，人们对频次、可预测性等语言输入中的分布信息（distributional information）比较敏感（Howes & Solomon, 1951; Forster & Chambers, 1973）。传统的“词汇—规则”模型（Pinker, 1991; Pinker & Ullman, 2002）认为，频次只是整体存储的语言形式（词）的特征，短语是由词通过计

算的方式组合而成的，因此这种模型也就无法预测到短语的频次效应。与此观点相左的几个模型，如联结主义模型（Elman, 1991）、语言知识的范例模型（Pierrehumbert, 2001）、基于用法的理论（Bybee, 2006）等都认为，不同颗粒度的语言材料其实是以相似的方式表征和加工的，它们都要受到语言经验的作用和影响。因此，从这些理论模型的视角出发，频次效应似乎不仅存在于词汇层面，多词短语同样也可能具有频次效应。Arnon 和 Snider（2010）利用短语判断任务对频次分别为高、中、低三段的若干四词短语进行了研究，结果发现多词短语的加工确实也存在着频次效应，从而证实大脑中存储着短语的频次信息。这为心理词典中存在语块的表征提供了一定证据，也为人们重新认识心理词典的存储单位提供了新的视角。句子语境中的词语阅读研究也促使人们开始关注比词更大的语言单位。Prior 和 Bentin（2007）通过一项记忆研究表明，在句子语境中阅读词语会激发一种联想作用，这种作用能够将共现的词语联系在一起。这就是词汇加工的句子语境效应（Morris, 2006）。句子语境效应打破了孤立地研究单个的词的传统，使人们从词与词之间的关系入手，考察词汇加工乃至更大的语言单位的加工特征。

综上所述，作为一个大于词而小于句子的语言单位，语块具有特殊的重要性。虽然词在心理词典中的表征地位已经得到了比较充分的证实，但是这并不意味着只有词才是心理词典的存储单位。语块是否可能也是心理词典中的一个单位？即，语块是否和词一样，也是整体存储和提取加工的呢？确证了语块的心理现实性，也就表明人类的语言加工不止以词为单位，表明语块在语言加工中也有相当重要的地位。因此，语块是否整体存储于心理词典之中，是否具有独立的心理表征，这是非常值得研究的一个

问题。

第二节 语块的心理现实性

Ellis 和 Sinclair (1996) 认为, 工作记忆对语言习得和语言使用中信息的保持、巩固、存储和提取有重要的制约作用。心理学记忆理论认为, 大脑的工作记忆容量是有限的, 这个容量大约为 7 ± 2 个单位 (Miller, 1956)。相比之下, 人类的语言活动对认知加工提出了极高的要求。根据 de Bot (1992) 的描述, 英语母语者的平均语速大约为 150 词 / 分钟, 峰值时可达 300 词 / 分钟——这就意味着说话人每秒钟要从 3 万词左右的心理词典中进行 2 到 5 次的词汇选择, 如此高难度的加工任务显然会超出工作记忆的负荷, 因此, 人们不可能仅仅通过语法规则和词汇选择完成言语加工活动。事实上, 基于用法的语言学理论 (Jurafsky, 2002; Pawley, 2007) 认为, 人们在大量语言经验的基础之上能够对语言进行组块化 (chunking) 处理, 并将这些组块存储在长时记忆之中, 从而补偿工作记忆容量的不足 (Conklin & Schmitt, 2012; Siyanova, 2010)。此外, 越来越多的研究也表明, 人们的心理词典中并不只有词, 语块这种大于词的单位很可能也是作为一个整体存储在心理词典中。这类研究大都是基于短语的频次效应 (Arnon & Snider, 2010) 设计的。其基本逻辑是, 如果语块是作为一个整体存储加工, 那么较之于控制条件下的非语块或一般短语, 其加工速度显然会更快, 加工准确性也显然会更高, 即语块具有加工优势。