

# 词块的语义信息加工与认知心理研究

郭高攀

(东华理工大学 外国语学院 江西 南昌 330013)

**摘要:**采用文献综述法,概述了最新研究中的词块语义信息加工特点,并从认知心理和语义加工层面,以影响词块语义信息加工的主要因素为突破口,归纳出频率效应、语音短时记忆容量、词汇知识量和工作记忆广度四个最主要因素。由于后两个因素在相关研究中缺乏足够数据,继而文章采用实证研究法,通过生成性词汇知识测试和运算广度任务测试分别对测试者进行高分组和低分组分配,考察两组高分组和低分组对词块语义加工的影响。经过产出性语篇词块测试、t检验和方差关联度分析,结果表明:词汇知识量和工作记忆容量对词块语义加工过程都有影响,但词汇知识量的影响力较弱;且两因素间表现为非线性关系;关联度对词块的习得效果解释力也较小。

**关键词:**词块;语义信息加工;词汇知识量;工作记忆容量

**中图分类号:**H313 **文献标识码:**A **文章编号:**1674-3512(2015)01-0047-06

郭高攀.词块的语义信息加工与认知心理研究[J].东华理工大学学报:社会科学版,2015,34(1):47-52.

Guo Gao-pan. The semantic information processing and cognitive and psychological research of formulaic language acquisition [J]. Journal of East China Institute of Technology (Social Science) 2015, 34(1): 47-52.

## 1 背景

如何能像母语学习者一样地道而又流利地表达?解决这样的问题是我们二语学习者共同的心愿。虽不能盖以全貌,但可以肯定的是这种效果的最大特点是大量程式化的序列结构和词汇化的句式<sup>[1]</sup>,即所谓的习惯语原则<sup>[2]</sup>进行处理语言。基于这样原则,话语者能最大程度明白无误地表达既定意义,受众也能尽可能地理解既定通常所要表达的意义,组码和解码也都相对容易。因此,词块研究成为大家热议的焦点。

大多数国外相关研究主要关注词块序列的有效提取和加工效率,更有趣的是尤其注重习语的探讨<sup>[3-5]</sup>。对于习语的研究,可以总结为两个主要侧重方向:一是习语的显性意义和隐形意义启动速率对比;二是习语与非习语的语义加工速率对比。这种研究倾向引来颇多非议,首先对于我国二语学习者来说,他们很少有机会在实际语言交际过程中接

触到大量的习语,特别是隐形意义表达式;其次,在检验的语料中习语的显性意义和隐形意义存在很多模棱两可的情况,对于语义加工速率或因素的考查也存在很多争议。虽然国内最近两年从实证角度研究词块的文献逐渐多了起来,却鲜有人从认知心理和语义加工层面实证研究影响词块语义加工过程的主要因素。

## 2 词块语义信息加工特点概述

对于词块语义信息加工模式,专家早有论断,认为词块的语义部分是整体加工模式<sup>[6]</sup>,且在神经学领域也得到认同。大量的实证研究以及实验工具也佐证了这种加工模式的独特优势,如反应时任务;自定间隔式阅读;监控式复述;眼动跟踪;脑电记录。整体储存与加工的论断最直接的数据支撑来源于以上研究中词块与非词块的对比结果:用于实验组的高频词块序列比控制组的低频非词块序列反应速度快,加工时间少。用于实验的语料也多集中在两个单词序列(如 sort of and kind of)、三词SVO句式(如 I like it and I keep it);四词序列(如 don't have to worry; in the middle of);长句子中包含的词块序列(如 don't worry about it)。

收稿日期:2014-09-20

基金项目:江西省高校人文社会科学项目(YY1202)。

作者简介:郭高攀(1981—),男,湖北天门人,讲师,主要从事词汇习得、认知语言学研究。

对于词块语义信息的整体储存和加工,Wray从认知心理学角度总结出以下三个主要特点:1)词块序列内的各个单元是同步进行处理;2)语义序列直接与语音形式进行映射,整体加工途径中语义和语音形式同步;3)由于序列内各单元同步处理且反复使用,整个序列的语义加工过程变成自动加工和提取过程<sup>[1]</sup>。

部分研究者认为,母语学习者单词的认知处理能力最大限度在8~10个单词之间,然而,我们常常发现二语学习者也能毫不犹豫地說出较长甚至超过10个单词的句子,如>You can lead a horse to water, but you can't make him drink. 这表明学习者长时记忆中大量的词块积累可以有效补偿短时记忆量的不足。

综上所述,多种认知和心理实验的技术运用,各种观察和个案分析,结果都表明词块的整体储存和加工特点,无疑这个结论也佐证了掌握大量词块对于教学 and 实际交际大有裨益。然而,词块大量存在于自然的言语中,却并非天生存在于自然的大脑中。于是,我们想知道的是,如何才能有效掌握词块?词块与哪些因素最相关?解决以上问题才能真正让我们知道如何更有效促进二语学习。

### 3 影响词块序列加工过程的主要因素

#### 3.1 频率效应

大脑记忆中既有词块整体形式的心理表征,也有组成词块的单个单词的心理词汇表征。Wray和Conklin & Schmitt都认为这主要是由于频率效应的作用<sup>[4,7]</sup>。频率效应决定了各心理词汇表征间联系的紧密程度、复杂度和抽象性。认知语言学理论中的基于用法和基于范例的两种模型都表明,语言习得的最基本单元是建构,学习者的学习过程实质上是掌握一系列大量的建构序列。词块建构序列和非词块建构序列的加工过程是类似的,而影响他们加工速率的是频率。同时,我们也可以从联接主义的视角看到相似的观点。词汇与词汇表征并非孤立存在,而是以网络状形式呈现。各联接网络间使用的频率决定了联接的强度,进而也影响了学习者提取和加工的方式。

#### 3.2 语音短时记忆容量

语音短时记忆(phonological short-term memory,简称PSTM)是有效习得词汇和语法单位重要的指标。实际上,语言习得本质上是词语序列的学习,

而PSTM中词语序列组建能力是词块习得和语法构建的重要因素<sup>[8]</sup>。PSTM在言语表达过程中,对表层结构的容量和持续时间产生影响,学习者在有限时间内分析词语序列。因此,学习者在PSTM上的个体差异体现在对词语序列和语法结构的掌握上。O'Brien et al.指出PSTM能够有效帮助表达能力较差的学习者掌握一定的表达技巧,且对于表达能力较强的学习者,也能有效促进准确使用功能性词汇<sup>[9]</sup>。Hummel通过探索性因素分析讨论二语学习者的学习情况,总结了三个相关因素,即PSTM、学能和二语水平<sup>[10]</sup>。其中,PSTM与二语水平相关系数为0.35,而与二语词汇水平则达0.36。Wen对40位二语学习者表达能力中的PSTM和工作记忆的影响进行研究,发现PSTM和工作记忆都与句法复杂度、词汇多样性相关<sup>[11]</sup>。进而得出,PSTM的高容量意味着较高二语表达水平,词语的准确性、用量和复杂度都有相应提高。以上研究表明了PSTM对于词块学习的意义,它能有效将长时记忆中词块序列的心理表征的各个单元内联接、强化和启动。不过,PSTM也会因为学习者学习水平和语言类别存在个体差异。

#### 3.3 词汇知识量和工作记忆广度

词汇量是词汇能力的重要衡量标准之一,于是各种词汇量化指标和等级也随之而生。然而,这些量化的考查只是广度上的,而深度上的词汇知识(如词汇搭配和运用能力)才是词汇知识量的核心。况且,词汇量的增加和词汇知识中其他的维度并不是同步的<sup>[12]</sup>。词汇不仅仅是指单个词,而合成词、短语甚至句子也可以是词汇表中的词汇项,是许许多多的词块集合。随着语料库的兴起,人们认识到自然话语中90%都是由短语结构等多词单位结构来实现的。那么,词汇量和词汇知识量、词块加工过程之间有什么关系?

词块是一种概念化的心理机制。多词单位经过高频复现后,词块形式整体进行解码和加工处理,从而使语言使用者在短时记忆中处理更多的信息。多词单位中的词项目一旦同现后,即在记忆中彼此关联,进行关联学习。所以,当该词项中任一项目被提取时,其他项目也一同复现。语言在输入中,两个高频同现的词汇可以在长时记忆中得到关联。然而,工作记忆广度对词块语义加工有什么影响?词汇知识量和工作记忆广度之间存在正相关吗?

以往研究对此两个因素涉及较少,这是本研究力图论证的重点。

## 4 词汇知识效应和工作记忆效应实验设计

### 4.1 研究对象

实验选取某高校本科非英语专业大一年级四个平行班级 245 人,大三年级四个平行班级 239 人,共计 484 人参加实验过程。随机选取大一两个平行班级和大三的两个平行班级共 244 人参加词汇知识测试,另外四个平行班级共 240 人参加工作记忆容量测试。虽然选取研究对象为不同层次,但是人员分组是根据词块知识效应和工作记忆效应实验需要进行。

### 4.2 实验材料和步骤

整个实验设计分两个主要步骤:一是通过词汇知识量的考查将受试进行分组,进而检验对词块语义提取效果的影响;二是通过工作记忆测试将受试进行分组,然后检验对词块语义提取效果的影响。

由于词汇知识量并不是词汇量等级表给定的单个单词这么简单,而是包括词汇的广度和深度,并且语境中的运用才是最重要的界定指标。鉴于此,文章选取控制的生成性词汇知识测试<sup>[13]</sup>,将受试分为词汇知识量高分组和低分组。测试所用材料为 90 个完整句子,目标词用横线空出。该测试为避免多个目标词填入,会给出目标词的前几个字母。90 个测试目标词分别属于 5 个词频级别(2 000, 3 000, University Word List, 5 000 和 10 000),每个 1 分,满分 90 分,研究中设定高分的 25% 处和低分的 25% 处确定高分组和低分组。所有受试都是在课堂时间完成,大约半小时。

为了分析工作记忆容量对词块的语义提取过程的影响,研究对受试进行工作记忆测试。测试的工具为运算广度任务<sup>[14]</sup>,但是根据实验目的的需要,中国学习者的记忆力效应应该用母语(中文)测试比较合理,实验参照了李红<sup>[15]</sup>的中文词组。首先,给受试者在电脑屏幕上呈现一串运算数据,如  $(5-3) \times 6 = 12?$  要求受试者朗读数据串并判断结果的正误,同时按下键盘上的“Y”或者“N”;随即数据串消失后,新的一页出现,呈现一个中文词组,如“皮肤”,受试者被要求朗读,随即研究者呈现下一页,出现新的一串运算数据,以此类推。共有 60 个运算数据串,相应的配以 60 个中文词组,被分成 15 个系列组,每组含有 2 至 6 个不等的运算串和词

组。受试者在没有记忆词组的暗示下,按照词组呈现顺序回忆词组内容并在准备好的纸张上写下。正确按照每个系列组中词组的呈现顺序记下一个词组计 1 分,记忆维度可以是 0~60 分。研究中设定高分的 25% 处和低分的 25% 处确定高记忆组和低记忆组。每个受试者大约需要 15 分钟,由于人数众多且耗时量大,故测试都安排在非上课时间,为了让学生认真和注意力集中,我们安排了考勤和期末总评奖励制度。

在检验词块语义加工过程中,考察方式采用产出性测试<sup>[16]</sup>。产出性测试篇章如:

I \_\_\_\_\_ is clear that the more you use English,

the better you will learn it. There is no disagreement about this. (this is obvious)

[Answer: It is clear]

为了让测试者给出符合语境的目标词块且不造成歧义,词块的开头字母已经给出,且在右边给出了符合语境的解释。

测试所使用的词块也参照了 Schmitt et al. 选用的标准<sup>[16]</sup>,选用的词块符合一定的高频复现,同时满足学生所学教材和阅读的各种主题材料,包括口语、书面和专业文献各种内容,避免所用材料的片面性。参照的语料库选用 CANCODE (Cambridge and Nottingham Corpus of Discourse in English) 进行口语材料的词块频率检测,BNC 语料库则进行书面语材料中的词块频率检测,频率满足每百万词 50 至 100 频率间。基于以上文献参考,语料库词频运用和参照学生所学材料等因素考虑,最终确定 20 个词块。

## 5 实验结果与分析

首先,研究考察了第二语言词汇知识量对语篇中产出性词块语义加工过程的影响情况。将词汇知识测试的分数进行降序排序。244 人的 25% 应是 61,而第 61 名的分数为 72,所以判断高分组为 72 分以上。同理,判定低分组为 52 分以下。经过独立样本 t 检验(表 1),表明高分组和低分组之间存在显著性差异。说明高低组间有较好的区分度。接着,让分好组的受试者参加产出性词块测试。所有测试者都在规定时间内完成并提交测试材料,而且低分组的受试并没有像我们预测中不能按时完成或完成后质量不高。经过对材料的数据统计并

采用独立样本 t 检验(表 2),结果表明这两组成绩的均值存在差异性  $p = 0.045 < 0.05$ 。但从数据分析中不难发现,彼此间的差异性并不是很明显。这也说明三年级比一年级的学生在词汇知识量上虽然占有一定优势,但是对于语料中篇章的词块学习和掌握情况并没有呈现绝对优势。

高词汇知识量并不能一定有效产生词块的习

表 1 词汇知识量测试成绩高分组与低分组统计量

高分组与低分组		N	均值	标准差	均值的标准误差
词汇知识量成绩	高	61	78.065 6	4.419 16	.565 82
	低	61	43.836 1	6.621 13	.847 75

表 2 词汇知识量测试成绩分组后进行的产出性词块测试

Levine 检验				t 检验				
F	Sig.	t	df	Sig. (双侧)	均值 差值	标准 误差	95% 置信区间 下限 上限	
词块测试		2.142	114.51	.045	1.147 5	.535 69	.086 39	2.208 6

其次,研究还考察了工作记忆容量对语篇中产出性词块语义提取影响情况。实验中,受试 240 人的 25% 应是 60,而第 60 名的分数为 50,所以判断高分组为 50 分以上。同理,判定低分组为 39 分以下(表 3)。通过独立样本 t 检验,表明工作记忆容量的高分组和低分组之间存在显著性差异 ( $t(120) = 29.75, p = .000 < .01$ )。说明高低组间有较好的区分度。经过分组后进行的产出性词块测试,并对数据整理后进行独立样本 t 检验(表 4),结果表明工作记忆容量的高分组与低分组间存在显著差异性 ( $p = .002 < 0.05$ )。这说明工作记忆容量对语篇中产出性词块语义提取影响较大,记忆的

表 3 工作记忆容量测试成绩高分组与低分组统计量

高分组与低分组		N	均值	标准差	均值的标准误差
记忆容量成绩	高	60	52.700 0	2.125 71	.274 43
	低	60	31.300 0	5.149 63	.664 81

表 4 工作记忆容量测试成绩分组后进行的产出性词块测试

Levine 检验				t 检验				
F	Sig.	t	df	Sig. (双侧)	均值 差值	标准 误差	95% 置信区间 下限 上限	
词块测试	1.099	.297	3.227	118	.002	1.483 3	.459 7	.572 9 2.393 6

最后,研究还对词汇知识与记忆容量关联度(表 5)进行分析。通过 Means 过程的单因素方差分析,我们发现组间存在显著性差异,说明四个水平中,至少有一组与其他组有差异 ( $P = .003 <$

得和运用,而词块却是语篇中主要的构成部分。因此目前高年级学生写作质量不高,将原因归咎于单词量的识记也不妥。词块是语义和语用很好的结合,在使用中整体提取的优势明显,有效简化输出任务,降低工作记忆压力。所以积累和使用大量的词块可以极大地提高写作产出质量<sup>[17]</sup>。

广度对词块的语义加工有显著的促进作用。同时,我们也可以推断,工作记忆容量可以有效促进长时记忆中信息的提取和加工。

工作记忆容量取决于信息能否被快速提取的状态。根据认知心理学中工作记忆理论,通常人们只能一次记忆 5~7 个基本信息单位,超过的信息量会在大脑中逐渐消退。如果这 7 个基本信息单位以较大的词块单位进行存储,并能在广泛的语料中增加使用频率,这将非常有利于词块的掌握和促进语言学习产出的质量。比如听力、口译中,如果学习者能以词块为单位进行记忆,这将增大工作记忆的容量,降低认知负荷,产生良好缓解效应。

0.05)(表 6)通过以上记忆工作容量的独立样本 t 检验结果中我们可以预测这一组应该是记忆工作容量高分组。同时,结果中 Linearity 和 Deviation from Linearity 也可以分析出各变量间同时存在线

性关系 ( $0.039 < 0.05$ ) 和非线性关系 ( $0.008 < 0.05$ )。这说明词汇知识量高分组和低分组间总体还是线性走向,记忆工作容量高分组和低分组间也是线性走向。但是词汇知识量与记忆工作容量间却是呈 V 型走向的。表 7 中关联度测量表证明以上分析的合理性。由于 R 值远远小于 1 ( $R = 0.130$ ) 表明词汇知识量与记忆工作容量间关联度并不高; R 方也说明这个关联度对词块测试结果的解释力较小。

表 6 词汇知识与记忆容量关联度的 ANOVA 分析

		平方和	df	均方	F	显著性
(组合)		107.758	3	35.919	4.754	.003
词块成绩 *	组间					
	线性	32.459	1	32.459	4.296	.039
词汇与记忆	线性偏差	75.299	2	37.650	4.983	.008
	高低组					
	组内	1798.412	238	7.556		
	总计	1906.169	241			

表 7 词汇知识与记忆容量相关性度量分析

	R	R 方	Eta	Eta 方
词块成绩 * 词汇与记忆	.130	.017	.238	.057
高低组				

## 6 结论

通过对词块习得过程进行文献综述,概述了最新研究中的词块语义信息加工特点,并从认知心理和语义加工层面研究影响词块提取与语义信息加工的主要因素,归纳出词块序列加工过程中的频率效应、语音短时记忆容量、词汇知识量和工作记忆广度四个最主要因素。由于词汇知识量和工作记忆广度在相关研究中缺乏足够数据,继而文章采用实证研究法,通过生成性词汇知识测试和运算广度任务测试分别对所有测试者进行高分组和低分组分配,分别考察两组高分组和低分组对词块语义加工的过程。经过产出性语篇词块测试、t 检验和方差关联度分析,结果表明,词汇知识量和工作记忆容量对词块语义加工过程都有影响,但词汇知识量的影响力较弱;且词汇知识量和工作记忆容量间表现为非线性关系,关联度对词块的习得效果解释力较小。

### [参考文献]

[1] Wray, A. What Do We (Think We) Know About Formulaic Language? An Evaluation of the Current State of Play [J]. *Annual Review of Applied Linguistics*, 2012 (32): 231-254.

表 5 词汇知识量与记忆容量关联度描述性统计

词汇与记忆高低组	均值	N	标准差
词汇知识量高分组	12.3607	61	3.26615
词汇知识量低分组	11.2131	61	2.61480
记忆容量高分组	12.3667	60	2.64874
记忆容量低分组	10.8833	60	2.37994
总计	11.7066	242	2.81237

[2] Sinclair, J. M. Collocation: A progress report. In R. Steele & T. Threadgold (Eds.), *Language topics: Essays in honour of Michael Halliday* [C]. Amsterdam, the Netherlands: John Benjamins, 1987.

[3] Underwood, G., Schmitt, N., & Galpin, A. The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences* [C]. Amsterdam, the Netherlands: John Benjamins, 2004.

[4] Conklin, K., & Schmitt, N. The Processing of Formulaic language [J]. *Annual Review of Applied Linguistics*, 2012 (32): 45-61.

[5] Siyanova-Chanturia, A., Conklin, K., and Schmitt, N. Adding more fuel to the fire: An eye-tracking study of idiom processing by native and nonnative speakers [J]. *Second Language Research*, 2011(27): 1-22.

[6] Wray, A. *Formulaic language and the lexicon* [M]. Cambridge, UK: Cambridge University Press, 2002.

[7] Wray, A. *Formulaic language: Pushing the boundaries* [M]. Oxford, UK: Oxford University Press, 2008.

[8] Ellis, N. C. Formulaic Language and Second Language Acquisition: Zipf and the Phrasal Teddy Bear [J]. *Annual Review of Applied Linguistics*, 2012 (32): 17-44.

[9] O'Brien, I., Segalowitz, N., Collentine, J., & Freed, B. Phonological memory predicts second language oral fluency gains in adults [J]. *Studies in Second Language Acquisition*, 2007(29): 557-581.

[10] Hummel, K. M. Aptitude, phonological memory, and second language proficiency in nonnovice adult learners

- [J]. *Applied Psycholinguistics*, 2009(30): 225-249.
- [11] Wen, Z. *Working memory and second language learning* [M]. Bristol, UK: Multilingual Matters, 2011.
- [12] 吴旭东, 陈晓庆. 中国英语学生课堂环境下词汇能力的发展[J]. *现代外语*, 2000(4): 349-360.
- [13] Laufer, B., & Nation, P. A vocabulary-size test of controlled productive ability [J]. *Language Testing*, 1999(16): 33-51.
- [14] Conway, A. R. A., & Engle, R. W. Individual differences in working memory capacity: More evidence for a general capacity theory [J]. *Memory*, 1996(6): 577-590.
- [15] 李红. 语义提取: 基于第二语言词汇能力角度的研究[M]. 北京: 高等教育出版社, 2007.
- [16] Schmitt, N., Dörnyei, Z., Adolphs, S., and Durow, V. Knowledge and acquisition of formulaic sequences: A longitudinal study [C]. In Schmitt, N. (ed.), *Formulaic Sequences: Acquisition, Processing, and Use*. John Benjamins Press, 2004.
- [17] 丁言仁, 戚焱. 词块运用与英语口语和写作水平的相关性研究[J]. *解放军外国语学院学报*, 2005(5): 49-53.

## The Semantic Information Processing and Cognitive and Psychological Research of Formulaic Language Acquisition

GUO Gao-pan

(*College of Foreign Languages, East China Institute of Technology, Nanchang 330013, China*)

**Abstract:** Based on the process of literature review, this article outlines the latest research on the semantic information processing characteristics of formulaic language, and find out the main factors that affect semantic processing as a breakthrough. We conclude the most important four factors: frequency effects, phonological short-term memory capacity, the amount of vocabulary knowledge and working memory span. Since the last two factors lacking of adequate research data, this article uses empirical research method. Throng productive vocabulary knowledge testing and operational span tasks, the high and low groups are grouped. Based on the productive discourse chunks test, t test and variance correlation analysis, the results show that: both of the two factors have an impact on the semantic processing process, but the influence of the vocabulary knowledge is weak; there is a non-linear relationship between factors and the explanatory power of the correlation effects on the acquisition of chunks is small.

**Key Words:** formulaic language; semantic processing; vocabulary knowledge; working memory span