

频率和语言水平对公式化语言整体加工的作用

李红¹, 廖道蓉²

(1. 重庆大学 语言认知及语言应用研究基地, 重庆 400044; 2. 西华大学 外国语学院, 四川 成都 610039)

摘要: 研究主要考察公式化语言频率及语言水平是否影响公式化语言在中国英语学习者二语心理词汇的整体表征和提取。文章选取高低水平两组受试, 他们先后完成了反应时实验和二语词频检测任务。使用 E-Prime 软件和 Praat (4.3.04 版) 语音分析软件对数据进行采集和分析, 其结果仅显示了公式化语言频率对整体加工的作用, 没有发现学习者的语言水平对公式化语言整体加工的影响。

关键词: 公式化语言; 二语心理词汇; 语言水平; 频率

中图分类号: H319.1 **文献标志码:** A **文章编号:** 1008-5831(2009)05-0124-05

一、问题的提出

公式化语言(formulaic sequences), 包括各种预制的或半预制的语块, 是语言的一个重要组成部分。Wray^[1]把公式化语言定义为: 一个由单词或成份组成的连续或非连续的预制的序列。这些语言序列是以整体的形式在心理词汇中储存的, 人们在提取它们时, 不是通过语法规则即时生成这些预制或半预制的语块。很多语言学家认为日常语言大部分都是公式化语言。现有的研究^[2-3]主要基于语料库对这一语言现象进行描述性分析和研究, 从心理语言学角度开展的研究则较少。但是对公式化语言在心理词汇中的表征和提取的研究对进一步理解公式化语言的认知加工过程却有着重要的意义。

在一语心理词汇研究中, 对于规则复杂词形(如规则屈折词形, 派生词形, 甚至短语等)的存取方式主要有两大理论模型:(1)双处理模型^[4](dual-processing model); (2)单处理模型^[5](single processing model)或称为基于频率的处理模式(frequency-based processing model)。Pinker^[4]认为规则复杂词形(如 formed, careful)在人的大脑中不能以整词的形式存取, 人们每次提取这些规则复杂词形时需要通过规则在线组合而成(即 form 和 -ed; care 和 -ful 都分开储存在人的大脑里)。而 Bybee 认为心理词汇在大脑中的表征和提取形式主要由其词频决定, 并认为高频的规则复杂词形在心理词汇中具有整词表征的特性(即 formed 以一个整词的形式储存)。当高频规则复杂词形的词汇强度胜过其词汇连接时, 则可以被整体提取; 当低频规则复杂词形低于其词汇连接时则没有整词效应^[6]。一些研究^[7]支持双处理模型, 而较多的实验结果, 证明高频规则复杂词形在心理词汇中有整词效应^[8-9]。

收稿日期: 2009-02-08

基金项目: 重庆市教委人文社科重大项目(07SK159); 重庆大学语言认知及信息处理研究所 2004 年专项研究基金“基于人工神经网络的英汉双语心理词汇研究”

作者简介: 李红(1962-), 女, 重庆人, 重庆大学外语学院教授, 博士, 主要从事二语习得研究。
欢迎访问重庆大学期刊社 <http://qks.cqu.edu.cn>

Sosa 和 MacFarlane^[10]认为如果高频规则复杂词形在心理词汇中有整词效应,那么更大的语言单位(即公式化语言)在心理词汇中也可能以整体的形式存取。针对英语为本族语者的研究显示了公式化语言相对于非公式化语言在心理词汇中的加工优势;公式化语言的频率是影响其在心理词汇中表征和提取的一个重要因素之一^[11]。Sosa 和 MacFarlane^[10]以英语本族语者为受试观察到高频双字词短语的整体加工效应。受试的任务是在含有“of”的双字词短语的电话录音中听取句中的“of”。他们发现受试对高频双字词短语的反应时显著慢于低频双字词短语的反应时。但是也有一些实验并未发现公式化语言的整体加工效应^[12]。因此,公式化语言的整体加工效应需要进一步研究。

笔者考察了频率以及学习者的语言水平对公式化语言在二语心理词汇中的整体加工作用。研究问题:(1)公式化语言的频率是否影响其在二语心理词汇中的表征和提取?如果会,相对于低频公式化语言而言,高频公式化语言能否以整体的形式在二语心理词汇中存取?(2)学习者的语言水平是否会影

响公式化语言在二语心理词汇中的表征和提取?如果会,相对于较低水平的学习者而言,相同频率的公式化语言是否在高水平学习者的二语心理词汇中的整体加工效应更容易获得?

二、研究方法

(一) 受试

重庆大学 42 名英语专业一年级学生以及 39 名一年级和二年级英语专业的研究生参加了实验。根据英语专业一年级第一学期期末的综合英语考试成绩,选取了分数为 72 至 78 分段的 42 名低水平受试。高水平组则选取通过英语专业八级证书考试的研究生。一年级研究生专业八级的平均分为 64.23;二年级研究生专业八级的平均分为 65.23。两个年级的平均分没有显著性的差异($p = 0.43$)。除去未按要求操作的受试,低水平组进入数据统计的受试为 35 人,高水平组 36 人。

(二) 实验材料

反应时实验的材料基于 Sosa 和 MacFarlane^[10]中的 24 个含“of”的双字词公式化语言(不包括习语,见表 1)。

表 1 Sosa 和 MacFarlane (2002) 研究中的 24 个含“of”的实验材料

低频频率 (1-58)	中低频频率 (175-281)	中高频频率 (308-479)	高频频率 (889-3592)
sense of	care of	couple of	kind of
piece of	because of	part of	lot of
sums of	kinds of	most of	one of
each of	bit of	all of	out of
example of	any of	think of	sort of
colleague of	much of	type of	some of

Sosa 和 MacFarlane 根据“Switchboard Corpus”语料库把这 24 个含有“of”的双字词短语分成四组。为了不让受试的听力水平影响他们听取句中的“of”,我们对实验材料进行了简化。把这 24 个含有“of”的公式化语言分别镶嵌在 24 个句子结构较为简单的句子中(如 They couldn't go outing yesterday because of the heavy rain.)。在重庆大学任教的一名外籍教师的帮助下,我们把这 24 个句子录为 24 个独立的语音文件^①。然后把这 24 个句子分成四个频率组。利用 Praat (4.3.04 版)语音分析软件剪辑掉每句录音前后的空白音部分之后,对这四组录音句子的长度进行单因素方差分析(低频组句子长度为 4.64 秒;中低频组为 4.45 秒;中高频组为 3.81 秒;高频组为 3.84 秒),结果表明各频率组的句子录音长度没有显著差异: $F(3, 20) = 2.93, p = 0.059 >$

0.05。另外,为了让受试在正式实验前熟悉实验操作过程,实验还有四句镶嵌有“of”的公式化语言的句子录音作为实验练习材料。

另一组实验材料是为二语词频检测任务而设计的。该任务的目的是检查受试对表 1 中所列的双字词公式化语言的使用频率。我们让受试根据自己的语言使用情况来判断每个含有“of”的公式化语言的频率是属于哪个类别。

(三) 实验步骤

受试坐在电脑前听取实验句子中的“of”,若受试听到“of”则立即按下键盘上的“1”键;若到句子结束时还未听到“of”,受试则按下键盘上的“2”键。在正式实验开始前,受试首先阅读实验指令,在较为熟悉实验任务后,受试开始实验练习。实验练习和正式反应时实验程序设计相同。受试在操作熟练后进

①录音按 Sosa 和 MacFarlane(2002 年)的要求进行:(1)句中“of”前不能有明显停顿;(2)句中“of”的起始音能够被分辨;(3)“of”不能位于句首和句尾。

入正式实验。全部程序用2002年美国心理学软件公司出版的E-Prime(version 1.0)软件设置。反应时实验结束后,受试进行二语词频检测任务。每个受试完成反应时实验和二语词频检测任务大概需要20分钟。

三、结果分析

(一)二语词频检测结果分析

由于目前中国没有相关的含“of”的公式化语言词频的学习者语料库,所以本实验依据Sosa和MacFarlane^[10]实验中的一语词频来检测二语词频与母语词频之间是否存在差异。

根据受试对实验材料判断的结果把24个含有“of”的公式化语言分成四个频率组(即低频,中低频,中高频及高频)。低水平和高水平组的结果分别列在表2和表3中。

表2 低水平组词频检测结果

低频	中低频	中高频	高频
sense of	care of	most of	kind of
couple of	part of	each of	lot of
sums of	kinds of	piece of	one of
type of	bit of	out of	because of
sort of	example of	think of	all of
colleague of	much of	any of	some of

表3 高水平组词频检测结果

低频	中低频	中高频	高频
care of	kinds of	most of	lot of
sense of	couple of	piece of	all of
example of	any of	kind of	out of
type of	sort of	each of	because of
sums of	bit of	part of	one of
colleague of	much of	think of	some of

由表2、表3可见,两个水平组的词频检测结果很相似。虽然两个水平组词频检测结果同母语频率组(见表1)有些差异,如sort of, type of, each of, piece of和because of,但我们得到的大部分二语词频结果和Sosa和MacFarlane采用的母语频率分布很相似。由于本实验中二语词频检测任务是在小范围内进行,并主要基于受试的直觉判断,所以后面的结果分析还是基于表1中的频率分组,而二语词频检测结果只是用于解释本实验的数据结果。

(二)反应时实验分析

1. 正确率数据分析

我们分别对两个水平组的正确率数据进行了单因素的方差分析,结果显示在高水平组中,四个频率段的正确率之间不存在显著差异($F = 1.160$; $p = 0.327$)。同样,在低水平组中,四个频率段的正确率之间也不存在显著差异($F = 1.424$, $p = 0.238$)。我

们还分别对两个水平组在每个频率段的正确率数据进行了独立样本t检验,都未显示显著差异(低频 $t = -0.579$; $p = 0.565$; 中低频 $t = 0.684$; $p = 0.496$; 中高频 $t = -0.220$; $p = 0.827$; 高频 $t = -1.254$; $p = 0.214$)。

表4 高低水平组在每个频率段的正确率(%)

组名	频率段			
	低频	中低频	中高频	高频
高水平组	82.28	79.58	86.97	82.83
低水平组	79.94	82.43	86.03	76.06

2. 反应时数据分析

我们分别对两个水平组的正确反应时数据进行了分析。如图1所示,在低水平组中,低频公式化语言中“of”的平均反应时为641.89毫秒,中低频为665.62毫秒,中高频为658.06毫秒,高频为754.93毫秒。单因素方差分析结果表明该组受试对四个频率段的平均反应时之间存在显著的差异($F = 4.896$, $p = 0.002 < 0.05$)。Scheffe事后检验分析显示,受试对高频公式化语言中“of”的平均反应时显著慢于低频公式化语言中“of”的平均反应时($p = 0.009 < 0.05$)和中高频率的公式化语言中“of”的平均反应时($p = 0.022 < 0.05$),但是在高频和中低频率的公式化语言中“of”的平均反应时之间不存在显著差异($p = 0.062$)。

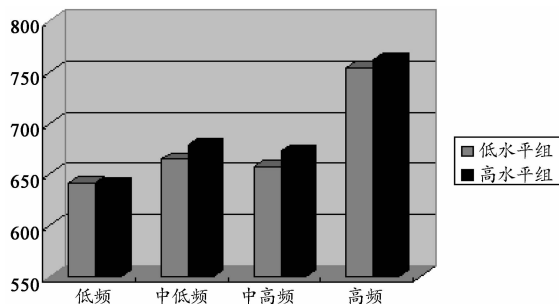


图1 高低水平组四个频率段的平均反应时

在高水平组中,低频公式化语言中“of”的平均反应时为640.49毫秒,中低频为679.39毫秒,中高频为673.39毫秒,高频为763.36毫秒。同样,我们对该组的正确反应时数据进行了单因素的方差分析,结果显示四个频率段的平均反应时之间也有显著差异($F = 5.456$, $p = 0.001 < 0.05$)。Scheffe事后检验分析显示高频的公式化语言中“of”的平均反应时显著慢于低频的平均反应时($p = 0.002 < 0.05$)和中高频率的平均反应时($p = 0.035 < 0.05$),但是高频率的公式化语言中“of”的平均反应时并没有显著慢于中低频率的平均反应时($p = 0.076 > 0.05$)。

化语言中“of”的平均反应时是否有显著的差异,我们又分别对每一频率段中两个水平组的平均反应时进行了独立样本t检验分析,结果并未发现显著的差异(低频 $t=0.042$; $p=0.967$); 中低频 $t=-0.368$; $p=0.713$; 中高频 $t=-0.580$; $p=0.562$; 高频 $t=-0.267$; $p=0.789$)。

四、讨论

(一)关于反应时实验的正确率结果

本实验中受试对公式化语言中“of”的正确率分析结果与 Sosa 和 MacFarlane^[10]的研究结果不一致,两个水平组的正确率都明显高于 Sosa 和 MacFarlane 实验中观察到的平均正确率(45%)。就频率效应而言, Sosa 和 MacFarlane 发现受试对母语双字词短语中“of”的正确率随着频率的增加而降低。他们认为这是因为他们实验中采用自然的对话录音以及频率效应。也就是说,如果高频双字词短语能以整体的形式储存在心理词汇中,人们听高频双字词短语中的“of”要比听低频中的“of”难,这样高频组中的正确率就相对低。

但是本实验并未显示频率效应,两个水平组都一致显示四个频率段的正确率之间不存在显著差异。正确率没有随着公式化语言的频率增加而降低,并且,就同一频率的公式化语言而言,高水平受试对“of”反应时的正确率也没有低于低水平的受试。一个可能的原因是本实验采用的24个含of的公式化语言的句子实验材料既不长也不复杂。因此,对于两个组的受试而言,听辨出句中的“of”相对较为容易,正确率就相对较高,并且也使高水平组受试对四个频率段的公式化语言中“of”的平均反应时的正确率和低水平组受试的正确率较为接近。

(二)关于反应时实验的反应时结果

在本实验中,无论是低水平组还是高水平组,受试对高频公式化语言中“of”的平均反应时都显著慢于低频以及中高频公式化语言中“of”的平均反应时,但没有显著慢于中低频率的平均反应时。导致这样结果的原因可能是本研究实验材料在母语频率与二语使用频率之间存在差异。在 Sosa 和 MacFarlane 的研究中,“because of”的频率属于中低频率段。而根据本实验受试的二语词频检测结果(见表2和表3)“because of”都属于高频。事实上受试对“because of”中的of的平均反应时都很慢:低水平组为842.20毫秒,高水平组为877.20毫秒,都明显慢于其所在的中低频率段的平均反应时(低水平组对中低频率公式化语言中“of”的平均反应时为665.62毫秒;高水平组对中低频率公式化语言中“of”的平均反应时为679.39毫秒)。如果把“because of”从中低频组转到高频组,并再次对

这两个频率的公式化语言中“of”的平均反应时做配对样本t检验分析,两个英语水平组都有显著性的频率效应(低水平组: $t=-3.661$, $p=0.000<0.05$;高水平组: $t=-4.065$, $p=0.000<0.05$)。

本实验的反应时结果显示了公式化语言的频率在二语心理词汇中对整体加工的作用,高频的公式化语言倾向于以整体的形式在二语心理词汇中表征和提取。

就语言水平效应来讲,反应时的分析结果并未发现显著的差异。虽然高水平受试对各频率段公式化语言中“of”的平均反应时稍慢于低水平受试(除低频组外),但是差异并不显著。因此本实验结果未能证实学习者的语言水平对公式化语言整体加工的作用,也未能发现同一频率的公式化语言更容易以整体的形式储存在语言水平相对较高的学习者心理词汇中。

五、结论

本研究初步显示了公式化语言的频率是影响其在二语心理词汇中整体储存和提取的因素之一,高频公式化语言倾向于以整体的形式储存在二语心理词汇中,但没有证实二语学习者的语言水平对公式化语言在二语心理词汇中的整体表征和提取作用。

由于公式化语言是英语的一个主要组成部分,根据一些研究者^[2-3]的观点,语言产出不是句法上的基于规则的加工过程,而是从记忆中提取更大的短语单位(即公式化语言)。我们应在英语学习过程中,注意公式化语言的习得,尤其要注意公式化语言的频率以及使用等问题。

本研究还存在以下局限:由于目前还没有标注中国英语学习者使用含有“of”的公式化语言的专门语料库,因此本研究的结果主要基于英语本族语的词频。虽然本实验的二语词频检测任务结果表明这些双字词公式化语言的二语词频分布和一语词频分布很相似,但是也有一些差异,可能会影响到本研究的结果。另外,在本研究中,虽然受试对高频含有“of”的公式化语言的平均反应时显著慢于其他频率组的平均反应时,但是我们并不知道这些公式化语言能以整体的形式储存在二语心理词汇中的最低频率限度是多少,这需要进一步的研究。

参考文献:

- [1] WRAY A. Formulaic language and the Lexicon[M]. Cambridge: Cambridge University Press, 2002.
- [2] NATTINGER J R, DECARRICO J S. Lexical phrases and language teaching [M]. Oxford: Oxford University Press, 1992.

- [3] SINCLAIR J. Corpus, Concordance, Collocation [M]. Oxford: Oxford University Press, 1991.
- [4] PINKER S. Rules of language [J]. Science, 1991, 253: 530 - 535.
- [5] BYBEE J. Regular morphology and the lexicon [J]. Language and Cognitive Processes, 1995, 10(5): 425 - 455.
- [6] 李红, 缪道蓉. 规则屈折词形在心理词汇中的表征与提取—基于词频还是规则[J]. 外国语言文学研究, 2004(4): 44 - 48.
- [7] PRASADA S, PINKER S. Generalization of regular and irregular morphological patterns [J]. Language and Cognitive Processes, 1993(8): 1 - 56.
- [8] ALEGRE M, GORDON P. Frequency effects and the representational status of regular inflections [J]. Journal of Memory and Language, 1999(40): 41 - 61.
- [9] FAROQI-SHAH Y, Thompson C K. Regular and irregular verb inflections in agrammatism: Dissociation or association [J]. Brain and Language, 2003(87): 9 - 10.
- [10] SOSA A V, MACFARLANE J. Evidence for frequency - based constituents in the mental lexicon: Collocations involving the word of [J]. Brain and Language, 2002(83): 227 - 236.
- [11] UNDERWOOD G, SCHMITT N, Galpin, A. The eyes have it: an eye-movement study into the processing of formulaic sequences [C]// schmitt. Formulaic sequences: Acquisition, processing and use. Philadelphia: John Benjamins Publishing Company, 2004: 153 - 172.
- [12] SCHMITT N, UNDERWOOD G. Exploring the processing of formulaic sequences through a self - paced reading task [C]// SCHMITT. Formulaic sequences: acquisition, processing and use. Philadelphia: John Benjamins Publishing Company, 2004: 173 - 189.

Effects of Frequency and Language Proficiency on the Holistic Representation and Access of Formulaic Sequences

LI Hong¹, MIAO Dao-rong²

(1. College of Foreign Language, Chongqing University, Chongqing 400044, China;

2. School of Foreign Language, Xihua University, Chengdu 610039, China)

Abstract: This study investigates whether formulaic sequences' frequency and the language proficiency of EFL Chinese learners would affect the holistic representation and access of formulaic sequences in the L2 mental lexicon. Two groups of subjects with different levels of English proficiency were chosen, and they were asked to complete a reaction time task and a L2 frequency checking task. E-prime software and Praat (version 4.3.04) speech analysis software were adopted in the present study. The findings revealed the effect of frequency but they did not show the effect of language proficiency in the holistic status of formulaic sequences in the L2 mental lexicon.

Key words: formulaic sequences; L2 mental lexicon; language proficiency; frequency

(责任编辑 胡志平)