

• 研究前沿(Regular Articles) •

## 语块的心理现实性\*

易 维 鹿士义

(北京大学对外汉语教育学院, 北京 100871)

**摘 要** 语块是一个大于词的多词序列, 是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与意义的组合。阅读加工研究和听觉加工研究表明, 语块和词一样也是心理词典的一个单位, 是整体存储加工的, 语块具有心理现实性。未来的研究可以对语块的界定、语块及其成分词的关系、语块的动态发展以及二语者语块的心理现实性问题进行深入的探讨。

**关键词** 语块; 心理现实性; 认知加工

**分类号** B842

### 1 引言

长期以来, 语言一直被视为由词汇和语法这两个不同的模块构成的一个系统。单个的词是由语素、音素、字母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。在本文中, 我们采用的是“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, 例如习语、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, 语块具有一定的频率性, 语块的分布受到频率等因素的影响(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。尽管不同研究中的统计数据各有差异, 但语言使用中存在着大量的语块是显而易见的。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

收稿日期: 2013-02-01

\* 国家社会科学基金项目(基金号: 12BYY059)资助。

通讯作者: 鹿士义, E-mail: lushiyi@pku.edu.cn

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, 语块对第一语言和第二语言习得都具有重要的作用。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系, 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, 并对习语的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, 人们对频率、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, 频率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率效应。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, 频率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率效应。Arnon 和 Snider (2010)利用短语判断任务对频率分别为高、中、低三段的若干 4 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率效应, 从而证实大脑中存储着短语的频率信息, 这为心理词典中存在语块的表征提供了一定的证据, 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。语块是否也有可能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。

## 2 语块整体存储和加工的心理现实性

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、存储和提取加工有重要的制约作用。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  $7 \pm 2$  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, 峰值时可达到 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,人们的心理词典之中并不只有词,语块这种大于词的单位很可能也是作为一个整体进行加工,存储于心理词典之中的。这些实验研究大部分是基于短语的频率效应(Arnon & Snider, 2010)进行设计的,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,其加工速度要显著更快,加工准确性也会显著更高,即语块具有加工优势。

### 2.1 阅读中的语块加工优势

许多研究都表明,语块在阅读中具有加工优势。Tremblay, Derwing 和 Libben (2009)利用三个自定速阅读实验,将实验语句分别以逐词、逐块和逐句等三种不同的方式呈现给被试。研究者假设,一个9词语句,如果其中包含一个整体存储加工的4词语块,那么被试在阅读加工时就只需要计算6个单位,因而其阅读速度也将显著快于由9个彼此独立的词构成的普通语句。研究结果证实了上述假设,他们发现,逐块呈现和逐句呈现条件下,包含语块的句子都要比不包含语块的句子在阅读加工上显著更快。

自定速阅读只是自然阅读的一种模拟。研究发现,自然阅读中包含了10-15%的回视(Rayner & Pollatsek, 1994)。自定速阅读迫使被试的眼球只能按从左至右的方式做前进式运动,这种情况与实际阅读中的眼动情况存在较大差异,并可能影响到实验观察和实验结果。有鉴于此,采用眼动技术来研究语块的阅读加工,实验效果无疑会好得多。Underwood, Schmitt 和 Galpin (2004)利用眼动技术考察了语块在阅读中的末词(*terminal word*)加工优势。他们比较了语块中的末词和非语块条件下目标词的加工情况,发现母语者被试对语块末词的注视次数要显著少于非语块条件下的目标词。该结果一定程度上支持了语块整体存储加工的假设,但实验并未发现母语者在加工语块末词时存在跳读,也未能揭示末词加工在注视时长上的优势,因而说服力还比较有限。为此, Siyanova-Chanturia, Conklin 和 Schmitt (2011)从语义加工的

角度入手,利用眼动技术进一步考察了语块在阅读中的加工优势。他们分析了21个高频习语在字面义和习语义用法条件下相对于控制短语的加工优势,结果发现,无论是按字面义还是习语义使用,习语的整体阅读时间都要比一般短语显著更快,注视次数也都要比控制短语显著更少。习语具有结构稳定性和语义凝固性的特点,也是语块研究中的一个重要部分,因此该实验进一步证实了前述两项研究(Tremblay et al., 2009; Underwood et al., 2004)的结论,即语块在阅读加工中具有较为明显的加工优势,语块在阅读加工中是作为一个整体存储加工的。

阅读中的语块加工优势还得到了病理语言学研究的支持。Arcara 等人(2012)考察了左忽视性阅读障碍病人对固定二项式短语(*irreversible binomials*)的阅读加工情况。忽视性阅读障碍通常是由于大脑右半球受损而引起的,典型的情况多为左忽视性阅读障碍。病人在阅读字母、词和词串时,他们会典型地忽视左边的部分,因而时常出现遗漏、误读、替换等错误。研究者认为,固定二项式短语在心理词典中很可能是整体存储和加工的,因此他们预测,忽视性阅读障碍病人在阅读固定二项式短语时的表现应该要比非固定二项式短语(*reversible binomials*)更好。该实验发现被试在阅读固定二项式短语时的表现确实要比非固定二项式短语显著更好,从而证实了固定二项式短语作为语块整体存储和加工的假设。

### 2.2 听觉监控中语块的加工优势

Ellis (2002)认为,语言加工对频率信息非常敏感, Wray (2002)也认为频率是界定语块的一项基本标准。相比于一般短语,高频短语(语块)具有明显的加工优势,这一点在听觉监控实验中已经基本得到了证实。

Sosa 和 MacFarlane (2002)利用词汇监控范式验证了包含“of”的语块的心理现实性。他们首先将包含“of”的2词语块嵌入到自然会话材料之中,然后要求被试在听到“of”一词时尽快做出按键反应。研究者假设,如果高频的“of”词串是整体存储加工的语块,那么被试在对“of”进行监控时,将会有个分解加工(*decomposition*)的过程,也就是说被试需要从整个词串中将“of”重新分离出来。因此,相比于频率较低的一般词串,高频词串的监控反应速度将会显著更慢,准确率也将显著

更低。实验结果证实了这一假设, 从而表明高频的“of”词串确实是整体存储和提取加工的。

由于 Sosa 和 MacFarlane (2002) 采用的“of”在许多高频词串中语音脱落的情况比较明显, 可能在一定程度上影响到实验结果, 因此 Kapatsinski 和 Radicke (2009) 选择了语音脱落较少的“up”, 采用相同的实验范式对包含“up”的词串进行了实验研究。他们发现, 频率对监控任务表现的影响呈现“U”型效应, 即在听觉监控实验中, 高频词串(语块)和低频词串中“up”的监控难度都要显著高于中等频率的词串。此外, 实验还表明, 高频词串的判断难度在所有词串中是最大的。这就说明, 高频短语具有整体加工的特性, 也就是我们所说的语块。基于听觉监控范式的这两项研究共同表明, 语块具有一定的频率性, 高频词串构成的语块是整体存储加工的。

### 2.3 ERP 研究中语块加工引发的电位反应

相比于自定速阅读、眼动技术和词汇监控等范式, ERP 研究无疑是观察语块在线加工最为直接、最为精确的一种手段。研究表明, 语块加工能够引发一系列的相关电位反应, 同样支持语块作为整体存储加工的观点。

Siyanova (2010) 结合眼动阅读和 ERP 技术对二项式短语的加工进行了深入研究。眼动研究发现, 二项式短语的在线加工同样存在着频率效应。高频二项式短语(语块)在阅读时, 其首次注视时间、总阅读时间和注视次数均要显著低于频率较低的控制短语。接下来, Siyanova 进行了三项 ERP 实验以进一步检验高频二项式短语整体加工的假设。ERP 实验发现, 高频二项式短语比低频控制短语引发了波幅显著更大的 P300 反应, 由于该反应与模板匹配相关, 因此这就表明, 被试大脑中形成了对高频二项式短语的表征。此外, 该实验还发现, 在语义联系均比较强的情况下, 低频控制短语(e.g. spoon and fork)比高频二项式短语(e.g. knife and fork)能引发波幅显著更大的 N400 反应。N400 的波幅与语义畸异程度相关, 这也就表明, 高频二项式短语相对于低频控制短语在大脑中具有更强的语义联系, 高频二项式短语具有加工的优势。因此, 眼动研究和 ERP 研究的发现进一步支持了高频二项式短语作为语块整体存储加工、具有加工优势的观点。

此外, Tremblay 和 Baayen (2010) 结合行为实

验和 ERP 技术也考察了 4 词语块加工的整体性。概率等信息能引发包括 P1, N1 和 N2 在内的一系列 ERP 电位反应, 它们会随着注意的分配、词汇频率等因素的不同而发生变化。研究者认为, 由于 N1、P1 和 P2 反应都发生在 200~250 ms 的范围内, 而词汇通达一般发生在 200 ms 以内, 所以, 如果实验引起了上述电位反应, 我们就可以肯定语块是整体存储和提取加工的——因为在这么短的时间内, 大脑不可能提取词 4 个独立的词, 更不可能通过规则计算将它们组合起来。该实验发现了 N1 和 P1 这两个电位的显著变化, 从而印证了 4 词语块整体存储和加工的假设。

## 3 研究展望

语块整体存储加工的心理现实性虽然基本得到了证实, 但仍然还存在着一些尚待解决的问题。

### 3.1 语块的识别和界定

在语言理解和语言产生过程中, 大脑对多种数据信息都会比较敏感。其中, 频率被普遍认为是造就语块加工优势的一个决定性因素(Ellis, 2002; Wray, 2002)。但是, 许多研究也表明, 频率并不能作为判定语块的唯一标准。Schmitt, Grandage 和 Adolphs (2004) 研究发现, 从语料库中抽取出来的高频语块, 在母语者的大脑中并不都具有心理现实性。Simpson-Vlach 和 Ellis (2010) 的研究则发现, 在语块加工方面, 母语者和二语者对不同的数据信息产生敏感——母语者对互信息(MI)较为敏感, 而二语者对频率信息较为敏感。因此几乎可以肯定, 频率并不能作为界定语块的唯一标准。

Thiessen, Kronstein 和 Hufuagle (2013) 提出了一个基于记忆的数据学习模型, 这个模型由提取(extraction)和整合(integration)两个模块构成, 并综合考虑了条件统计量(conditional statistics, 如预测性、概率等强度信息)、分布统计量(distributional statistics, 如频率)和基于线索的统计量(cue-based statistics, 如词汇边界)在数据学习中的作用。在这个模型中, 作者认为语块的提取主要是通过概率和预测性等强度信息实现的, 频率主要是用来对不同的语块进行比较, 从中概括出语言的结构信息。

此外, 计算机自动抽词研究也为我们从语料

库中提取语块作为实验材料提供了借鉴。自动抽词是文本信息处理中的重要课题之一,当前比较通行的做法是通过评估候选字符串内部结合紧密度来判断该串成词与否。罗盛芬和孙茂松(2003)在考察了频率、Z分数、T分数和互信息(MI)等9种常用统计量在汉语自动抽词中的表现之后,发现互信息的抽词能力最强。英文短语和汉语词的自动抽取原理是类似的,Pecina和Schlesinger(2006)列举了语块链接强度的82种计算方法,并尝试结合多种方法构建出二词语块的自动抽取方法。他们的研究表明,不同的语块对不同的统计学参数敏感,语块的自动抽取并不存在一个唯一的最佳方案。

因此,总的来说,除了频率,语块的界定和提取很有可能还涉及到概率、预测性、语义链接强度等不同的数据信息,如何在诸多参数的基础上构建出一个语块的界定标准,则有待于进一步研究。

### 3.2 语块及其成分词的关系

虽然语块是整体存储加工的,但这并不等于说语块不存在内部结构。Bybee(2006)认为,由于经常使用,极高频的那些语块可能会逐渐规约化而失去其内部结构。Weinert(1995)也曾指出,虽然语块可能是整体存储加工的,但语块内部的成分词却有可能是以线性或非线性的不同方式组织在一起的。成分词的表征和语块的表征是相互竞争还是相互促进的呢?Kapatsinski和Radicke(2009)的研究发现,当整个语块的频率足够高时,语块与其构成词之间在表征加工上是一种相互竞争的关系。此外,Tremblay等(2009)也认为,“存储”这一概念的界定还不太清楚,语块作为一个整体,其内部的成分词既有可能是通过组合知识联系在一起,也有可能彼此不分界限,没有内部结构。以上观点也就意味着,语块可能存在着不同的子类,不同子类的语块,其成分词之间的链接强度可能有所差异。由于成分词之间链接强度的不同,在整体表征的同时,有些语块的成分词可能被激活,有些语块的成分词则可能完全不被通达。

上述实验分别从视觉加工和听觉加工,语言理解和语言产生,行为实验和ERP实验等角度对2到4个词构成的语块进行了研究,大部分的实验研究都是基于语块的频率效应进行的,即证明了高频的语块相对于低频的语块具有加工上的显著

优势。大脑中存在着多词语块的心理表征。

大部分的实验都证实了语块加工的频率效应,但这并不直接意味着这些语块都是以整体的方式表征、并且存储为不可分析的一个整体的。Arnon和Snider(2010)的研究发现,短语的频率效应并不仅仅存在于极端高值与极端低值之间,事实上,这种效应存在于频率分布的整个连续统中。前文的实验基本上都是将频率极高的语块和频率较低的控制短语进行对比研究,因此,结合其它多项ERP研究的发现,我们基本上可以肯定,这些频率较高的语块确实是以整体的方式表征的,具有较强的心理现实性。但是,至于频率连续统中的其它语块是否也是以整体的方式表征,这还有待于进一步的研究。

### 3.3 语块的动态发展及加工模型

目前的语块研究中(无论是心理语言学研究还是语言学研究),纵向研究仍比较少。Bannard和Matthews(2008)的研究发现,2~3岁的儿童就已经能够在心理词典中存储词串信息了,但是在这个年龄段之后,语块的获得与发展又会经历一个什么样的过程呢?此外,成人存储在大脑中的语块,会不会因为某种原因(如病理原因)而衰退或丧失呢?

针对儿童的语块发展问题,Wray和Perkins(2000)进行过一项纵向研究,并建立了一个模型。他们认为,随着语言能力的发展,基于语块的整体性加工和基于规则的分析性加工在儿童语言中的比重不断发生变化。总的来说,1~2岁期间,儿童的语言基本上是以整体存储的语块为主;2~8岁期间,儿童的分析性知识不断发展,此时语块在他们的语言产出中比重不断下降;8~18岁期间,整体存储的语块在整个语言活动中的比重开始回升,但整体上还是低于分析性的基于规则和计算产出的语言;18岁之后,语块在语言产出中的比重逐渐稳定下来,并一直保持高位状态。

成人语块的发展研究主要集中在病理语言学方面。van Lancker Sidtis和Postman(2006),van Lancker Sidtis(2012)对左、右脑分别受损的成人被试进行了研究,结果发现,右脑受损的被试在自然产出的语言中,语块的比重明显要低于大脑功能正常的被试。在对一系列失语症病人进行研究的基础上,van Lancker Sidtis(2012)提出了语块的“双加工模型”(Dual-process Model)。“双加工模

型”认为, 语言加工包含了语块模式和语法模式这两个截然不同的模式, 两者不断作用, 相互影响。语块模式以语块为单位, 以整体的方式进行, 因而无需经过语法规则的作用; 语法模式则是以单个的词为单位, 利用计算的方式对语言进行加工处理。另外, 该研究还发现, 语块理解和语块产出涉及到不同的脑区——语块理解由左、右脑同时承担, 而语块产出则主要依赖于右脑的皮层下通路进行。

总而言之, 从儿童语言获得与发展的角度来说, 语块的动态发展过程仍然需要更多的心理学研究参与进来, 揭示其中的规律, 建立若干模型。从成人语块研究的角度来说, 病理语言学的研究以及脑功能正常被试的研究都还有待于进一步深入, 语块在成人脑海中的心理表征是一个动态的过程, 这个过程仍需得到更进一步的揭示。

### 3.4 二语者语块的心理现实性

现有研究基本上肯定了语块在母语者大脑中存在心理表征, 并且以整体的方式存储加工。但是, 语块是否在二语者大脑中具有心理现实性, 现有的研究并未得出一致的结论。Conklin 和 Schmitt (2008) 的研究发现, 高水平的二语者(达到硕士入学要求的英语水平, 雅思、托福成绩合格者)与母语者一样, 在阅读高频习语时也具有整体加工优势。此外, Jiang 和 Nekrasova (2007) 利用语法判断任务, 发现二语者被试在判断语块是否合乎语法时也能获得整体加工的优势, 因而同样支持语块在高水平的二语者大脑中具有心理现实性。但是另一方面, 不少研究(Underwood et al., 2004; Schmitt & Underwood, 2004; Siyanova et al., 2011)也并未发现二语者在加工目标语语块时具有加工优势, 加上专门针对二语者语块加工的研究甚少, 所以在语言习得领域, 二语者语块的心理现实性问题仍有待于进一步研究。

### 参考文献

李兴珊, 刘萍萍, 马国杰. (2011). 中文阅读中词切分的认知机理述评. *心理科学进展*, 19, 459-470.

罗盛芬, 孙茂松. (2003). 基于字串内部结合紧密度的汉语自动抽词实验研究. *中文信息学报*, 17(3), 9-14.

Arcara, G., Lacaïta, G., Mattaloni, E., Passarini, L., Mondini, S., Benincà, P., & Semenza, C. (2012). Is “hit and run” a single word? The processing of irreversible binomials in neglect dyslexia. *Frontiers in Psychology*, 3, 11.

Arnon, I. (2009). *Starting Big-The role of multi-word phrases in language learning and use*. Unpublished doctoral dissertation, Stanford University.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62(1), 67-82.

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, 19(3), 241-248.

Biber, D., Johnsson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written english*. London: Longman Pearson Education Limited.

Bobrow, S. A., & Bell, S. M. (1973). On catching on to idiomatic expressions. *Memory and Cognition*, 1, 343-346.

Bybee, J. L. (2006). From usage to grammar: The mind's response to repetition. *Language*, 82, 711-733.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29, 72-89.

Conklin, K., & Schmitt, N. (2012). The Processing of Formulaic Language. *Annual Review of Applied Linguistics*, 32, 45-61.

De Bot, K. (1992). A bilingual production model: Levelt's speaking model adapted. *Applied Linguistics*, 13(1), 1-24.

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, 24, 143-188.

Ellis, N. C., & Sinclair, S. G. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology*, 49, 234-250.

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, 7, 195-225.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, 20, 29-62.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, 12, 627-635.

Howes, D. H., & Solomon, R. L. (1951). Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology*, 41(6), 401-410.

Jiang, N., & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal*, 91, 433-445.

Jurafsky, D. (2002). Probabilistic modelling in psycholinguistics: Linguistic comprehension and production. In R. Bod, J. Hay, & S. Jandry (Eds.), *Probabilistic linguistics* (pp. 1-50). Cambridge, MA: MIT Press.

Kapatsinski, V., & Radicke, J. (2009). Frequency and the

- emergence of prefabs: Evidence from monitoring. In R. Corrigan, E. A. Moravcsik, H. Ouali, & K. M. Wheatley (Eds.), *Formulaic language: Acquisition, loss, psychological reality, and functional explanations* (pp. 499–520). Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97.
- Morris, R. K. (2006). Lexical processing and sentence context effects. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 377–401). San Diego: Academic Press.
- Nattinger, J. R., & Decarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Pawley, A. (2007). Developments in the study of formulaic language since 1970: A personal view. In P. Skandera (Ed.), *Phraseology and culture in English* (pp. 3–45). Berlin, New York: Mouton de Gruyter.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191–226). London: Longman.
- Pecina, P., & Schlesinger, P. (2006). *Combining association measures for collocation extraction*. Paper presented at the Proceedings of the COLING/ACL on Main conference poster sessions, Sidney, Australia.
- Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. L. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure* (pp. 137–157). Amsterdam: John Benjamins.
- Pinker, S. (1991). Rules of language. *Science*, 253(5019), 530–535.
- Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, 6(11), 456–463.
- Prior, A., & Bentin, S. (2007). Word associations are formed incidentally during sentential semantic integration. *Acta Psychologica*, 121(7), 57–71.
- Rayner, K., & Pollatsek, A. (1994). *The psychology of reading*. London: Routledge.
- Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 127–151). Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Schmitt, N., & Underwood, G. (2004). Exploring the processing of formulaic sequences through a self-paced reading task. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 173–189). Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Shaoul, C., & Westbury, C. (2011). Formulaic sequences: Do they exist and do they matter? In G. Jarema & G. Libben (Eds.), *The mental lexicon* (vol. 1, pp. 171–196). Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31, 487–512.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Siyanova, A. (2010). *On-line processing of multi-word sequences in a first and second language: Evidence from eye-tracking and ERP*. Unpublished doctoral dissertation, University of Nottingham.
- Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, 27, 251–272.
- Sosa, A. V., & MacFarlane, J. (2002). Evidence for frequency-based constituents in the mental lexicon: Collocations involving the word “of”. *Brain and Language*, 83, 227–236.
- Swinney, D. A., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior*, 18, 523–534.
- Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin*, 139(4), 792–814.
- Tremblay, A., & Baayen, H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 151–173). London: The Continuum International Publishing Group.
- Tremblay, A., Derwing, B., & Libben, G. (2009). Are lexical bundles stored and processed as single units? *Working Papers of the Linguistics Circle of the University of Victoria*, 19, 258–279.
- Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 153–172). Amsterdam, Philadelphia: John Benjamins Publishing Company.
- van Lancker Sidtis, D. (2012). Two-track mind: Formulaic and novel language support a dual-process model. In M. Faust (Ed.), *The Handbook of the Neuropsychology of Language* (pp. 342–367). Oxford: Blackwell.
- van Lancker Sidtis, D., & Postman, W. A. (2006). Formulaic

- expressions in spontaneous speech of left- and right-hemisphere damaged subjects. *Aphasiology*, 20(5), 411–426.
- Weinert, R. (1995). The role of formulaic language in second language acquisition: A review. *Applied Linguistics*, 16, 180–205.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language & Communication*, 20, 1–28.

## The Psychological Reality of Formulaic Language

YI Wei; LU Shiyi

(School of Chinese as a Second Language, Beijing University, Beijing, 100871, China)

**Abstract:** Formulaic language is composed of multi-word sequences that are form-meaning pairings stored in the mind of language users as a result of long-term language contact and extensive language experience. Psychological experiments focused on reading and auditory processing provide support for the existence of formulaic language; these studies show that formulaic language is stored and processed as a whole in our mental lexicon in the same way that words are. Future research should involve the following topics related to formulaic language: the identification of formulaic language, the relationship between formulaic language and the component words, and finally, the dynamic development of formulaic language, second language speakers and their formulaic language.

**Key words:** formulaic language; psychological reality; cognitive processing