

基于复杂网络分析的晚期汉英双语者心理词库结构探究

陈瑶 周榕

摘要：本研究结合语义流畅任务和复杂网络分析，探索晚期汉英双语者的双语词库结构，通过构建词汇关联网络、计算网络参数（网络节点数、边数、平均度、平均聚类系数、平均最短路径、小世界属性参数和模块化参数）和生成网络可视化图，考察词库结构的特征与差异。结果发现：①双语词库具有小世界属性和模块化属性；②L1 词汇网络比 L2 词汇网络结构更为优化，这与被试的 L2 使用频次和习得年龄相关；③词汇网络节点的组织方式在一定程度上支持激活扩散模型。本文据此为中国英语词汇教学提出相关建议：顺应学习者心理词库结构的发展趋势，采用 L2 词汇与图片或实物直接相连的教学方法，以促进 L2 词库结构的动态发展。

关键词：复杂网络分析；语义流畅任务；心理词库；晚期汉英双语者；双语词库

[中图分类号] H04/H3

DOI: 10.12002/j.bisu.504

[文献标识码] A

[文章编号] 1003-6539 (2024) 01-0142-19

引言

语言系统是一种复杂的网络结构，采用复杂网络分析来研究语言现象可使语言研究更为精实（陈芯莹、刘海涛，2011；刘海涛，2011；Chen & Zhou, 2022）。复杂网络分析是一种跨学科研究方法（陈芯莹，2015），研究者基于真实语料考察语言系统的网络特征，从而增进对语言结构的了解（刘海涛，2011）。在国外，运用复杂网络分析探究语言现象的研究已经相当丰富，涉及语音学（如 Vitevitch, 2008）、语义学（如 Steyvers & Tenenbaum, 2005）和句法学（如 Ferrer-i-Cancho et al., 2004）等领域。国内的复杂网络分析在语言研究中的应用还主要停留在句法层面（高松等，2010；刘海涛，2010；陈芯莹、刘海涛，2011），涉及其他层面的研究仍不多见；相关研究主要以语料库文本

[基金项目] 本研究受 2020 年广东省广州市哲学社会科学发展“十三五”规划一般课题“穗港澳青年幽默话语识解机制及和谐关系建构研究”（项目编号：2020GZYB30）资助。

为研究数据,因此只能探究语言表面结构,而不能探索语言背后的心理机制。若能以心理学范式为复杂网络分析提供数据来源,将会提升复杂网络分析在语言研究中的丰富度。

心理词库被喻为网络系统(Castro & Siew, 2020),在复杂网络技术发展起来后,许多研究(如Wilks & Meara, 2002; Steyvers & Tenenbaum, 2005; Vitevitch, 2008; Borodkin et al., 2016)都结合该研究手段和相应的心理学范式(如语义流畅任务)探究词库的结构特征。本研究也沿用该思路,通过语义流畅任务收集被试的词汇产出,并借助复杂网络分析进行数据统计,尝试将网络科学运用在实证数据的统计中,由此探究晚期汉英双语者双语词库的结构特征及结构差异,以期能为外语词汇教学提供些许启示。

一、研究背景

心理词库常被认为是以词汇为节点、以词汇间的关联程度为边线的网络系统(Castro & Siew, 2020)。基于此,Collins & Loftus(1975)提出了激活扩散模型,认为节点间根据词汇间的语义关联程度相互联结。然而,Bock & Levelt(1994)认为,词汇节点并不只以语义相似度进行联结,还会以语音和句法-词汇形态层面的相似度进行扩散式的联结分布。双语词库结构领域的研究者在此基础上讨论了两种语言词汇网络结构的异同(熊永红等,2011;孙继平、孙秀丽,2014;徐歌,2016)。

双语词库结构表征较为复杂,研究者难以用精准的仪器进行测量,因此,语义流畅任务常被应用于双语词库结构的探索(Antón-Méndez & Gollan, 2010)。语义流畅任务要求被试在有限的时间内就某个语义类别多而快地产出词汇,然后主试按照聚类 and 转化两个维度对词汇产出进行归类和数据分析,这种分析方式使得该任务成为探究双语词库结构的一扇窗户(Roberts & Le Dorze, 1997; Rosselli et al., 2002)。然而,该统计方式容易受到主试主观因素的影响,导致结果的客观性受到质疑。为弥补此不足,相关领域的研究(如Wilks & Meara, 2002; Kenett et al., 2011; Borodkin et al., 2016)多使用复杂网络分析来替代传统分析方式,这是因为:其一,复杂网络分析与传统分析方式得出的结果相似(Borodkin et al., 2016);其二,复杂网络分析囊括了数学和统计物理学的多种算法,统计过程不涉及人工分析带来的偏差,从而使结果更为可靠。

复杂网络分析是一种基于图论的研究方法,它可以为认知结构(如词汇网络)的模拟提供多种缜密算法(Siew et al., 2019)。复杂网络分析将系统视为

网络结构,网络内的节点为独立个体,边线长短代表节点间关系的疏密程度。通过该研究方法,研究者可以从微观、中观和宏观3个层面探索系统的结构特征(Siew et al., 2019)。

在微观层面,研究者可以通过网络可视化图呈现的节点组织方式来考察网络的微观特征。

在中观层面,研究者可以通过模块化参数(Newman, 2006)对节点的聚类分布和聚类程度进行考察。模块化参数主要观察网络可被分为多个子网络的程度,当网络的模块化参数大于0.3时,意味着网络具有模块化属性(Newman, 2006; Chen et al., 2018)。具有模块化属性的网络更倾向聚类成不同的子网络,进而实现子网络间的协调配合。模块化参数越大,意味着网络的模块化程度越高,网络就越倾向聚类为多个子网络,从而实现节点间更程度的分工合作。

在宏观层面,研究者可以通过网络参数观察整体结构特征。例如,可以由网络节点数和边线数来观察网络的规模大小,也可以由平均度观察节点互为联结的紧密程度。平均度指网络中节点平均联结的度量,度表示一个节点与其他节点之间的联结数,而平均度则是所有节点的度的平均值(陈芯莹, 2015)。平均度越高,节点间的联结就越紧密。此外,研究者还可以通过对比实证网络^①和随机网络^②的平均聚类系数和平均最短路径来观察该系统是否具有小世界属性(Watts & Strogatz, 1998)。相较于随机网络,具有小世界属性的网络平均聚类系数更高、平均最短路径数值更小(Watts & Strogatz, 1998; Hills et al., 2010; Siew et al., 2019),这种网络内的节点联结紧密,节点间的信息可以高效地联结通达,是一种较为优化的结构(Bullmore & Sporns, 2012; Siew et al., 2019)。平均聚类系数是指网络内随意的两个节点互为相邻的概率,平均最短路径是指网络内随意的两个节点间最短路径的平均值(Siew et al., 2019)。此外,本研究还通过补充小世界参数(Humphries & Gurney, 2008)来考察实证网络的小世界属性。小世界参数是指平均聚类系数和平均最短路径的均衡值,当数值大于1时,表示网络具有小世界属性(Humphries & Gurney, 2008; Borodkin et al., 2016)。

目前,复杂网络分析在心理词库研究中的应用集中于对单语者词库的探索,如语音网络(Vitevitch, 2008; Kello & Beltz, 2009; Arbesman et al., 2010)、语义网络(Steyvers & Tenenbaum, 2005)、正字法网络(Kello & Beltz, 2009; Siew

① 实证网络指由实验数据构建而成的网络。

② 随机网络指与实验数据样本大小一致,但由随机算法构建而成的网络。

et al., 2019) 等, 仅有少数研究 (Wilks & Meara, 2002; Borodkin et al., 2016) 将其运用到对双语词库的探究上。这些研究发现, 单语者和双语者词库均具有小世界属性, 这表明小世界属性是心理词库结构的共性 (Wilks & Meara, 2002; Steyvers & Tenenbaum, 2005; De Deyne & Storms, 2008; Borge-Holthoefer & Arenas, 2010; Kenett et al., 2011; Borodkin et al., 2016)。此外, Wilks & Meara (2002) 通过对比双语词库的网络密度^①, 发现 L1 (母语) 词库比 L2 (第二语言) 词库节点联结更为紧密, 由此他们认为 L1 词库结构更为复杂、优化。然而, 这项研究仅参考了网络密度, 没有考察其他网络属性, 这在一定程度上降低了结果的可靠性 (Borodkin et al., 2016)。为了进一步探索双语词库结构, Borodkin et al. (2016) 基于 Wilks & Meara (2002) 的研究思路从小世界属性和模块化属性角度观察双语词库结构, 丰富了对网络属性的探索, 结果发现: 相较于 L1 词库, L2 词库模块化的程度更低, 节点间的联结更为紧密。这表明, L2 词库内的节点尽管联结更为紧密, 但由于节点聚类程度较低, 语义信息的激活仅局限于某个子网络而无法实现子网络间的有效协作, 阻碍了语义信息的高效通达, 因此相较而言, L1 词库更为优化。上述研究 (Wilks & Meara, 2002; Borodkin et al., 2016) 虽在节点联结强度上的结果有所不同, 但一致发现 L1 词库比 L2 词库结构更为优化。研究者将此结果归因于双语者较低的 L2 使用频次和较晚的 L2 习得年龄。

然而, 现有研究多以双语均为印欧语系语言的双语者为研究对象, 鲜有研究关注以汉藏语系语言为 L1、以印欧语系语言为 L2 的双语者 (如汉英双语者), 前者双语语言距离较近, 而后者双语距离较远, 后者的双语词库是否会呈现出有别于前者的结构特征, 我们尚未得知。此外, 目前仍鲜有研究从复杂网络分析的角度探索以汉藏语系语言为 L1 的双语者心理词库结构。因此, 为了增进对此类双语者词库的了解, 本研究以晚期汉英双语者为研究对象, 探索此类双语者的词库结构, 并考察语言距离对词汇网络结构的影响。在我国, 汉英双语者多在出生后首先习得汉语, 六七岁之后 (母语成为主导语言后) 才开始学习英语, 因此他们被称为晚期双语者 (Fabbro, 1999)。

基于前人的发现及本文的研究目的, 笔者提出以下研究问题: ①晚期汉英双语者的双语词库结构有何特征? 双语词库是否具有小世界属性? 词库内节点以何种方式进行联结? ②晚期汉英双语者双语词库结构有何差异? L1 词库是否比 L2 词库优化? 为了解答上述问题, 本研究沿用 Borodkin et al. (2016) 中的

① 网络密度指节点间互为联结的紧密程度, 数值越大则紧密程度越高。由于该数值与平均度所考察的网络特征有重合之处, 因此本研究仅观察平均度。

研究思路,结合语义流畅任务和复杂网络分析来探索晚期双语者词库的结构特征和结构差异。研究结果可验证网络科学在心理词库探究中的有效性,促进对双语词库结构的认识,并根据词库特征提出改进二语词汇教学模式的相关建议。

二、实验方法

1. 研究思路与实验设计

本研究的实验范式是语义流畅任务,即要求被试用不同语言(汉语或英语)在相继的1分钟内产出不同语义类别(“动物”或“蔬菜水果”)下的词汇。选用这两个语义类别是因为:其一,它们是语义流畅任务中最常用的语义类别(Acevedo et al., 2000);其二,它们在不同语言文化中有着较高的清晰度和稳定的统一性(Ardila et al., 2006);而蔬菜和水果合二为一是为了避免这两个类别在植物学定义和日常使用中的歧义(Kavé, 2005; Borodkin et al., 2016)。本研究为了避免前1分钟语言和语义类别顺序对后1分钟词汇结果的影响,对被试词汇产出的语言类别和语义类别顺序进行了拉丁方平衡。

2. 被试

广东省某高校32名晚期汉英双语者(其中包括29名女生、3名男生)参与了本次实验,他们均是英语专业一年级研究生,年龄相近(介于22—26岁之间),都通过了英语专业八级考试,L2水平较高,能较为熟练地使用汉英两种语言。

实验前,被试需完成二语词汇量测试VST-A(Nation & Beglar, 2007),为了避免疲劳效应,笔者将题目按比例随机减半到70道小题,统计被试答对的题目数量并乘以200,得出被试最终词汇量结果。实验为拉丁方设计,在L2词汇产出部分,将被试随机分为两组(一组产出“动物”类别的词汇,另一组产出“蔬菜水果”类别的词汇),因此,需要对比这两组被试L2词汇量大小,以确定他们L2词汇产出的结果是否具有可比性。结果显示,两组被试的英文词汇量大小无显著差异($t=0.476$, $p=0.641$)。

3. 实验程序

正式实验前,主试引导被试完成1组练习,要求被试在1分钟内快而多地用1种语言(汉语或英语)产出属于“家具”或“家人”类别的词。被试在词汇产出期间不可产出专有名词,亦不可产出以不同词素结尾的同一单词(如复数形式等)。被试在完成练习并经确认熟悉流程后,进入正式实验。

在正式实验中,主试引导被试先后用不同语言在1分钟内产出属于“动物”

或“蔬菜水果”类别的词汇。被试的所有词汇产出均由主试用纸笔以及录音机记录,待被试完成任务后,主试当场向被试确认其词汇产出的情况,以防数据错漏。

4. 数据处理与统计

实验结束后,主试将收集到的 4 组数据分别整理为英文-动物组、英文-蔬果组、汉语-动物组和汉语-蔬果组。排除被试不正确的及重复出现的词汇后,将其转成 4 组发生矩阵,每组矩阵的第 1 行是所有被试产出的词汇,第 1 列是被试的序号,列与行交汇的每个单元格代表该被试词汇产出的情况,若该被试产出某个单词,则在该单词的单元格上标“1”,无则标“0”。随后使用 R 语言软件(3.5.1 版本)来构建实证网络和随机网络、处理网络、计算网络参数并生成网络可视化图。

(1) 词汇关联网络的构建与网络的处理

本研究的网络节点为被试的词汇产出,边线权重是词与词之间的关联系数(即产出词汇 a 会连带产出词汇 b 的概率),关联系数体现词与词之间的相互作用(Kenett et al., 2013; Borodkin et al., 2016),边线权重的计算采用皮尔逊相关系数算法,将发生矩阵转化为对应的相关系数矩阵(Kenett et al., 2013)。

转换后的相关系数矩阵被视为邻接矩阵,由此构建 4 个有权重无向的词汇关联网络(英语-动物词汇网络、英语-蔬果词汇网络、汉语-动物词汇网络和汉语-蔬果词汇网络)。

然而,词汇间相关系数值通常较低,这样会导致网络局部拓扑关系和结构特征的疏漏。为弥补此缺陷,本研究采用基于层次聚类算法的最小生成树算法来对网络进行处理(Kenett et al., 2013)。

(2) 网络参数的计算

本研究计算节点数、边数、平均聚类系数、平均最短路径、小世界属性参数和模块化参数,详见表 1;并生成网络整体和局部可视化图,见下文图 1—图 4,其中图 1 和图 2 分别为英语动物、蔬菜水果词汇网络图,图 3 和图 4 分别为汉语动物、蔬菜水果词汇网络图。此外,各图中 a 为整体图,b 为最小生成树算法处理后的局部图。

(3) 随机网络的生成与随机网络参数的计算

为了验证网络的可靠性,本研究构建 Erdős-Rényi 随机网络(Erdős & Rényi, 1960)来证明实证网络参数与随机网络参数在统计上是否存在显著差异。随机网络由与实证网络相同数量的节点和 1 个固定概率(0.5)构建而成,固定概率指随意的两个节点互相联结的概率。本研究对此进行了 1000 次随机模

拟, 取该 1000 个图的参数均值作为随机网络的参数数值, 并将随机网络与实证网络的参数进行独立样本 t 检验, 以比较两者的差异性。

三、结果与分析

1. 网络参数对比分析

词汇网络参数数值如表 1 所示, 4 个词汇网络中的聚类系数均大于随机网络的相应参数, 平均最短路径也比随机网络的小, 并且两者的数值差异不大, 小世界属性参数也大于 1, 因此, 4 个词汇网络均具有小世界属性。

此外, 词汇网络的模块化参数大于 0.3, 表明词汇网络均具有模块化属性 (Chen et al., 2018)。同时, L1 词汇网络的模块化参数大于 L2 词汇网络 (动物词汇网络: $Q_{\text{汉语}}=0.701>Q_{\text{英语}}=0.696$; 蔬果词汇网络: $Q_{\text{汉语}}=0.713>Q_{\text{英语}}=0.690$), 这表明 L1 词汇网络的模块化程度更高。

另外, 由表 1 可知, 被试 L1 词汇网络的节点、边线数量比 L2 词汇网络的多 (动物词汇网络: $N_{\text{汉语}}=45>N_{\text{英语}}=26$, $E_{\text{汉语}}=947>E_{\text{英语}}=293$; 蔬果词汇网络: $N_{\text{汉语}}=42>N_{\text{英语}}=26$, $E_{\text{汉语}}=836>E_{\text{英语}}=323$), 这表明 L1 词汇网络的规模更大。被试 L1 词汇网络的平均度比 L2 词汇网络的大 (动物词汇网络: $K_{\text{汉语}}=42.098>K_{\text{英语}}=22.538$; 蔬果词汇网络: $K_{\text{汉语}}=39.810>K_{\text{英语}}=24.846$), 这表明 L1 词汇网络内节点相连的紧密程度高于 L2 词汇网络。

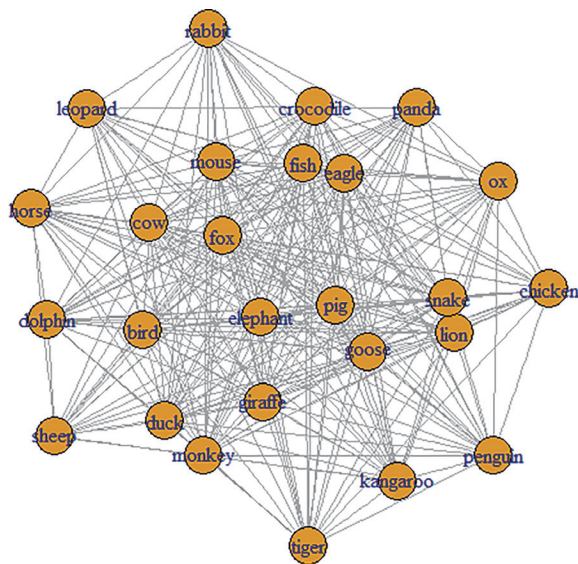
总之, 相较于 L2 词汇网络, 双语者 L1 词汇网络中节点联结更为紧密, 模块化程度更高, 更倾向聚类为不同的子网络, 这使得词汇信息的传递更为高效, 意味着 L1 词汇网络的结构更为优化。

表 1 4 个网络的主要参数情况

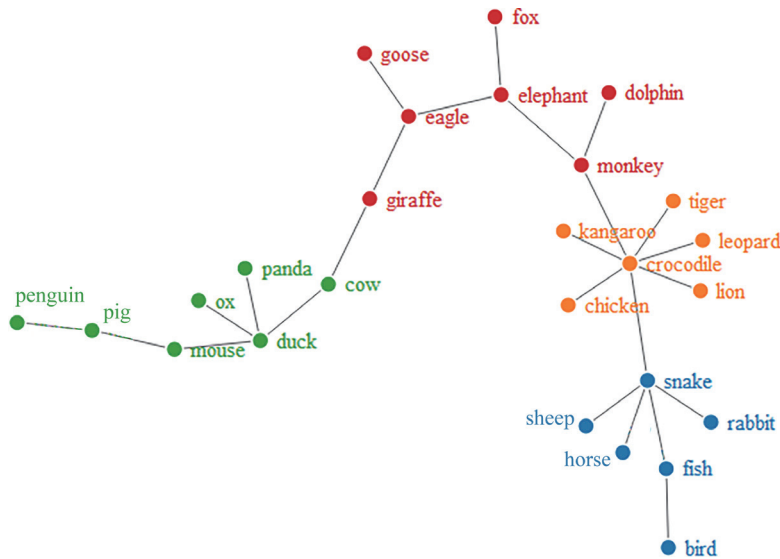
网络类型	N	E	K	CC	CC_{ran}	$ASPL$	$ASPL_{\text{ran}}$	S	Q
英语-动物	26	293	22.538	0.917	0.867	1.098	1.361	1.310	0.696
英语-蔬果	26	323	24.846	0.994	0.956	1.006	1.334	1.380	0.690
汉语-动物	45	947	42.098	0.963	0.935	1.043	1.364	1.345	0.701
汉语-蔬果	42	836	39.810	0.974	0.948	1.029	1.358	1.356	0.713

注: 表 1 中 N 为网络节点数, E 为边数, K 为平均度, CC 为网络平均聚类系数, CC_{ran} 为随机网络的平均聚类系数, $ASPL$ 为网络平均最短路径, $ASPL_{\text{ran}}$ 为随机网络平均最短路径, S 为小世界属性值, Q 为模块化参数。

除了网络参数，本研究还获得了网络可视化图，以此宏观地观察词汇网络结构特征。通过对比双语词汇网络的整体图（图 1a 和图 3a 对比，图 2a 和图 4a 对比）可知，双语者 L1 词汇网络比 L2 词汇网络的规模大（节点和边数更多），节点分布也更为密集。此外，从网络的局部图可以看出，节点的组织模式虽分布较为散乱，但在一定程度上以语义、语音及句法-词汇形态层面聚类，该聚类模式将在本文的讨论部分进一步加以说明。

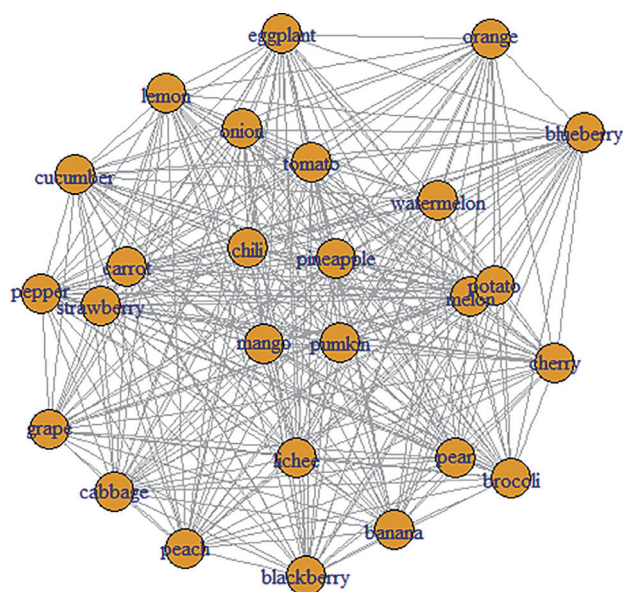


a. 英语-动物词汇网络整体图

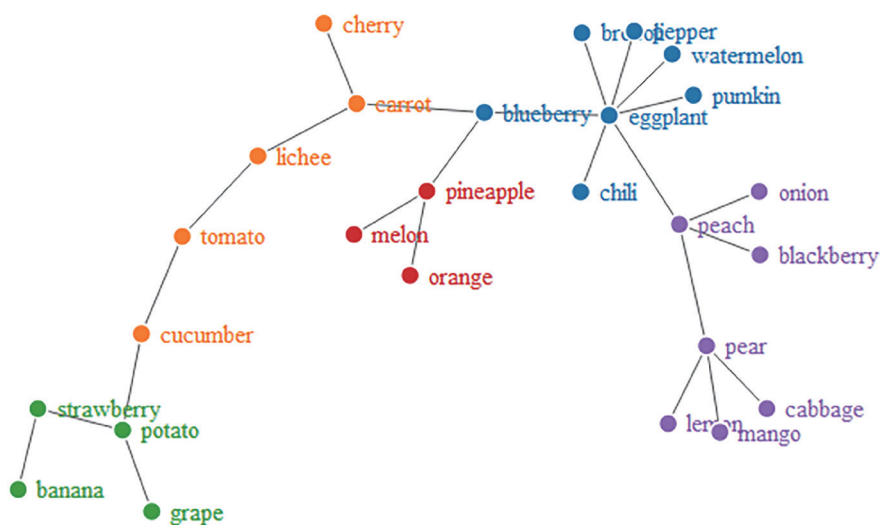


b. 英语-动物词汇网络局部图

图 1 英语-动物词汇网络的整体图与局部图

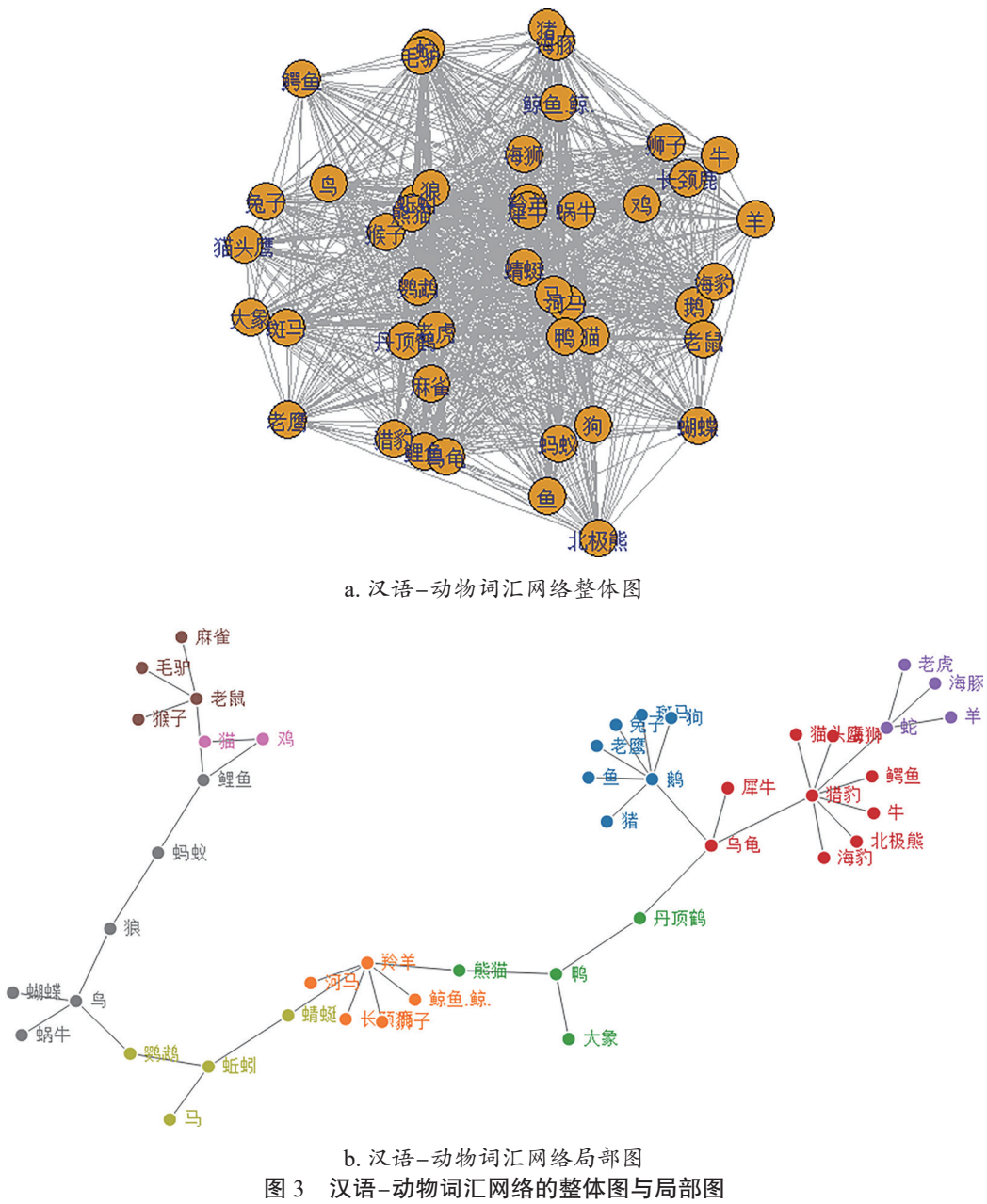


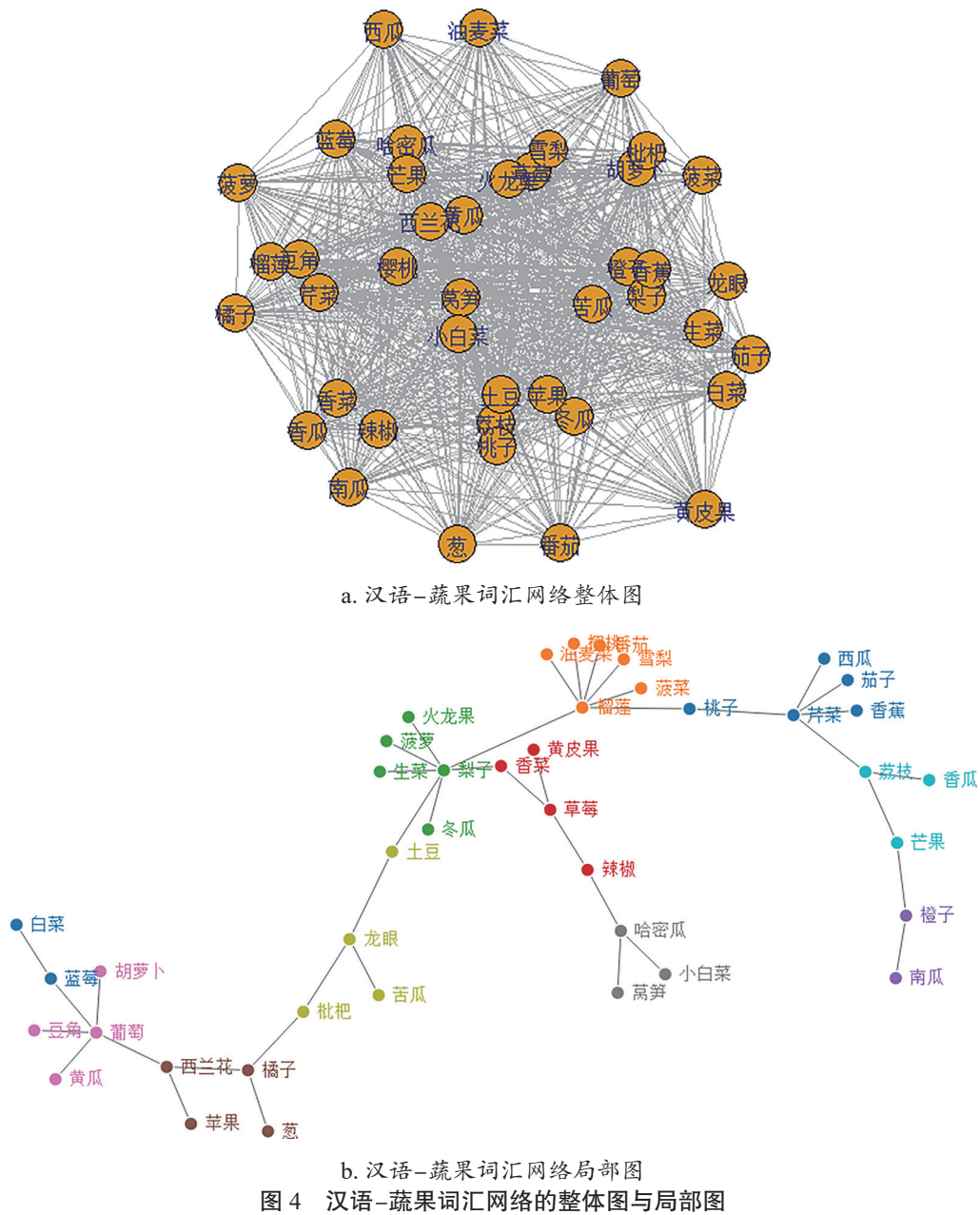
a. 英语-蔬果词汇网络整体图



b. 英语-蔬果词汇网络局部图

图2 英语-蔬果词汇网络的整体图与局部图





2. 网络验证

本研究采用 Erdős-Rényi 随机网络模拟方法 (Erdős & Rényi, 1960), 验证实证网络结果是否为有趋向性的而非随机产生的网络数据 (Borodkin et al., 2016)。随机网络参数数值如表 2 所示。由表 2 中的数据可知, 随机网络与实证网络参数间存在显著差异 (平均聚类系数: $t=-28.392$, $p<0.001$; 平均最短

路径: $t=23.311$, $p<0.001$; 模块化参数: $t=-87.233$, $p<0.001$; 平均度: $t=-2.589$, $p<0.05$)。

表 2 Erdős-Rényi 随机网络参数表

网络类型	平均聚类系数	平均最短路径	模块化参数	平均度
英语-动物	0.497	1.500	0.096	12.538
英语-蔬果	0.497	1.502	0.096	11.923
汉语-动物	0.499	1.499	0.078	23.111
汉语-蔬果	0.500	1.499	0.079	21.476

四、讨论

本研究采用语义流畅任务和复杂网络分析来探究晚期汉英双语者的双语词库结构。对应研究问题,下文将从双语词库的结构特征和结构差异这两个方面对研究结果进行讨论。

1. 晚期汉英双语者双语词库的结构特征

从宏观层面的平均聚类系数、平均最短路径和小世界参数可知,晚期汉英双语者的双语词汇网络均具有小世界属性,这意味着汉英词库内节点间的联结程度紧密,节点间的距离较短,词汇语义信息均能较为高效地联结通达,表明两种语言的词汇网络均为优化的网络结构(Siew et al., 2019)。这与已有研究结果相符(Wilks & Meara, 2002; Borge-Holthoefer & Arenas, 2010; Kenett et al., 2011; Borodkin et al., 2016),进一步验证了小世界属性是心理词库结构的共性。同时,该结果也丰富了前人对双语词库研究的发现(Wilks & Meara, 2002; Borodkin et al., 2016),即对于双语语言距离较近和语言距离较远的双语者而言,双语词库均具有小世界属性,这表明语言距离并不会对词库的小世界属性产生影响。

从中观层面的模块化参数可知,双语词库具有模块化属性,这表明晚期汉英双语者的双语词库均可根据节点间的关联程度进行聚类分布,从而实现网络有效的分工协作。

从微观层面的网络可视化图可知,双语词汇网络内的节点主要以语义为线索聚类联结。例如,在英语-动物词汇网络(如图 1b)中,节点 tiger、crocodile、leopard 和 lion 相连,这是因为这 4 种动物均为凶猛的食肉类动物,

且均为野生动物，因此，被试倾向将四者同时联想。当被试联想到这些动物的食肉性质时，便会进而联想起处于食物链较低端的动物 chicken；而当被试联想到这些动物的野生特性时，便会联想起同属野生动物类别的 kangaroo。又如，在英语-蔬果词汇网络（如图 2b）中节点 cherry、carrot、lichee、tomato 和 cucumber 相联，这是因为前 4 种蔬果的表皮均为红色。而当被试联想到节点 tomato 时，便会激活该节点属于蔬菜类别的属性，进而提取出同为常见蔬菜的节点 cucumber。汉语词汇网络结构也是如此，在汉语-动物词汇网络中（如图 3b），由于词汇共享昆虫类别属性，节点“蝴蝶”与“蜗牛”相连，节点“蚯蚓”与“蜻蜓”相连。此外，词汇节点除了以语义为线索相连，还会以语音相似度为线索互为联结。例如，在汉语-动物词汇网络（如图 3b）中，由于共享语音“bào（豹）”，节点“海豹”和“猎豹”相连。另外，句法-词汇形态层面信息也会影响节点的聚类模式。例如，在英语-动物词汇网络（如图 1b）中，由于节点共享首字母 p，penguin 和 pig 相连。汉语词汇网络也是如此（如图 4b），偏旁一致的词汇节点会互为聚类，如共享偏旁“艹”的“苹果”会与“葱”“葡萄”相连。综上所述，晚期汉英双语者双语词库内的词汇节点主要以语义、语音和词汇形态等因素相互联结，而激活扩散模型也认为心理词库内的词汇以概念层（语义）、语音层和句法-词汇形态层（词汇形态或句法特征）为线索互为相连，因此，晚期汉英双语者双语词库词汇组织方式支持激活扩散模型的假设。

2. 晚期汉英双语者双语词库的结构差异

从宏观层面的网络参数可知，L1 词汇网络节点的联结更为紧密，表明 L1 词库的复杂度更高，这与 Wilks & Meara（2002）的结果相符，与 Borodkin et al.（2016）的结果不一致。后者研究发现，L2 词库内节点的联结更为紧密。这种结果上的差异可能是由不同实验被试的 L2 使用频次不同导致的：Borodkin et al.（2016）研究中的被试虽和本研究的被试一样，均在习得 L1 后才开始学习 L2，但 Borodkin et al.（2016）的被试是在 L2 环境下学习和生活的，L2 在日常生活中的使用频次较高，而本研究的被试是在 L1 环境下学习和生活的，仅在特殊场合（如语言课堂）使用 L2。这就造成本研究被试的 L1 使用频次远高于 L2，随着被试与 L1 的接触日益增多，他们词库内的 L1 词汇会随着汉语文化的发展不断进行概念层面的细分和结构层面的拓展，从而导致 L1 词汇网络的节点联结更为紧密。同时，该结果也表明语言使用频次比语言习得年龄更能对词汇联结的紧密程度产生影响。

从中观层面的模块化参数可知，L1 词库的模块化程度更高，这与 Borodkin et al.（2016）的发现相符，意味着与 L2 词汇网络相比，L1 词汇网络内的节点

更倾向聚类成不同的子网络,由此关联程度高的节点便会聚类分布,子网络内的节点可更为高效地传递信息,同时也方便单个子网络与其他子网络间的信息交流,从而实现子网络间的分工协作和语义信息的快速通达,更有利于网络的动态发展。而 L2 词汇网络的发展则较为滞后,这可能是因为中国英语课堂倾向以 L1 作为 L2 词汇教学的主要媒介,从而导致双语者在 L2 习得初期借助 L1 的词汇网络来构建相应的 L2 词汇网络。但随着被试 L2 水平的提升以及他们对 L2 文化的进一步了解,原先在 L1 词汇网络结构中聚类的节点会作出相应的重组,使得网络内的节点聚类更符合 L2 文化背景下的概念表征。由于本研究中的被试在日常生活中缺乏与 L2 的接触,L2 使用频次较低,因此,他们的 L2 词汇结构仍处于重组的过程中,网络内节点会为了概念重组而打破原先网络中的聚类群落,逐步过渡到与 L2 文化因素相吻合的聚类部落,这导致网络内节点的分布较为零散,模块化程度较低。此外,L2 词汇网络的模块化程度不如 L1 词汇网络的另一个原因可能是 L1 是知识体系构建的主要工具,因此 L1 词汇网络可以随着被试知识体系的扩展而不断发展;而 L2 较少参与知识系统的拓展,这导致其发展较为滞后。

从上述网络参数和微观层面的可视化图可以发现,晚期双语者 L1 词汇网络的规模更大,节点联结程度和模块化程度更高,因此 L1 词库更为优化,这一结果与 Wilks & Meara (2002) 和 Borodkin et al. (2016) 的发现一致。Borodkin et al. (2016) 将 L1 词库结构更为优化的结果归因于语言习得年龄和语言使用频次,本文也基于这两个因素对研究结果进行了阐释。如上文所述,本研究的被试学习 L2 的年龄晚于 L1,同时使用频次也低于 L1,因此,尽管被试是较为熟练的双语者,但他们的 L2 水平还是远不如 L1 水平;被试 L1 词汇网络的构建不仅早于 L2,而且还得到了更为充足的语言输入,因而促进了 L1 词汇网络的动态发展和优化。此外,本文认为,L1 词汇结构更为优化的结果还可由非选择性激活假说加以解释。根据非选择性激活假说 (Kroll & Tokowicz, 2005),被试在词汇产出时会自动激活目标语言,因此,在 L2 词汇提取的过程中,L1 词汇信息同时被激活,而由于被试 L2 水平相对薄弱,容易受到 L1 激活的干扰,导致被试需要在有限的时间内对 L1 进行较大程度的抑制而无法流畅地产出 L2 词汇;而在 L1 词汇提取过程中,虽然 L2 词汇信息被激活,但 L1 能对其进行有效的抑制 (Grosjean & Li, 2013),因此被试能较好地完成 L1 词汇的产出任务,这导致基于产出性任务数据构建而成的 L2 词汇网络不如 L1 词汇网络结构优化。

结语

本研究结合复杂网络分析和语义流畅任务,对晚期汉英双语者词库的结构进行了探索,结果发现:①双语词库均具有小世界属性和模块化属性;②双语者 L1 词汇网络比 L2 词汇网络的结构更加优化;③双语词汇网络间节点的组织方式支持激活扩散模型假设。

本文研究结果与前人研究结果 (Borodkin et al., 2016; Wilks & Meara, 2002) 整体相符。本研究表明双语者的双语语言距离并不会对词库结构的属性和差异造成影响,同时也展示了复杂网络分析在双语词库探究中的稳定性和可行性。

本研究可以为英语词汇教学提供相应的建议。在词汇教学过程中,教师可以根据学习者心理词库的结构特征来改进教学模式。例如,由于双语者的词汇网络具有小世界属性,教学者可根据此采用复杂网络分析中的贪心演算法^① (Borgatti, 2006) 或者大语言模型 (如 ChatGPT) (宋飞等, 2023), 提取学习材料中的词汇中心节点,让学习者先掌握中心词汇,再引导他们顺带习得与中心词汇在语音和语义层面相联结的词汇,帮助学习者高效科学地构建和完善 L2 词库。此外,如上文所述,以 L1 翻译为主的传统词汇教学方式可能是导致 L2 词汇网络模块化程度较低的原因,因此,为了减少 L1 对 L2 词库构建的干扰,教师可以尝试使用 L2 词汇和图片或实物直接相联的教学方法来加强 L2 词汇与概念之间的联系 (吴诗玉等, 2017), 进而帮助学生提高 L2 词库的模块化程度,促进 L2 词库结构的动态发展。另外,本研究还认为, L2 词库的模块化程度不高的另一个原因可能是因为 L2 较少参与知识结构的构建过程,因此,可以增加 L2 在知识结构构建过程中的比重。例如,可考虑将 EMI 教学模式引入课堂 (让教师使用英语教英语以外的其他科目,如历史、地理和生物等), 有利于完善双语者 L2 词库的结构。当然,这些教学建议是否可行,还需要未来的研究结合复杂网络分析和英语词汇教学法进一步加以探索。

总之,本研究的结果展示了复杂网络分析在心理词库探究中的可行性,并从结构特征和结构差异层面对晚期汉英双语者词库进行探讨,增进了对双语词库结构的了解,也为英语词汇教学提供了一些建议。但是,本研究作为复杂网

^① 贪心演算法是指提取中心节点的一种常用算法。该算法首先随机提取部分节点,然后观察这部分节点的提取是否会影响整体网络的中心度,如果中心度提升了,则将提取的节点随机和整体网络的节点进行替换。这个过程迭代进行,直到中心度下降,迭代停止,此时提取的节点则为中心节点 (Borgatti, 2006)。然而,中心节点并不是本研究的关注点,因此本文不加以赘述。

络分析在汉英双语词库研究中的初探, 存在以下缺点: 第一, 本研究的被试数量不多, 如若增加被试数量, 估计可以提高研究结果的推广度; 第二, 没有考虑以 L2 为知识系统构建用语的被试, 如香港的汉英双语者 (香港的大学课堂主要以英语为教学用语, L2 是他们知识体系构建的主要工具), 对他们而言, L2 词库模块化的程度是否会更高, 或者是否会与 L1 词库模块化程度相当? 这些问题均有待进一步探讨; 第三, 仅以词与词之间的关联系数作为网络的边线权重, 而心理词库是一个非常复杂的系统, 词与词之间的关联可能还存在错综复杂的关系, 因此, 未来在构建词汇网络时可将词与词之间的多重联系 (如句法关系) 加以考虑, 从而多维度地丰富对词库结构的探索。希望本文可以抛砖引玉, 吸引更多的研究者使用复杂网络分析探讨双语词库的结构以及二语词汇习得的发展进程。

参考文献:

- [1] ACEVEDO A, LOEWENSTEIN D A, BARKER W W, et al. Category fluency test: Normative data for English- and Spanish-speaking elderly [J]. *Journal of the International Neuropsychological Society*, 2000 (7): 760-769.
- [2] ANTÓN-MÉNDEZ I & GOLLAN T H. Not just semantics: Strong frequency and weak cognate effects on semantic association in bilinguals [J]. *Memory & Cognition*, 2010 (6): 723-739.
- [3] ARBESMAN S, STROGATZ S H & VITEVITCH M S. The structure of phonological networks across multiple languages [J]. *International Journal of Bifurcation and Chaos*, 2010 (3): 679-685.
- [4] ARDILA A, OSTROSKY-SOLÍS F & BERNAL B. Cognitive testing toward the future: The example of Semantic Verbal Fluency (ANIMALS) [J]. *International Journal of Psychology*, 2006 (5): 324-332.
- [5] BOCK K & LEVELT W. Language production: Grammatical encoding [C] // GERNSBACHER M A. *Handbook of Psycholinguistics*. San Diego: Academic Press, 1994: 945-984.
- [6] BORGATTI S P. Identifying sets of key players in a social network [J]. *Computational & Mathematical Organization Theory*, 2006 (1): 21-34.
- [7] BORGE-HOLTHOEFER J & ARENAS A. Semantic networks: Structure and dynamics [J]. *Entropy*, 2010 (5): 1264-1302.
- [8] BORODKIN K, KENETT Y N, FAUST M, et al. When pumpkin is closer to onion than to squash: The structure of the second language lexicon [J]. *Cognition*, 2016 (1):

60–70.

- [9] BULLMORE E & SPORNS O. The economy of brain network organization [J] . *Nature Reviews: Neuroscience*, 2012 (5) : 336–349.
- [10] CASTRO N & SIEW C S Q. Contributions of modern network science to the cognitive sciences: Revisiting research spirals of representation and process [J] . *Proceedings of the Royal Society A: Mathematical, Physical, and Engineering Sciences*, 2020 (6) : 1–25.
- [11] CHEN H, CHEN X & LIU H T. How does language change as a lexical network? An investigation based on written Chinese word co-occurrence networks [J] . *Plos One*, 2018 (2) : 1–22.
- [12] CHEN Y & ZHOU R. The mental lexicon features of the Hakka-Mandarin dialect bilingual [J] . *Brain Sciences*, 2022 (12) : 1–20.
- [13] COLLINS A M & LOFTUS E F. A spreading-activation theory of semantic processing [J] . *Psychological Review*, 1975 (6) : 407–428.
- [14] DE DEYNE S & STORMS G. Word associations: Network and semantic properties [J] . *Behavior Research Methods*, 2008 (1) : 213–231.
- [15] ERDÖS P & RÉNYI A. On the evolution of random graphs [J] . *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, 1960 (1) : 17–61.
- [16] FABBRO F. *The Neurolinguistics of Bilingualism: An Introduction* [M] . London: Psychology Press/Taylor & Francis, 1999.
- [17] FERRER-I-CANCHO R, SOLÉ R V & KÖHLER R. Patterns in syntactic dependency networks [J] . *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 2004 (5) : 1–8.
- [18] GROSJEAN F & LI P. *The Psycholinguistics of Bilingualism* [M] . New York: John Wiley & Sons, 2013.
- [19] HILLS T T, MAOUENE J, RIORDAN B, et al. The associative structure of language: Contextual diversity in early word learning [J] . *Journal of Memory and Language*, 2010 (3) : 259–273.
- [20] HUMPHRIES M D & GURNEY K. Network ‘small-world-ness’: A quantitative method for determining canonical network equivalence [J] . *Plos One*, 2008 (4) : 1–10.
- [21] KAVÉ G. Phonemic fluency, semantic fluency, and difference scores: Normative data for adult Hebrew speakers [J] . *Journal of Clinical and Experimental Neuropsychology*, 2005 (6) : 690–699.
- [22] KELLO C T & BELTZ B C. Scale-free networks in phonological and orthographic

- wordform lexicons [C] //PELLEGRINO F, MARSICO E, CHITORAN I, et al. *Approaches to Phonological Complexity*. Berlin: Walter de Gruyter Mouton, 2009: 171–190.
- [23] KENETT Y N, KENETT D Y, BEN-JACOB E, et al. Global and local features of semantic networks: Evidence from the Hebrew mental lexicon [J] . *Plos One*, 2011 (8) : 1–14.
- [24] KENETT Y N, WECHSLER-KASHI D, KENETT D Y, et al. Semantic organization in children with cochlear implants: Computational analysis of verbal fluency [J] . *Frontiers in Psychology*, 2013, 4: 1–11.
- [25] KROLL J F & TOKOWICZ N. Models of bilingual representation and processing: Looking back and to the future [C] //KROLL J F & DE GROOT A M B. *Handbook of Bilingualism: Psycholinguistic Approaches*. Oxford: Oxford University Press, 2005: 531–553.
- [26] NATION I S P & BEGLAR D. A vocabulary size test [J] . *The Language Teacher*, 2007 (7) : 9–13.
- [27] NEWMAN M E J. Modularity and community structure in networks [J] . *Proceedings of the National Academy of Sciences of the United States of America*, 2006 (23) : 8577–8582.
- [28] ROBERTS P M & LE DORZE G. Semantic organization, strategy use, and productivity in bilingual semantic verbal fluency [J] . *Brain and Language*, 1997 (3) : 412–449.
- [29] ROSSELLI M, ARDILA A, SANTISI M N, et al. Stroop effect in Spanish-English bilinguals [J] . *Journal of the International Neuropsychological Society*, 2002 (6) : 819–827.
- [30] SIEW C S, WULFF D U, BECKAGE N M, et al. Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics [J/OL] . *Complexity*, 2019: 2108423. (2019–1–17) [2019–10–1] . <https://doi.org/10.1155/2019/2108423>.
- [31] STEYVERS M & TENENBAUM J B. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth [J] . *Cognitive Science*, 2005 (1) : 41–78.
- [32] VITEVITCH M S. What can graph theory tell us about word learning and lexical retrieval? [J] . *Journal of Speech, Language, and Hearing Research*, 2008 (2) : 408–422.
- [33] WATTS D J & STROGATZ S H. Collective dynamics of ‘small-world’ networks [J] . *Nature*, 1998, 393: 440–442.

- [34] WILKS C & MEARA P. Untangling word webs: Graph theory and the notion of density in second language word association networks [J]. *Second Language Research*, 2002 (4): 303-324.
- [35] 陈芯莹. 从复杂网络看汉字句法功能的演变 [J]. 山西大学学报 (哲学社会科学版), 2015 (2): 48-54.
- [36] 陈芯莹, 刘海涛. 汉语句法网络的中心节点研究 [J]. 科学通报, 2011 (10): 726-731.
- [37] 高松, 颜伟, 刘海涛. 基于树库的现代汉语动词句法功能的计量研究 [J]. 汉语学习, 2010 (5): 105-112.
- [38] 刘海涛. 语言复杂网络的聚类研究 [J]. 科学通报, 2010 (27-28): 2667-2674.
- [39] 刘海涛. 语言网络: 隐喻, 还是利器? [J]. 浙江大学学报 (人文社会科学版), 2011 (2): 171-182.
- [40] 宋飞, 郭佳慧, 曲畅. ChatGPT 在汉语作为外语教学中的应用体系及实践 [J]. 北京第二外国语学院学报, 2023 (6): 110-128.
- [41] 孙继平, 孙秀丽. 心理词库研究视角下的二语词汇习得——一项基于词汇联想网络体系的实证研究 [J]. 外语研究, 2014 (1): 42-46.
- [42] 吴诗玉, 马拯, 胡青青. 中国英语学习者词汇与概念表征发展研究: 来自混合效应模型的证据 [J]. 外语教学与研究, 2017 (5): 767-779.
- [43] 熊永红, 黄元, 王云龙. 激活扩散模型与英语词汇教学中的常规关系建构 [J]. 外语教学, 2011 (4): 61-64.
- [44] 徐歌. 基于构建式词汇联想的二语心理词库组织模式及重组研究 [J]. 西安外国语大学学报, 2016 (1): 76-79.

收稿日期: 2019-10-23

作者信息: 陈瑶, 博士, 华南师范大学外国语言文化学院, 510631, 研究方向: 心理语言学。电子邮箱: yaochenscnu@gmail.com

周榕 (通讯作者), 华南师范大学外国语言文化学院教授, 博士生导师, 510631, 研究方向: 心理语言学。电子邮箱: zhourong@m.scnu.edu.cn

A Study of the Mental Lexical Structures of Late Chinese-English Bilinguals Based on Complex Network Analysis

Chen Yao / Zhou Rong

(South China Normal University, Guangzhou 510631, China)

Abstract: This study employed semantic fluency task and complex network analysis to investigate the lexical structures of late Chinese-English bilinguals. By obtaining network visualization and network parameters, including average nodes,

BEIJING DIER WAIGUOYU XUEYUAN XUEBAO

北京第二外国语学院学报

1979 年创刊

ISSN 1003-6539 CN 11-2802/H

2024 年第 1 期（第 46 卷第 1 期）（总第 297 期）

2024 年 2 月出版

主 管：北京市教育委员会 主 办：北京第二外国语学院
主 编：计金标 副主编：程 维 邱 鸣
执行主编：周长银 执行副主编：许传华

特约顾问：

鲍川运（美国蒙特雷国际研究院） 陈 平（澳大利亚昆士兰大学）
戴炜栋（上海外国语大学） Esther Geva（University of Toronto）
胡壮麟（北京大学） 黄友义（中国外文局）
陆俭明（北京大学） 潘海华（香港中文大学）
沈家煊（中国社会科学院） 王克非（北京外国语大学）
王 宁（上海交通大学） 文秋芳（北京外国语大学）

编委会主任：计金标 编委会副主任：程 维 邱 鸣

特约编委：

曹顺庆（北京师范大学） 陈众议（中国社会科学院）
程 工（浙江大学） 胡建华（中国社会科学院）
黄国文（华南农业大学） 蒋洪新（湖南师范大学）
刘和平（北京语言大学） 陆建德（中国社会科学院）
马会娟（北京外国语大学） 马秋武（北京语言大学）
苗兴伟（北京师范大学） 聂珍钊（浙江大学）
冉永平（广东外语外贸大学） 束定芳（上海外国语大学）
王初明（广东外语外贸大学） 王东风（中山大学）
王铭玉（天津外国语大学） 王守仁（南京大学）
王文斌（北京外国语大学） 文 旭（西南大学）
徐一平（北京外国语大学） 杨金才（南京大学）
张 辉（南京师范大学） 周领顺（扬州大学）
庄智象（上海外国语大学）

校内编委：

程 维 金振杰 李洪波 李美霞 李向民 林允清 刘学慧 刘 燕
司显柱 隋 刚 王柯平 肖 凌 许传华 杨 玲 张惠芹 张喜华
周长银 周 烈

average degree, average clustering coefficient, average shortest-path length, small-world index, and modularity index, this study characterized the lexical structure from the perspectives of structural features and structural difference. First, it was found that the bilinguals’ lexicons were small-world and modular in nature. Second, compared with the L2 lexical network, the L1 counterpart was better organized, which may be attributed to the influence of L2 proficiency and the age of L2 acquisition. Third, the results concerning how words within the network are connected supported the hypothesis of the spreading activation model. Based on the study’s findings, the strategies recommended for teaching English vocabulary in China include emphasizing alignment with learners’ developing mental lexicon and using methods that directly connect L2 vocabulary with images or real objects to enhance dynamic L2 vocabulary growth.

Keywords: complex network analysis; semantic fluency task; mental lexicon; late Chinese-English bilinguals; bilingual lexicon

（责任编辑：魏 鸣）

征订启事

本刊通过邮局及中国国际图书贸易集团有限公司（海外）征订发行，每期定价 18 元，国内邮发代号：80-630，国外发行代号：M6997，敬请广大订户与邮局、中国国际图书贸易集团有限公司联系，或微信扫描下图二维码，前往中国邮政网上营业厅订阅本刊。

《北京第二外国语学院学报》是经国家新闻出版总署批准，北京市教育委员会主管，北京第二外国语学院主办，在国内外公开发行的高校学报（双月刊）。

除传统的“语言学研究”“翻译研究”“外国文学研究”“外语教学研究”等固定栏目外，《二外学报》还将根据各期的关注重点，设立“名家论坛”“学术争鸣”“学术前沿”“博士新论”“重点项目追踪”“文化研究”“国别研究”“‘一带一路’与国家语言战略研究”“典籍翻译与中国文化‘走出去’”等机动栏目。另外，在传统栏目中会不定期设立反映各研究领域最新成果的专栏。

