

基于大数据的语义实证研究新进展^{*}

浙江大学 王 琿 吴义诚

提要:任何人类语言都是音 - 义的结合体。语音形式听得见,把语音转写出来的书写形式也看得见,然而语言意义既看不见又摸不着,故语义研究比语音、音系、句法、词法等要困难得多。对语义的形式刻画即形式语义学研究已属不易,对其进行实证研究似乎更是难上加难。然而,国际学界基于大数据的语义研究近期取得了突破性进展,主要体现在三个方面:1)语言符号的音 - 义结合关系并非完全任意;2)单词的长度与其语义内容呈正相关关系;3)单词按照“近邻链”的省力原则不断发展中新的意义。本文拟就相关成果进行评介,以期为国内同行开展基于汉语的语义研究提供有益的启示。

关键词:音 - 义联系、非任意性、词义扩展

[中图分类号] H030 [文献标识码] A [文章编号] 1000-0429 (2020) 04-0620-09

1. 引言

众所周知,语言是人类社会发展到一定阶段的产物,是人类对周围世界有规律的认识所产生的结果。人类的认知活动与所创造的语言符号之间的匹配是否完全任意?人类语言的词汇有限,但意义无限,因此人们需要从有限的词汇发展出无限的意义,然而词汇的语义内容是如何扩展的?针对这些涉及语义与认知的重大问题,国际学界基于大数据进行了实证研究(参见沈园 2017),并取得了突破性进展。相关成果主要体现在三个方面:1)语言符号的音 - 义结合关系并非完全任意;2)单词的长度与其语义内容呈正相关关系;3)单词按照“近邻链”的省力原则不断发展中新的意义。本文将对这些研究成果进行评述¹,以期为国内同行开展语义研究课题提供有益的启示。

* 本研究是国家社科基金重大项目“基于大型语料库的汉语弱组构性历时演变与语言演变规律研究”(18ZDA292)的阶段性成果。《外语教学与研究》编辑部及匿名审稿专家提出了宝贵的修改意见,谨致谢忱。

1 本文评述的文献均刊于基础学科领域顶级期刊 *Proceedings of the National Academy of Sciences of the United States of America*(《美国科学院院报》)。

2. 音 - 义结合中普遍存在的非任意性

语言中音 - 义的结合关系自古以来一直是人们感兴趣的话题,哲学家、语言学家及人类学家等都曾对此进行过不同程度的探讨。例如,苏格拉底 (Socrates, B.C. 469-399) 认为语符声音与其指称对象 (referent) 之间的关系是贴合的;印度最古老的哲学著作《奥义书》(*Upanishads*)² 记载着不同的元音及辅音与其所代表的意义之间的关系是有规律可循的;“俄国科学之父”罗蒙诺索夫 (Mikhail Lomonosov, 1711-1765) 认为具体的声音承载着具体的意义。但此类观点在当时并未引起人们足够的重视。约一个世纪之后,“现代语言学之父”索绪尔在其《普通语言学教程》中提出了语言符号音 - 义结合的任意性观点,并将语符的任意性视作语言学中最根本的原则,是整个语言系统的基础 (Saussure 1916);随着《普通语言学教程》的影响力日益增长,音 - 义结合的任意性观点逐步深入人心,在语言学领域似乎具有不可动摇的地位。

尽管如此,仍有不少语言学家如丹麦的 Jespersen (1922/1959)、英国的 Firth (1930/1964)、美国的 Bolinger (1949, 1950, 1980) 等对音 - 义结合的任意性观点不断提出质疑,国内学者如许国璋 (1988)、胡壮麟 (1994)、王寅 (1999) 等也对任意性原则持不同看法。Firth (1930/1964) 曾造了一个新词 phonestheme (“联觉音组”)来刻画语符在音 - 义方面的系统配对关系,例如,英语中的 /gl/ 音一般表示“能见度”、“亮度”,因此含有此音的单词多携带此义,如 glass、glare、glaze、gleam、glimmer、glint、glisten、glitter、glory 等。Bolinger 更是旗帜鲜明地提出“语符并非是任意性的”、“语素并非是最小的意义单位”(Bolinger 1949, 转引自 Magnus 1999: 12) 等观点,比语素更小的单位如音素也可以表达意义,如 crash、mash、bash、lash、slash、gnash、splash 等词都与声音有关,因为其中的 -ash 表示噪音 (Bolinger 1980: 20)。

近年来,语言符号的非任意性问题再次成为一项热门课题。非任意性 (non-arbitrariness) 指单词的读音与意义之间的象似性,即通过具体声音可推测出其携带的语义内容,其中拟声词和感叹词是语符非任意性的典型例证。除此之外,语言中大量基本词汇 (名词、形容词等) 的声音与语义内容或许也存在着不同程度的非任意性。以往学者对非任意性特征的考察因所选语料较少,未能引起学界足够的重视。为弥补这一缺陷, Blasi *et al.* (2016) 对全球 6452 种语言进行了广泛且深入的考察,按照 Swadesh list (Swadesh 1952, 收录了出现频率在全球语言中均较高的 100 个词项), 从受试语言中分别选取与 100 个词项具有相同意

² 《奥义书》源自《吠陀》经典的最后一部分,主要讲述人与宇宙的关系,曾入选“有史以来最有影响的一百部著作之一”(胡作玄 2004: 159-162)。

义的单词,基于大量的语言事实,证实了人类语言中普遍存在具有非任意性特征的音-义结合现象。下面重点介绍 Blasi *et al.* (2016) 的研究。

为描述音-义结合非任意性这一特征,Blasi 等人首先探讨了语言的象音现象(ideophone)及象似性(iconicity)。例如,梅尔语(Mel)常用重复单个单词的方式表明其数量之多,如 hábá 一词意为“人”,hábá-hábá-hábá 则表示“许多人”;再如,人们通常会下意识地使用特定的嘴型和手势来表达与其相关的意义,如在发英语的 you 音时,嘴型外凸,表示在外的“你”。为验证这一现象,Blasi 等人令一组母语为英语的受试从其不认识的英语单词中选取对应的反义词,结果显示,该组受试的正确率高于随机水平,这说明人们对音义之间的联系较为敏感,即人们在认知活动中会把具体的声音与特定的意义结合起来。

另外,Blasi 等人对这些基本词汇的发音进行了定量描述,统计出 74 种适用于其研究的 6452 种语言中的音-义结合关系,包括概念关系和符号象征关系。比如:表达 small 的意义时,多数语言常会与发音短促、嘴型较小的前高元音 /i/ 相联系,如英语中的 petty、little、bit、sip 等词,其中一些概念意义可能与不止一个符号相关,如前高元音通常会触发腭音化,因此在表达 small 这个概念时,也会出现与清辅音中的硬腭齿龈擦音 /s/ 相结合的现象,如英语中的 silk、slice、slim、slip 等词;除表事物特性的词具有音-义结合的非任意性外,作者还发现与人体身体部位相关的词同样具有音-义结合的非任意特征,比如与 tongue 相关的概念通常会出现舌侧音 /l/, 如英语中的 loud、laugh、call 等词。为探究这些语言为何有着类似的音-义结合关系,他们对语言分布的地理区域及其相似性进行了调查,结果表明:人们普遍能够使用特定声音来表达特定意义,这与人类居住的地理区域远近并无多大关联。这些散居在世界各处的人为何不约而同地使用相似的声音来表达特定的意义呢?他们认为,这些音-义结合的非任意性特征可能是由最初仅有口语形式的原始语言通过演化而保留下来的,但是是否存在最原始的史前语言因缺乏典籍资料不得而知。最后他们抛出了一个引人深思的开放性问题:既然语言分布的地域差异及原始语言种系发生所遗留的特征不能解释为何全球人类均使用类似声音来表达特定意义,那么这种普遍存在的音-义结合非任意性特征可能与其他因素有关,如声音的象征意义(symbolism)、象似性(iconicity)或人类通感(synaesthesia)(Ramachandran & Hubbard 2001)。

Blasi 等人的研究对象虽包括汉语,但对如何探究汉语的音-义结合现象并未给予描述,我们猜测他们可能是把汉语转换成拼音字母后与其他字母语言共同研究,这未免淡化了汉语作为表意文字的特征。表意文字本身的字形便可表达一定意义,因此这类文字更多的是形义结合,但汉语的音-义结合现象根据

作者们的研究也呈现出了非任意性特征。这意味着汉语不仅仅是形 - 义的结合体,还可能是音 - 形 - 义的结合体。例如,英语中表示“细小”义的词一般带有发音时口型窄小的 s 音,如 small、seed、shack、slice、slight、slim、sip、souvenir 等;同理,汉语里表示“细小”义的词带有发音时口型窄小的 x 音,如细、隙、溪、小、狭、瑕、纤、星等。当然,汉语具体在哪些词语上以及多大程度上体现了音 - 义结合的非任意性特征有待国内语言学工作者进行更深入细致的研究。除此之外,作者在对超过全球 2/3 的语言进行计量分析时,分析方法降低了因语言差异或选词特殊而造成的过分音义非任意现象,研究结果可信度高,证实了语符音 - 义结合的确普遍存在着非任意性特征,这对语言学界流行的语符任意性观点来说无疑是个巨大的挑战。

以往的语义分析通常按照从大到小的次序,把一个语义单位进行细致切分直至语素,不再探究其他更为细致的语义单位。如今这项基于大数据的语义研究打破了这一藩篱:由于具体声音往往承载着一系列相关联的特定意义,因此有可能音素才是语言中最小的意义单位。这一发现至少说明人类语言的高度复杂性。毋庸置疑,任意性和象似性皆是客观存在的事实(参见沈家煊 1993; 朱永生 2002 等),但后者具体体现在哪些层面上以及文化因素等对音义结合关系有多大影响,这些都是我们亟待深入研究的课题。

3. 单词长度与语义内容

语义即语言的意义,这种意义是抽象的,存储于人的大脑中,既看不见也摸不着,属于认知心理范畴上的概念。但这种意义可以通过具体的声音或文字形式表达出来,因此语义研究即是研究这些声音和文字符号所携带的意义以及意义与其所指客体之间的关系,通常包括对词汇意义、语法意义、句子意义三个层面的研究。在这三个层面中,词汇是最基本的内容,句子意义及语法关系均要通过具体词汇的有序组合才得以实现,因此本文中语义内容的研究主要关注词汇(以下所言词汇均指实词)意义的研究。

哈佛大学语言学家齐夫(Zipf 1949)发现,人类语言中使用频率高的单词往往较短。人们在日常交际中,倾向于把使用频率高的意义赋予短词形式之上,也即是短的实词常含有较多语义内容,以此达到交流的高效性目的,即言语交流的最省力原则(Principle of Least Effort)。我们知道,交际的高效性并不仅仅因为所使用的词汇形式短,还与交际双方为成功达到交际目的所做的努力有关。参与交际的人群会对会话做出推理,这种推理使交流变得精简,正如 Levinson (2000: 29) 所言: Inference is cheap, articulation expensive。由于具体的声音可能携带某种特定的意义,而短实词势必只有较少的音节,发音时间上也较短;

那么这些音节少的短词,是否有能力在语言演变的历史长河中依靠单薄的音素去创造丰富的语义内容呢?

针对上述问题, Piantadosi *et al.* (2011) 基于互联网上一个万亿词的大型词库,对词汇的使用频率、词汇长度及语义内容三者之间的关系进行了考察。他们假设词汇长度的编码取决于该词汇在语境中所具有的平均语义内容,语义内容决定词长观点与齐夫的使用频率决定词长观点的不同之处在于:首先,这样一个以语义内容决定词长的词汇系统可能不是最精简的,因其不会刻意缩短那些语义内容极其丰富的词汇,即便有其他更为精简的词汇形式可代替;其次,基于语义内容决定词长更能够让该项语义内容在交际中保持稳定。针对捷克语、荷兰语、英语、法语、德语、意大利语、波兰语、葡萄牙语、罗马尼亚语、西班牙语和瑞典语共 11 种语言,他们从谷歌语料库中选取了使用最为频繁的 25,000 项词条。为了保证数据的有效性,所选词条也必须同时出现在 OPUS (Tiedemann 2009) 语料库中。对数据进行二元、三元、四元模型 (N-gram) 分析后的结果显示,词越长,其语义内容越多。为验证结果的信度,并且确保从谷歌语料库中选取的语料不带有人工选择因素,他们采用 BNC 语料库,复制了整个研究过程,最终得出与上述相似的结果,进一步验证了词汇长度与其语义内容呈现正相关关系。

Piantadosi 等学者的此项研究并未涉及汉语,研究对象为印欧语系的几种语言。我们认为,学界应提高对另一大语系——汉藏语系的关注度,尤其要加强对汉语的语义研究。另外,该团队在确定语义内容时,并未给予标准,未参照词典或相关词汇权威来解释词语所具有的意义,根据语料库及上下文推断语义内容未免有人为因素或者过度推断,如果文中提供了语义内容可参照的标准,相信其研究结果将更有说服力。

除此之外,上述研究成果向人们展示了词汇的长度与其语义内容呈现正相关关系,是意义编码的交际高效性缩短了词长,而非频率缩短了词长。这一观点修正了齐夫有关词长与词汇使用频率之间关系的认识,同时也从侧面证实了声音与意义的非任意关系,词汇越长,其音节及音素也会越多,因此有可能携带更多的意义,从而丰富了词汇的语义内容。例如 friend 与 friendliness、gentleman 与 gentlemanliness,后者较前者语义内容更多也更抽象。但这一发现似乎更适合屈折语和黏着语,因其可通过添加前后缀改变词汇的语义内容,对于像汉语这样的孤立语来说,词的形式是固定不变的,语义的改变需依靠不同的词汇组合来完成,如“朋友”与“友谊”,二者均为两字词语,但后者的语义内容显然比前者更为抽象。如此看来,不同语言里词汇的语义内容与词长关系不同 (Grzybek 2006)。对语义的研究应根据不同语系区别对待,不能一概而论。无论屈折语、黏着语还是孤立语,其语义内容的扩展可能有各自不同的方式,这些方式是否具

有共同的核心特征？由于人类语言的词汇有限，意义无限，人们需要从有限的词汇发展出无限的意义，接着我们就谈谈词汇的语义内容是如何发展的。

4. 语义内容的扩展机制

每种语言都面临着用有限词汇表达无限思想的压力，其中一种解决方法是创造新词，然而更为常用的方法则是从已知词汇上扩展出新意。词义扩展造成了词汇的一词多义现象，普遍存在于人类语言中，在词汇的历史发展中占据主要地位。一般认为，语义内容的扩展有着原型范畴（prototype）的特点（Rosch 1975），如 face 可指身体部位、面部表情、社会面子等等，多项语义内容之间总有某种关系，这些语义项均与 face 的本意有关。词汇的语义内容在不断扩展时，新义项也会与词汇的原始意义有关。但这种观点不能解释有些词汇的义项随时间的发展而慢慢淡化甚至消失的现象。随后的模型理论（exemplar theory）认为词汇原有的意义对新义项的产生具有示范效应，旧义项均为新义项的产生提供基础（Medin & Schaffer 1978；Nosofsky 1986）。Ramiro *et al.* (2018) 就词汇语义内容如何扩展的问题进行了探讨，试图提出一种能够阐释词汇语义内容扩展机制的计算方法。他们认为链条（chaining）结构能解释语义扩展的规律，链条连接使得新义项得以生成，且与原始义项高度相关，如此往复，便在语义空间内生成了一系列具有链条结构的语义内容项目（如图 1），

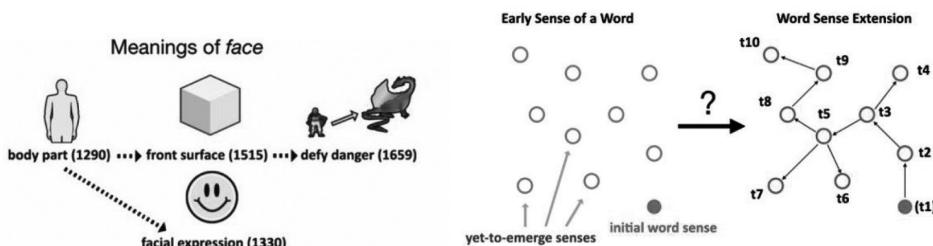


图 1. 词汇按“近邻链”原则发展出新义的路径（同上：2324）

如图 1 所示，face 的本义与身体部位有关，后隐喻为物体的前面、表面，进而发展成更为抽象的含义“面对”（面对危险等），链条结构之所以能够解释语义扩展的规律，是因其很好地诠释了认知最简化原则：从一个已知义项发展出的新义项需与此已知语义项有最大关联性（Sperber & Wilson 1995），此链条结构并非一维的线性结构，即并非所有的语义内容都是以一个链条连接点为基点去延伸新的语义内容，任何一项语义内容都可能成为一个基点，以此发展出多项语义内容。此观点符合人类语言“省力高效”的发展规律（Jespersen 1922/1959：

324), 并且能够解释部分语义内容随时间发展而慢慢消失的现象, 随着语义链条发展得越来越多, 形如繁茂的树枝, 那些结构单薄的语义链因丧失再生能力而被抛弃, 相应的语义内容因此也逐渐消失。

为解释链条结构在语义发展方面所具有的省力高效性, Ramiro 等人提出了“近邻链”算法模型 (Nearest-Neighbor Chaining Algorithm), 同时也根据以往语义内容发展的观点提出了几种算法模型: 随机算法 (Random Algorithm)、模型算法 (Exemplar Algorithm)、原型算法 (Prototype Algorithm)、始源算法 (Progenitor Algorithm)、现时算法 (Local Algorithm)。为探究上述六种算法模型中哪一种最符合语言发展的省力高效原则, 最能解释语义扩展的规律, 他们重点研究了随时间顺序先后出现的相邻语义内容之间的语义距离, 距离越短说明新词义的产生越“省力”。词汇选取 BNC 语料库中常用的 5000 个单词, 这些单词也必须存在于 *The Historical Thesaurus of English* 词典中, 因该词典可帮助确定词汇语义内容出现的历史顺序。他们首先假定“近邻链”模型在词汇语义内容的发展方面具有最省力特征, 其次测试该模型与其他五种模型在预测语义内容涌现的先后顺序方面的解释力, 最后对六种模型两两之间进行评估, 得出“近邻链”原则最能解释语义内容的扩展规律, 且符合语言发展的省力原则。

可见, 词义的历史扩展顺序可以预测, 按照“近邻链”的原则不断发展出新义。在新词义产生的过程中, 有些词义往往通过加长原有词汇的形式来表达更为抽象的意义, 即单词越长, 词义可能越多也越复杂。人类在交际时, 为了表达更丰富的思想, 需对已有词汇加以改造来创造新义, 新词义若想为听话人快速理解, 需要与旧语义内容有最邻近的关联, 这样词汇的新义才会因其认知最简化特征得以传承。词汇新义的产生不仅归于交际因素、认知最省力因素等, 隐喻、转喻、双关等修辞手段也会导致新义项的产生。另外, 决定词汇语义内容多少的, 不仅与人为命题有关, 也与词汇本身的特征 (发音难度、单词长短等) 有关, 由于人们在表达思想时, 倾向于选用形式较简单的词汇, 因此这些形式简单的词汇生成多项语义内容的能力 (相较于同等形式简单的词来说) 更强, 但又因有限的音素所表达的意义有限, 人们在需要表达复杂、抽象的意义时, 又不得不采用形式相对复杂、音素较多的词汇。因此, 词汇的音素、结构复杂程度、语义内容之间的关系并非是绝对的, 而是相对的。除此之外, 社会因素、文化差异等宏观因子也会影响词汇的语义内容, 如 queer 这个词, 随社会的发展, 其贬义用法 (如同性恋、娘娘腔) 已经消失, 只剩中性含义, 指行为怪异的人。

5. 结语

本文对国际学界基于大数据的语义研究成果进行了评介, 其中的新发现对

语义研究具有重要启示作用：

首先,音-义结合在人类语言里具有较为普遍的非任意性特征。这给自索绪尔(Saussure 1916)以来语言学界盛行的语符任意性观点带来了很大的挑战。理论方面,音-义结合的象似性将有助于人们探索语言的共同起源问题;应用方面,它有助于帮助人们学习不同的语言,在外语教学中能够提高学习者对外语的兴趣。

其次,词汇的语义内容按照“近邻链”的省力原则不断发展中出新义。人类的认知遵循“省力高效”原则,词汇新义的产生总会按照与其原有意义最接近的方向发展出新义,以求将人们为理解新义所做的认知努力降到最低。“近邻链”原则或许不是解释词义扩展最好的一套理论,但到目前为止,它最适合解释语义内容的扩展规律。

本文评介的这些成果的共同之处在于,它们都建立在大数据的跨语言事实基础之上,研究对象涉及不同语系的诸多语言,同时也涉及不同的社会背景及地域差异等。目前国内学界的语义研究相对比较单薄,音义学(phonosemantics)的研究更是几近空白³。在人类语言之林,汉语是一个巨大的文化宝库,蕴藏着十分丰富的音、形、义资源。希望学界能多从跨语言的视角开展语言学研究,发扬“大胆质疑、小心求证”的科学精神,在包括汉语语义研究的领域取得突破性进展。

参考文献

- Blasi, D., S. Wichmann, H. Hammarströmb, P. Stadlerc & H. Christiansen. 2016. Sound-meaning association biases evidenced across thousands of languages [J]. *Proceedings of National Academy of Sciences of the United States of America* 113: 10818-10823.
- Bolinger, D. 1949. The sign is not arbitrary [J]. *Boletín del Instituto Caro y Cuervo* 5: 56-62.
- Bolinger, D. 1950. Rime, assonance and morpheme analysis [J]. *Word* 2: 117-136.
- Bolinger, D. 1980. *Language: The Loaded Weapon* [M]. London: Routledge.
- Firth, J. 1930/1964. *Speech* (Reprinted in *The Tongues of Men and Speech*) [M]. Oxford: OUP.
- Grzybek, P. 2006. History and methodology of word length studies [J]. *Contributions to the Science of Text and Language: Text, Speech and Language Technology* 31: 15-90.
- Jespersen, O. 1922/1959. *Language: Its Nature, Development and Origin* [M]. London: Allen & Unwin.
- Levinson, S. 2000. *Presumptive Meanings. The Theory of Generalized Conversational Implicatures* [M]. Cambridge, MA.: The MIT Press.
- Magnus, M. 1999. *Gods of the Word: Archetypes in the Consonants* [M]. Kirksville: Thomas Jefferson University Press.

³ 值得一提的是,美国学者 Schuessler (2007)曾对古汉字的音-义结合关系进行过系统的探讨。

- Medin, D. & M. Schaffer. 1978. Context theory of classification learning [J]. *Psychological Review* 3: 207-238.
- Nosofsky, R. 1986. Attention, similarity, and the identification-categorization relationship [J]. *Journal of Experimental Psychology: General* 115: 39-57.
- Piantadosi, S., H. Tily & E. Gibson. 2011. Word lengths are optimized for efficient communication [J]. *Proceedings of National Academy of Sciences of the United States of America* 108: 3526-3529.
- Ramachandran, V. & E. Hubbard. 2001. Synesthesia — A window into perception, thought and language [J]. *Journal of Consciousness Studies* 8(12): 3-34.
- Ramiro, C., M. Srinivasan, B. Malt & Y. Xu. 2018. Algorithms in the historical emergence of word senses [J]. *Proceedings of National Academy of Sciences of the United States of America* 115: 2323-2328.
- Rosch, E. 1975. Cognitive representations of semantic categories [J]. *Journal of Experimental Psychology: General* 104: 192-233.
- Saussure, F. 1916. *Cours de Linguistique Générale* [M]. C. Bally & A. Sechehaye (eds.). Paris: Payot.
- Schuessler, A. 2007. *ABC Etymological Dictionary of Old Chinese* [M]. Honolulu, HI.: University of Hawaii Press.
- Sperber, D. & D. Wilson. 1995. *Relevance: Communication and Cognition* [M]. Oxford: Blackwell.
- Swadesh, M. 1952. Lexico-statistical dating of prehistoric ethnic contacts: With special reference to North American Indians and Eskimos [J]. *Proceedings of the American Philosophical Society* 96: 452-463.
- Tiedemann, J. 2009. News from OPUS—A collection of multilingual parallel corpora with tools and interfaces [A]. In N. Nicolov & K. Bontcheva (eds.). *Advances in Natural Language Processing* [C]. Amsterdam: John Benjamins. 237-248.
- Zipf, G. 1949. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology* [M]. New York: Addison-Wesley Press.
- 胡壮麟, 1994, 语言的可证性 [J], 《外语教学与研究》(1): 9-15。
- 胡作玄, 2004, 《影响世界历史的 100 部名著排行榜》[M]。天津:天津教育出版社。
- 沈家煊, 1993, 句法的象似性问题 [J], 《外语教学与研究》(1): 2-8。
- 沈 园, 2017, 形式语义学和实验方法 [J], 《当代语言学》(2): 234-245。
- 王 寅, 1999, 《论语言符号象似性——对索绪尔任意说的挑战与补充》[M]。北京:新华出版社。
- 许国璋, 1988, 语言符号的任意性问题——语言哲学探索之一 [J], 《外语教学与研究》(3): 2-10。
- 朱永生, 2002, 论语言符号的任意性与象似性 [J], 《外语教学与研究》(1): 2-7。

收稿日期: 2018-09-12; 修改稿, 2020-03-28; 本刊修订, 2020-05-24

通讯地址: 310058 浙江省杭州市 浙江大学外国语文化与国际交流学院

studies revealed some aspect(s) of their global constructional attribute.

A multidimensional data analysis of overseas dissemination of Chinese classics *Lun Yu* (p. 580)

QIN Hongwu et al. (College of Foreign Languages, Qufu Normal University, Qufu 273165, China)

Reception is an important dimension in translation assessment, yet it is very difficult to evaluate it on a text-data basis. This study, by text data mining, examines the effectiveness of communication by comparing 11 versions of *Analects*. It is found that, generally there is little difference between the Chinese and foreign translators and that the key difference lies in the differently translated essential concepts which make it difficult to form in the receptor culture identifiable and unanimous meanings. The comparison between the 11 English versions and the original English texts on Confucianism shows that all translated texts have strong marks of otherness, which makes them less reader-friendly. Therefore, the role of translation in cultural communication needs to be reexamined and redefined. The study concludes that ideal translation must serve effective cross-cultural dialogue by taking reader response into account.

On the public diplomatic practice in book translation and dissemination: A case study of the Franklin Book Programs in the U.S. and its implications (p. 594)

WU Yun (School of Foreign Languages, Tongji University, Shanghai 200092, China)

Book translation and dissemination is a durable, stable and effective means in public diplomacy. Since the beginning of the Cold War, the United States has influenced the thinking and public opinion of the people across the world by translating American books. Franklin Book Programs, which began in 1952, was an exemplar for international propaganda, which promoted the political, economic and cultural expansion of the U.S. in the Middle East and Southeast Asia. This paper, by analyzing the principle, mechanism and strategies of the Franklin Book Programs, examines how the U.S. culture and values infiltrated the targeted countries through book translation during the Cold War.

On a legitimate construal of semantic waves: A case study of scientific texts (p. 607)

LUO Zaibing (School of Foreign Languages, Southwest Jiaotong University, Chengdu 400715, China)

YANG Bingjun (School of Foreign Languages, Shanghai Jiao Tong University, Shanghai 200240, China)

Semantic wave is a dialogic focus between LCT and SFL, but previous studies are not explicit enough in delineating its profiling, structural attributes and functions, and construing its distinct types. This study is intended to bring forth a legitimate construal of semantic waves in scientific texts on two bases: delicate semantic scales of semantic gravity in terms of rhetorical units of discourse; semantic density scales in terms of fractal elements within a message. It is found that semantic profiling depends on delicate scales of SG \pm /SD \pm ; structural attributes and functions lie in a correspondence between abstract levels of contextual specificity and semantic complexities in meaning making; the type of semantic waves depends on its waving position and functional nature of metalanguage. This study provides a legitimate construal for semantic waves and contributes to their theoretical integration between LCT and SFL.

Advances in big data-driven semantic studies (p. 620)

WANG Hui & WU Yicheng (School of International Studies, Zhejiang University, Hangzhou 310058, China)

Languages are sound-meaning pairings in which sounds are audible and their transcriptions are visible. Meanings, however, are neither audible nor visible. Therefore, semantic studies are notoriously difficult, be it formal or empirical, compared to phonetics, phonology, syntax and morphology. Recently, significant advances have been made in big data-driven semantic studies abroad in three aspects: 1) the sound-meaning relationship in words is not totally arbitrary; 2) the orthographic length of a word is in positive correlation with its amount of information; 3) lexical sense extension can be accounted for by the cost-effective “nearest-neighbor chaining” principle. We will introduce and comment on these latest findings to open new avenues for semantic studies in Chinese.