

英汉一对多对应词意义数量的统计研究及其加工的优势效应

李 佳¹ 赵全备^{2✉}

1. 北京语言大学 高级翻译学院 202221296092@stu.blcu.edu.cn

2. 北京外国语大学 中文学院 zqb@bfsu.edu.cn

摘要: 研究基于联结主义理论,以英汉一对多对应词为研究对象,基于英汉双语平行语料库获取英语词对应的全部汉语词,并对汉语词按照词义进行了合并与归类,确定出英语词对应汉语词的意义数量,在此基础上,通过掩蔽启动范式下的真假词判断任务,考察了对应词意义数量对中国英语学习者一对多词加工的影响,研究主要发现:意义数量越多的词,被试加工速度越快,但错误率更高。研究为跨语言视角下的对应词加工研究提供了来自不同书写系统的实证依据。

关键词: 语料库, 英汉一对多对应词, 意义数量, 统计研究, 加工优势

A Statistical Study of the Number of Meanings of English-Chinese One-To-Many Translation-Equivalent Word Pairs and Their Processing Advantage

Jia Li¹ Quanbei Zhao^{2✉}

Beijing Language and Culture University, School of Translation and Interpreting 202221296092@stu.blcu.edu.cn

Beijing Foreign Studies University, School of Chinese Language and Literature zqb@bfsu.edu.cn

Abstract: In the present study we categorize all Chinese meanings corresponding to the 68 English words in a parallel corpus respectively and count the numbers of these translation-equivalent word meanings. To examine the impact of the number of meanings on Chinese EFL learners' processing of English-Chinese one-to-many translation-equivalent word pairs, we design and carry out a truth-value judgment task within the masked priming paradigm, whose results demonstrate that participants' response time benefit from the higher number of meanings, but the accuracy rate reverse. This finding reveals empirical evidence entailing different writing systems for the translation-equivalent word pair processing study under the cross-linguistic perspective.

Keywords: corpus, English-Chinese one-to-many translation-equivalent word pairs, the number of meanings, statistical study, processing advantage

1 前言

任何两种语言的词语,除了一些科技术语和专有名词以外,根本不存在一对一的关系(胡明扬,1997),跨语言一对多关系是更普遍的现象,这一现象的研究对第二语言的学习和双语词汇加工都有重要影响(Degani & Tokowicz, 2010; Boada, Sánchez-Casas, Gavilán, García-Albea, & Tokowicz, 2013; Degani, Prior, & Eddington, 2016)。由于这一现象的研究不仅涉及语言学、心理学等领域,也与翻译学、词典学等语言学的实际应用领域紧密相连。因此,具有重要的理论价值和实践意义(赵翠莲,2005)。

研究者常关注在字词识别中是否会出现一对多词识别的优势效应,即与一对一词相比,加工一对多词的反应时更短,准确率更高。一些研究发现存在一对多对应词加工的优势效应(Lin & Ahrens, 2010; 李海龙、王同顺, 2013; Haro, Demestre, Boada, & Ferre, 2017),然而一些研究发现一对多词加工具有劣势效应(Laxen & Lavour, 2009; Basnight-Brown, Kazanas, & Altarriba, 2018; Zhou, Chen, Feng, & Zhou, 2019; 王堃、王悦, 2019)。值得注意的是,这些研究中大多比较了一对一词与一对多词的加工情况,而对一对多词内部意义数量的划分关注较少,尚未得出一致的结论,并且已有意义数量的研究主要集中于单语视角,对跨语言意义数量的关注较为缺乏。Jager & Cleland (2014)以意义数量为6作为分界,研究发现意义数量对多义词的加工有促进作用,意义数量越多,被试对多义词的加工越快, Hara & Ferre (2018)的研究以意义数量为10

作为分界,发现意义数量的多少,并不影响多义词的加工。Zgusta (1971) 将对应词定义为“目的语言中的一个词汇单位,它与源语言中相对应的词汇单位具有同样的词汇意义”,英语和汉语的语言概念体系差异较大,两种语言的词汇之间也因此具有高度的“不对称性”,比较明显的就是词目与对应词的高度不对称(丁骏,2013),比如:英语词“know”既可以对应汉语词“知道”,也可以对应“认识”以及“确信”等。英汉一对多对应词意义数量的多少如何影响学习者对对应词的加工仍存在一定的争议,需要进一步的考察。

查阅词典和被试词汇联想是确定跨语言一对多对应词意义数量的主要方式,但前者由于更新速度较慢,往往难以完全反映出语言使用的动态过程,后者由于学习者短时间内能够想到的对应词数量有限,难以对一对多对应词的意义数量进行细致地划分。此外,语料库统计作为一种补充方法日益受到关注,基于语料库的研究方法是在真实的语言使用过程中获取语言材料,统计结果往往与学习者实际的语言认知规律相符。语料库统计方法还可以和心理实验法进行交叉验证并日益受到学者们的重视(Ellis, Frey, & Jalkanen, 2009; 熊文新、赵秀花, 2015; 陆军、卫乃兴, 2018; 房印杰、梁茂成, 2020),这是因为基于语料库对语言材料进行统计分析,可以加深对语言属性的认识,为心理实验提供适合的实验材料,心理实验可以对语料统计结果的心理现实性进行检验并通过操纵自变量而引起学习者的特定反应,以进一步探讨学习者心理词汇的加工情况。

研究以联结主义理论为基础,该理论强调语言习得的过程实际上是学习者通过真实语言材料获取语言知识并储存于心理词典中的过程,并主张语言的统计属性和语言学习者习得的规律有很强的相关性(MacWhinney, 1997, 2005; 邢红兵, 2016)。鉴于上述考虑,研究首先基于大规模的语料统计,获取英汉一对多对应词的意义数量,在此基础上,考察跨语言视角下意义数量对中国英语学习者对应词加工的影响,为相关的研究提供来自不同语言系统的实证依据。

2 基于语料库的英汉一对多对应词意义数量的统计研究

2.1 语料来源及处理

语料来源于厦门大学卢伟教授开发的英汉双语平行语料库,包括英汉双语句子对齐语料 215713 句对,其中英文约 329 万词次,中文约 537 万字次,语料选自各种英汉对照文本,主要包括日常生活、文学、科技、法律等语域的口语和书面语语料。本研究中主要考察的是动词本身,由于英语词往往是和介系词一同出现,英语词中动词搭配介系词的情况也纳入了统计范围中。

2.2 意义数量的确定

研究以 68 个常见高频英语词为研究对象,这是因为在实际语言使用过程中高频词的对应词数量更多,更容易形成跨语言一对多词。研究首先通过语料库统计出英语词所对应的全部汉语对应词,在此基础上,以《牛津高阶英汉双解词典》(第九版)为参照,对属于同一意义的对应词进行合并,每个意义之下既包含了中心意义,也涵盖了与之相关的同义词或近义词,如“run”的汉语对应词“跑”和“奔跑”,对属于不同意义的对应词按照不同的词义进行单独分类,形成不同的意义词条,如“run”的汉语对应词“奔跑”“管理”,从而确定出英语词所对应的汉语词的意义数量。

表 1 英汉一对多对应词意义数量表

英语词	汉语对应词的意义数量	英语词	汉语对应词的意义数量	英语词	汉语对应词的意义数量	英语词	汉语对应词的意义数量
move	26	buy	10	speak	8	drop	33
accept	18	help	8	finish	8	leave	26
ask	18	use	4	sing	6	remind	4
hold	26	solve	8	stop	15	eat	5
send	17	allow	13	learn	8	write	13
open	17	keep	32	agree	22	arrive	18
read	17	exist	10	avoid	29	expect	10
visit	14	enter	23	treat	22	bring	31
join	14	enjoy	13	close	15	jump	22
lose	14	bear	27	cry	8	serve	16
live	18	show	29	play	17	run	33
try	11	study	8	increase	6	handle	6
find	20	blow	19	save	11	feel	12
know	9	lead	18	swim	10	give	30
talk	8	meet	21	introduce	15	make	46
lift	31	tell	13	walk	15	tidy	3
take	36	Wait	11	call	26	kick	10

3 实验研究

3.1 研究问题

基于语料库的统计结果,研究主要考察的是英汉一对多对应词意义数量是否影响中国英语学习者对应词的加工。

3.2 实验设计

研究使用的是掩蔽启动范式下的真假词判断任务,采用单因素实验设计,意义数量的多少为组内自变量。因变量为受试对英汉对应词判断的反应时和错误率。

3.3 被试

来自北京某高校的 33 名本科生和研究生参与了本次实验,年龄范围为 17-24 岁。

3.4 研究材料

研究在自建英汉对应词库的 68 对词中,匹配笔画数、字母数、汉语词词频和英语词词频后,以对应词意义数量 14 为分界,选取 26 对词作为实验材料,其中意义数量较少组(意义数量小于 14)和意义数量较多组(意义数量大于或等于 14)各 13 对。另外,构建了 26 对假词作为填充材料。

表 2 实验材料相关数据平均值和标准差

组别	义项数	笔画数	字母数	汉语词词频	英语词词频
意义数量	8.462	18.308	4.231	3.180	3.473
较少组	(2.259)	(4.385)	(1.363)	(.574)	(.403)
意义数量	21.154	16.000	4.231	3.076	3.578
较多组	(5.595)	(4.301)	(.599)	(.657)	(.281)

实验材料的匹配结果显示,意义数量组间差异主效应显著, $t(12)=8.124, p<.001$, Cohen's $d=2.253$ 。笔画数 ($t(12)=-1.489, p=.162$)、字母数 ($t(12)=.000, p=1.000$)、汉语词词频 ($t(12)=-.537, p=.601$) 和英语词词频 ($t(12)=.627, p=.542$) 组间差异主效应均不显著。统计结果说明材料的相关变量值符合分组要求,控制变量得到了有效控制。

3.5 实验程序

研究采用 E-Prime 软件在电脑上依次呈现启动词和目标词,被试通过按键来判断呈现的汉语词是否为合法词,如果是真词按“J”键,假词按“F”键。半数被试的用手按此规定,半数被试的用手规定相反。

实验中注视点以红色 72 号字体的“+”在戴尔笔记本屏幕中央呈现 500ms,接着呈现一串“#/曩”作为掩蔽刺激¹,启动词在同样的位置呈现 50ms 后,接着呈现目标词,直至被试做出反应。SOA 时长与前人相关研究一致,为 50ms。在实验中,屏幕背景为白色,分辨率为 1024×768,刷新率为 60 赫兹,32 位颜色深度。主试用中文讲述要求,告知被试尽快尽量准确地判断刺激是否为中文的合法词,被试先完成 16 个词对判断练习直至确认其熟悉流程,再完成正式实验。之后,受试还需要填写第二语言背景问卷(Li, Zhang, Yu, & Zhao, 2020)。

3.6 数据分析

正式实验后剥离了 1 名由于程序出错未能完整记录的受试数据,最终有效人数为 32 人,共记录了被试 832 个反应。只有意义数量较少和较多组进入分析(排除填充词和练习词),删除错误判断(0.729%)和被试在各实验条件下超过±3 个标准差的反应时数据(1.875%),根据被试在每种条件下的反应,然后进行统计检验。

3.7 研究结果

表 3 被试在意义数量较少和较多组真假词判断的反应时和错误率

	反应时	错误率
意义数量较少组	664.384 (99.153)	0.002 (0.014)
意义数量较多组	638.886 (70.797)	0.014 (0.031)
意义数量效应	+25.497*	+0.012

注: *代表 $p<.05$

t 检验结果表明,意义数量较多组与意义数量较少组真假词判断反应时差异显著 ($t(31)$)

¹ 当汉语词启动英语词时,前掩蔽为“#”,被试能够观察到汉语启动词的存在(Wang, 2013),参考 Wang & Forster (2010) 和 Wang (2013),本研究中采用了频率极低的汉字“曩”代替“#”,以达到更好地掩蔽效果。

=2.284, $p=.029$, Cohen's $d=.404$), 同意义数量较多的词 ($M=638.886\text{ms}$) 相比, 被试对意义数量较少的词 ($M=664.384\text{ms}$) 判断更慢。

意义数量较多组与意义数量较少组真假词判断错误率差异边缘显著 ($t(31)=1.973$, $p=.057$), 同意义数量较多的词 ($M=0.014$) 相比, 被试对意义数量较少的词 ($M=0.002$) 判断错误率更低。

4 讨论

研究基于英汉双语平行语料库的统计, 获取了英语词对应汉语词的意义数量, 并通过掩蔽启动范式下的真假词判断任务, 发现对应词的意义数量影响中国英语学习者一对多对应词的加工, 意义数量越多的词被试加工速度越快, 这与Rodd & Gaskell (2002)、Beretta (2005) 以及Jager & Cleland (2014) 的研究结论一致, 意义数量多的对应词包含更多的语义表征节点, 在词汇加工过程中, 会激活更多的语义信息, 根据分布式表征模型, 在词汇提取的初期阶段, 某个词义对应的多个语义节点的激活会互相促进而非竞争, 因此, 意义数量多的对应词能被更快的识别。Haro & Ferre (2018) 与Rodd & Gaskell (2002)、Beretta (2005) 以及Jager & Cleland (2014) 的研究同样根据词典义项确定意义数量, 却发现意义数量多的词与意义数量少的词识别速度并无显著差异, 因此意义数量的确定方法可能对研究结果产生一定的影响。本研究尝试以联结主义理论为基础, 基于学习者实际的语言使用情况, 通过对语料的深入分析, 对英汉一对多对应词意义数量进行了考察, 在一定程度上可以避免词典以及主观评测确定意义数量的局限, 为跨语言一对多对应词的选取及意义数量的确定提供了一定的参考和借鉴。

此外, 研究发现意义数量多的词被试识别速度更快, 但错误率更高。这样的结果可能与被试的认知资源分配有一定的关联, 金楷吟、王建勤 (2017) 认为第二语言学习者的注意力资源是有限的, 在语言加工的过程中, 如果只关注加工速度 (流利性) 的话, 必然以牺牲准确性为代价。此外, 本研究采取的是掩蔽启动范式且启动词与目标词呈现之间的时间间隔仅为 50ms, 通过对被试的访谈, 我们得知被试无法看清启动词的内容甚至根本无法意识到启动词的存在, 因此, 被试是在无意识的状态下完成了对目标词的判断, 尽管实验指导语提醒被试尽可能又快又准地完成判断, 较短的启动词与目标词呈现之间的时间间隔更使得被试在完成实验任务时处于一种节奏较快、高度紧张的状态, 被试往往难以兼顾加工速度与准确性, 因此, 更多的语义节点相互促进促使被试提高了加工速度, 但却以错误率提升为代价, 更少的语义节点使得被试加工速度减慢, 但也促进了准确率的提升。

跨语言对应词的研究中, 目前对相似书写系统的语言关注较多, 而对不同书写系统的语言, 如汉英关注较少 (王晓琳等, 2019; Zhou, Chen, Feng, & Zhou, 2019)。不同于前人研究在考察意义数量因素时, 主要集中在单语视角, 本研究以英汉两种不同书写系统的语言为研究对象, 考察了意义数量因素对跨语言对应词加工的影响, 在一定程度上为相关的研究提供了参考和借鉴。但在本研究中只考察了在 L2-L1 方向意义数量对对应词加工的影响, 尚未结合 L1-L2 方向进行对比研究, 此外, 研究中选取的被试大多为非平衡的汉英双语者, 未来的研究中可以对年龄范围更大且二语水平更高的被试群体进行考察, 除了高校学生外, 其他类别的被试 (如译员等) 也可作为考察对象, 进一步探讨意义数量对不同受试群体对应词加工的影响。

参 考 文 献

- [1] Basnight-Brown D, Kazanas S, Altarriba J. Translation ambiguity in Mandarin-English bilinguals: Translation production differences in concrete, abstract, and emotion words. *Linguistic Approaches to Bilingualism*, 2018,10:1~28.
- [2] Beretta A, Fiorentino R, Poeppel D. The effects of homonymy and polysemy on lexical access: an MEG study. *Cognitive Brain Research*, 2005,24:57~65.
- [3] Boada R, Sánchez-Casas R, Gavilán J, García-Albea J, Tokowicz N. Effect of multiple translations and cognate status on translation recognition performance of balanced bilinguals. *Bilingualism: Language and Cognition*, 2013,16:183~197.
- [4] Degani T, Prior A, Eddington C. Determinants of translation ambiguity A within and cross-language comparison. *Linguistic Approaches to Bilingualism*, 2016,6:290~307.

- [5] Degani T, Tokowicz N. Ambiguous words are harder to learn. *Bilingualism: Language and Cognition*, 2010,13:299~314.
- [6] Ellis N, Frey E, Jalkanen I. The psycholinguistic reality of collocation and semantic prosody: Lexical access. In: Römer U, Schulze R, eds. *Exploring the Lexis-Grammar Interface (Studies in Corpus Linguistics)*, 35:89~114. John Benjamins, Amsterdam. 2009.
- [7] Haro J, Demestre J, Boada R, Ferre P. ERP and behavioral effects of semantic ambiguity in a lexical decision task. *Journal of Neurolinguistics*, 2017,44:190~202.
- [8] Haro J, Ferre P. Semantic Ambiguity: Do Multiple Meanings Inhibit or Facilitate Word Recognition? *Journal of Psychological Research*, 2018,47:679~698.
- [9] Jager B, Cleland A. Polysemy Advantage with Abstract but Not Concrete Words. *Journal of Psychological Research*, 2014,45:143~156.
- [10] Laxen J, Lavour J. The role of semantics in translation recognition: effects of number of translations, dominance of translations and semantic relatedness of multiple translations. *Bilingualism: Language and Cognition*, 2010,13:157~183.
- [11] Li P, Zhang F, Yu AY, Zhao XW. Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, 2020,23:938~944.
- [12] Lin C, Ahrens K. Ambiguity advantage revisited: Two meanings are better than one when accessing Chinese nouns. *Journal of Psychological Research*, 2010,39:1~19.
- [13] MacWhinney B. A unified model of language acquisition. In: Kroll JF, de Groot AMB, eds. *Handbook of Bilingualism: Psycholinguistic Approaches*, 49~67. Oxford University Press, Oxford. 2005.
- [14] MacWhinney B. Second language acquisition and the competition model. In: MacWhinney B, ed. *Tutorials in Bilingualism: Psycholinguistic Perspectives*, 113~142. Lawrence Erlbaum Associates, Hillsdale. 1997.
- [15] Rodd J, Gaskell G, Marslen-Wilson W. Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, 2002,46:245~266.
- [16] Wang X, Language dominance in translation priming: Evidence from balanced and unbalanced Chinese-English bilinguals. *The Quarterly Journal of Experimental Psychology*, 2013,66(4):727~743.
- [17] Wang X, Forster K. Masked translation priming with semantic categorization: Testing the Sense Model. *Bilingualism: Language and Cognition*, 2010,13:327~340.
- [18] Zgusta L. *Manual of Lexicography*. Mouton, Hague. 1971.
- [19] Zhou GW, Chen Y, Feng Y, Zhou R. Processing of translation ambiguous words by Chinese-English bilinguals in sentence context. *Journal of Psychological Research*, 2019,48:1133~1161.
- [20] 丁骏.英汉词典中的对应词之现状分析. *复旦外国语言文学论丛*,2013,(1):53~58.
- [21] 房印杰,梁茂成.中国英语学习者关系代词取舍研究—语料库与实验法的交叉验证. *外语与外语教学*,2020,312(3):34~43+147.
- [22] 胡明扬.对外汉语教学中语汇教学的若干问题. *语言文字应用*,1997,(1):14~19.
- [23] 霍恩比. *牛津高阶英汉双解词典(第九版)*.北京:商务印书馆,2018.
- [24] 金楷吟,王建勤.第二语言学习者汉语口语产出的认知加工机制. *华文教学与研究*,2017,67(3):1~10.
- [25] 李海龙,王同顺.从歧义效应看中国英语学习者歧义词心理表征的发展. *现代外语*,2013,36(4):379~386.
- [26] 陆军,卫乃兴.二语搭配隐性、显性知识接口关系研究——基于语料库的心理语言实验研究范式探索. *外语教学与研究*,2018,(6):885~897+960.
- [27] 王莹,王悦.中、英多义词的翻译识别劣势效应及其影响因素. *心理研究*,2019,12(2):120~127.
- [28] 王晓琳,等.词类型、翻译优势及具体性效应在翻译歧义词加工中的交互. *当代外语研究*,2019,(2):57~69.
- [29] 邢红兵.汉语作为第二语言的词汇习得研究.北京:北京大学出版社,2016.
- [30] 熊文新,赵秀花.英语特异组合及其在外语教学上的验证. *外语教学*,2015,(4):55~59.
- [31] 赵翠莲.多义词心理表征的发展模式—基于对中国 EFL 学习者的实验研究. *外语教学与研究*,2005,37(4):294~302.