

汉语变体中分析型致使构式变异研究*

——多分类逻辑斯蒂回归建模

田笑语 张炜炜

(上海外国语大学语言研究院, 上海 200083)

摘要: 以往对于汉语分析型致使构式的研究较少关注其在语言变体维度上的变异。本研究在认知社会语言学的框架下, 基于跨变体语料库, 使用多分类逻辑斯蒂回归考察了影响致使词“使”、“令”、“让”选择的句法、语义、语言变体等因素。研究发现: 1) “令”构式用于关系从句的概率显著高于其他小句类型, 其结果谓词倾向于使用形容词, 被使者常充当感事。2) 相较于“令”, “使”和“让”都更偏好及物动词作为结果谓词、排斥充当感事的被使者。3) 当被使者为定指或充当受事时“使”的概率显著提高, 而“让”更倾向于充当施事的被使者, 当被使者为客体时“让”的概率显著降低。4) “语言变体”与“小句类型”之间存在显著的交互效应: 中国大陆地区汉语中“使”和“让”在不同小句类型中的概率差异更大; 中国台湾地区汉语和新加坡汉语中“小句类型”对“使”和“让”概率的影响较小, 且“让”的概率总是高于“使”; 而三个变体在“令”的用法上未表现出显著差异。本文展示了认知社会语言学将语言内部和外部因素整合在同一框架内进行考察的优势, 并在更多数据和参项的基础上丰富了汉语分析型致使构式与汉语变体变异的研究。

关键词: 分析型致使构式; 构式交替; 变体变异; 认知社会语言学; 多分类逻辑斯蒂回归

DOI: 10.13458/j.cnki.flatt.004679

中图分类号: H0 文献标识码: A 文章编号: 1004 - 6038(2020) 03 - 0022 - 12

1. 引言

在“基于使用”的语言观(参见 Langacker, 1987; Barlow & Kemmer, 2000) 指导下, 认知语言学者致力于对影响语言使用的认知、社会因素进行统一考察, 并有意识地发展与之相符的方法论, 由此产生了认知社会语言学(Cognitive Sociolinguistics)(参见 Kristiansen & Dirven, 2008; Geeraerts et al., 2010; Reif et al., 2013)。在认知社会语言学的框架下, 学者分别从“一形多义”(semasiological) 或“一义多形”(onomasiological) 的角度(参见 Geeraerts et al., 1994: 5) 出发考察语言的变异。前者涉及同一语言形式在不同变体(lects)^① 中的语义分布差异, 如首都名语义分布的语言变体变异(Zhang et al., 2011)、双宾构式语义的语言变体变异(Colleman, 2010) 等; 后者关注近义(near-synonymous) 词或近义构式在不同变体之间的选择差异, 如荷兰语变体中的致使构式交替变异(Levshina et al., 2013a)、英语变体中的与格构式交替变异(Bernaisch et al., 2014; Röthlisberger et al., 2017) 等。

本文拟在认知社会语言学的框架下探讨汉语分析型致使构式的交替现象, 并考察在不同的汉语语言变体(包括中国大陆地区汉语、中国台湾地区汉语和新加坡汉语) 中致使词

* 基金项目: 本文是国家社会科学基金项目“基于概率模型的转喻识别英汉对比研究”(项目编号: 17CYY004) 的阶段性成果。

作者简介: 田笑语, 硕士生, 研究方向: 认知语言学, 语料库语言学; 张炜炜, 副研究员, 博士, 博士生导师, 研究方向: 认知语言学, 语料库语言学, 语言变异与变化。通讯作者邮箱: weiwei.zhang@shisu.edu.cn



现了出来,从而探索致使构式在不同维度上的变异。由于致使构式交替研究主要关注不同语言形式的选择——因变量为分类变量,研究者也常采用(混合效应)逻辑斯蒂回归来考察研究问题(如 Speelman & Geeraerts 2009; Levshina et al. 2013a 等)。

2.2 “使”、“令”、“让”致使用法的研究

关于汉语致使构式的研究主要包括三类:1) 考察“使、令、让”由使令词到致使词^③的语法化过程(如 徐丹 2003; 张丽丽 2005, 曹晋 2011); 2) 梳理分析型致使构式与其他表达致使概念的手段(如“V得”句、“把”字句、动结式等)之间的区别和联系(如 范晓 2000; 何元建 王玲玲 2002; 项开喜 2002; 熊仲儒 2004; 周红 2004; 牛顺心 2008 等); 3) 基于语料库考察“使、令、让”等致使词在语言实际使用中的表现(如 宛新政 2004; Miyake 2005; 牛顺心, 2007; 杨江锋 2016)。这三类研究在不同程度上涉及了“使”、“令”、“让”作为致使词的用法差异,主要有以下几点发现:

(1) “使”、“令”的语法化时间早于“让”。历时研究发现,“使”在现代汉语中只能表达致使义; 尽管“令”从先秦到清代一直以使令用法为主,但在现代汉语中该用法衰落了; “让”到民国以后才开始用于表达致使义(张丽丽 2005)。基于语料库的共时调查也表明,“使”和“令”已经成为纯粹的致使词,而“让”用作致使词的频率仍低于其用作使令词的频率(牛顺心 2007)。

(2) [“令人”+结果]构式呈现词汇化的趋势,并常用于关系从句中来修饰一个名词,如“令人振奋的消息”。张丽丽(2005)指出,“令”在宋代以后就以“描述性致使”用法为主,而“常用作名词修饰语、被使者通常是泛指的人”正是“描述性致使”的重要特点。“使”、“让”后常跟复杂的结构,而“令”构式的结果谓词通常是一个“光杆词语”,与“令+NP”构成一个四字短语(宛新政 2004: 77)。在普通话中,“令”构式作谓语的能力已经萎缩,而是经常作定语,其词汇化程度比“使”、“让”更高(牛顺心 2007)。

(3) “使”、“令”、“让”倾向于搭配不同的结果谓词。杨江锋(2016)使用构式搭配分析法中的共现词素分析(Collexeme Analysis)和多项显著共现词素分析(Multiple Distinctive Collexeme Analysis)(参见 Stefanowitsch & Gries 2003)对“使”、“让”、“叫”致使构式中的结果谓词进行了辨析,结果表明,“使”更倾向于与表示结果改变的动词共现,“让”则倾向于搭配表示运动和移动的动词。而“令”主要用于“描述性致使”,其另一重要特点正是“搭配感受类谓词”(张丽丽 2005)。

(4) 三者的分布存在语体差异。如果将语体宽泛地分为书面语和口语,那么“使”、“令”更多地用于书面语中,而人们在口语中倾向于选择“让”(宛新政 2004: 83; Miyake 2005; 杨江锋 2016: 67)。杨江锋(2016: 67-74)基于兰卡斯特汉语书面语语料库(LCMC)和自建的汉语口语语料库,进一步考察了“使”和“让”在21种不同语体中的用法,发现“让”多用于以人际互动为目的的典型口语语体,而“使”主要出现于侧重信息传递的典型书面语语体。

通过上述文献回顾可以看出,针对汉语分析型致使构式的研究已经取得了一定的成果,但同时也存在一些不足。首先,研究者大多使用内省法或仅借助于零星语料,然而,已有研究表明语言学家的直觉并不总是正确的(Sampson 2001: 14-26 2005: 16-20),并且语言具



有概率性(Bod et al. 2003: 2; 桂诗春 2004) ,因此有必要基于语料库等真实的语言使用数据对分析型致使构式进行研究。其次,少数基于语料库的研究多使用简单的百分比统计方法(如宛新政 2004; 牛顺心 2007) 或仅关注结果谓词的影响(如杨江锋 2016) ,还未见有学者系统分析句法、语义等多个因素对致使词选择的共同作用,也尚未有学者探讨不同汉语变体中致使词的分布差异。鉴于此,本文将基于跨变体语料库,使用多因素分析的方法对汉语分析型致使构式进行更加深入的考察。

3. 语料来源和标注

本研究语料来源于“中文十亿字标注语料库”(Tagged Chinese Gigaword Version 2.0 Corpus)^④(Huang 2009) ,该语料库共包含八亿余汉字的新闻文本语料,收集了来自台湾“中央新闻社”(中国台湾地区)、“新华社”(中国大陆地区)和《联合早报》报纸(新加坡)的新闻报道,三个地区的文本语体一致,且所选媒体具有较高的代表性,是研究汉语变体变异的理想语料库。由于语料库数据庞大,且在地区和时间上分布不均,为了研究的可操作性,并保证语料数量在变体间的平衡,我们从每个变体中分别抽取了时间较为相近的月份(2003年9月或10月)的两百万字左右语料建立子语料库。

首先,我们从子语料库中检索出包含“使”、“令”、“让”的索引行;然后,人工清除了不属于致使的用法;接着,我们从清理后的“使”、“令”、“让”索引行中分别随机抽取了30%的目标句进行语料标注。索引行总数和目标句抽样数目如表1所示:

表1 索引行总数和目标句抽样数目

	大陆地区	台湾地区	新加坡	合计	随机抽样的目标句
使	1429	660	1294	3383	1015(30%)
让	670	1956	2601	5227	1568(30%)
令	249	351	1022	1622	487(30%)
合计	2348	2967	4917	10232	3070

在参考前人研究(牛顺心 2007,2008; Levshina 2011)的基础上,我们对随机抽取的3070个索引行进行了27个自变量的标注^⑤,主要涉及致使构式致使者、被使者和结果谓词三个槽位的句法、语义特征,整个致使构式的特征,语言变体等参项。本研究因变量为致使词的选择,包含“使”、“令”、“让”三个值。

由于自变量数目较多,且某些自变量之间存在一定的多重共线性(multicollinearity),难以直接建立回归模型。我们首先使用随机森林(random forest)方法(Breiman 2001)找出对结果有显著影响的自变量,然后再用多分类逻辑斯蒂回归模型(Fox 2008)分析这些自变量的影响方向和强度(参见Strobl et al. 2009: 341)。随机森林的结果显示,对“使”、“令”、“让”的选择具有显著影响的自变量有以下五个(变量重要性排序依次降低)(田笑语,2020)。

[1] 语言变体(Variety)

该变量表示致使构式所在的语言变体,包括三个值:中国大陆地区(ml)、中国台湾地区(tw)和新加坡(sg)。

[2] 结果谓词的句法形式(PredSynt)

该变量表示结果谓词的句法形式,涉及五个值:及物动词(trans)、不及物动词(intrans)、形容词(adj)、系动词(copula)、习语/成语(idiom)。语例如下:

(2) 这项条例将使浙江失业保险范围覆盖所有企业。(PredSynt = trans)

(3) 任何一个修改都有可能使整个方案彻底崩溃。(PredSynt = intrans)

(4) “零障碍、低成本、高效率”的理念,使这个地理意义上的弹丸小县成了“政策不足服务补”的成功实践者。(PredSynt = copula)

(5) 城市,让生活更美好。(PredSynt = adj)

(6) 他的钢琴独奏令人叹为观止。(PredSynt = idiom)

[3] 小句类型(ClauseType)

该变量表示致使构式所在的小句类型,包括六个值:简单句(simple)、复合句(cpd)、复杂句主句(main)、状语从句(advbl)、补足语从句(compl)、关系从句(rltv)。语例如下:

(7) 那一趟旅程让她爱上印度。(ClauseType = simple)

(8) 将信息化手段引入农村中小学,不仅开阔了老师的眼界,也使生活在大山深处的孩子能够较早地接触和掌握一些现代科技知识。(ClauseType = cpd)

(9) 由于围观的群众很多,使场面一度失控。(ClauseType = main)

(10) 为使这项制度能与社会脉动相结合,有关部门将广纳各界建言。(ClauseType = advbl)

(11) 最近第四台成人节目出现迷药检验剂的广告,宣称可使迷药现形。(ClauseType = compl)

(12) 该协会认为董事部并没有提出令人信服的理由。(ClauseType = rltv)

[4] 被使者的语义角色(CeRole)

该变量表示被使者在被使事件中担任的语义角色,包括五个值:施事(agent)、受事(patient)、感事(expcer)、客体(theme)、受益(害)者(befry)。语例如下:

(13) 这将使建筑业者能更灵活地应付市场新需求。(CeRole = agent)

(14) 我们希望小泉首相重视日本遗弃化学毒剂伤人事件,使其尽快得到妥善解决。(CeRole = patient)

(15) 上海合作组织的成就足以令我们感到欣慰。(CeRole = expcer)

(16) 当今的底板、胶皮和胶水已经使球速越来越快。(CeRole = theme)

(17) 宫宝仁练就的双脚“神功”还使他赢得了美满的婚姻。(CeRole = befry)

[5] 被使者的定指性(CeDef)

该变量表示被使者是否为定指,包括两个值:定指(def)、不定指(indef)。当被使者是专有名词、代词、被指示词修饰,或可根据上下文确定时,将其标记为“定指”,其他情况标记为“不定指”。语例如下:

(18) 哈马斯立即宣布将让以色列付出沉重代价。(CeDef = def)

(19) 他的成就实在令人钦佩。(CeDef = indef)

4. 数据分析和结果

4.1 多分类逻辑斯蒂回归建模

逻辑斯蒂回归(logistic regression)适用于因变量是分类变量的情况,已被广泛应用于近义词或近义构式的辨析研究(如 Speelman & Geeraerts 2009; Levshina et al. 2013a; 张懂 许家金 2019 等)。逻辑斯蒂回归的自变量可以是一个或多个分类变量或数值变量。当因变量有两个值时,涉及二分类(binary/dichotomous)逻辑斯蒂回归,简称逻辑斯蒂回归;当因变量有两个以上的值时,则需要使用多分类(multinomial/polytomous)逻辑斯蒂回归。本研究因变量涉及“使”、“令”、“让”三个值,因此采用多分类逻辑斯蒂回归。

多分类逻辑斯蒂回归主要包括两种处理方式:第一种是将因变量的参考值与其他值进行逐一比较(one vs. one);第二种则是逐一将某个值与其他所有值进行比较(one vs. rest)。这两种方式都可以在 R 中实现,前者可以使用{nnet}包(Ripley, 1996; Venables & Ripley, 2002)或{mlogit}包(McFadden, 1973, 1974; Train, 2009),后者可以用{polytomous}包(Arppe 2008; 2009)。本研究利用神经网络算法包{nnet}中的multinom函数来完成建模。

我们将数据(3070 句目标句)的 70% 分为训练集,30% 分为测试集,利用训练集建立的回归模型来预测测试集中的因变量值,以评估模型的效率。在将上述五个自变量(参见第 3 节)及其两两交互效应作为自变量加入模型后,我们利用 Anova 函数对模型进行选择,最终得出最优模型如下:

Marker ~ PredSynt + Variety + ClauseType + CeRole + CeDef + Variety: ClauseType

最终模型包含 5 个主因素的同时,还包含了“语言变体”和“小句类型”(即 Variety: ClauseType)的交互效应。该模型的预测准确率为 70.26%,远高于一直选最高频次词“让”的概率(53.47%)和随机选择任意一个致使词的概率(33.33%)。模型对于“令”、“让”、“使”的预测力 C 值分别是 0.88、0.78、0.84,可见模型具有较强的预测力^⑥。

表 2 多分类逻辑斯蒂回归模型中自变量的主效应

自变量值	回归系数(p 值)	
	让	使
Intercept	3.03(0.010) **	3.93(< 0.001) ***
PredSynt = copula	1.38(0.037) *	1.52(0.025) *
PredSynt = idiom	-0.07(0.789)	-0.11(0.750)
PredSynt = intrans	-0.19(0.401)	0.37(0.201)
PredSynt = trans	1.61(< 0.001) ***	1.65(< 0.001) ***
CeDef = indef	-0.32(0.042) *	-1.12(< 0.001) ***
CeRole = befry	0.31(0.501)	0.97(0.042) *
CeRole = expcer	-1.87(< 0.001) ***	-2.28(< 0.001) ***
CeRole = patient	0.10(0.783)	1.53(< 0.001) ***
CeRole = theme	-0.93(0.009) **	0.63(0.088)

注: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; 本研究选择的显著性水平(α level)为 0.01。

4.2 模型结果

4.2.1 自变量的主效应

自变量的主效应如表2所示^⑦。模型因变量的参考值为“令”,因此相对于“令”而言,当回归系数为正时,选择“让”或“使”的概率增加;当回归系数为负时,选择“让”或“使”的概率减少。由表2可见,当结果谓词是及物动词(PredSynt = trans)时,相较于“令”,“使”和“让”的概率均显著增加(如例20和例21)。也就是说,“令”较为排斥及物动词,而更倾向于与形容词(PredSynt = adj,即该自变量的参考值)共现(如例22)。此外,当被使者为非定指时(CeDef = indef),“使”的概率显著降低,说话人倾向于选择“令”(如例22)。

(20) 他希望欧盟能够发挥有效作用,以使中东地区摆脱目前的局面。(PredSynt = trans)

(21) 此次欧洲之行是想让高以最佳状态打好混双。(PredSynt = trans)

(22) 这将会是非常令人乐观的进展。(PredSynt = adj)

当被使者的语义角色是感事(CeRole = expcer)时,相较于“令”,“使”和“让”的概率均显著降低,说明“令”构式中的被使者更倾向于充当感事(如例22)。当被使者是受事时(CeRole = patient),“使”的概率显著提高(如例23)。当被使者是客体时(CeRole = theme)时,“让”的概率显著降低。总体来看,“让”倾向于搭配作为施事(CeRole = agent,该变量的参考值)的被使者(如例24)。

(23) 政府宣布削减雇主公积金缴交率,预期将使私人房地产的销售大受影响。(CeRole = patient)

(24) 不扩大联合国的作用,就不可能让更多的国家出钱出力,支持伊拉克的重建。(CeRole = agent)

4.2.2 自变量的交互效应

除上述主效应之外,最终模型还显示,“语言变体”(Variety)与“小句类型”(ClauseType)两个自变量之间存在显著的交互作用,我们对其进行了可视化呈现,如图1所示。图中三列分别为大陆地区汉语(ml)、新加坡汉语(sg)和中国台湾地区汉语(tw)的情况,折线图的各个节点代表“小句类型”为特定值时选择“使”(第一行)、“让”(第二行)或“令”(第三行)的概率。

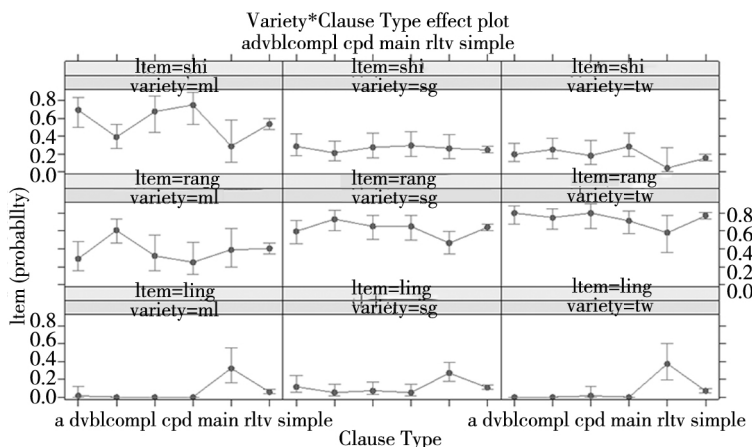


图1 语言变体与小句类型之间的交互作用

就选择“使”的概率而言(图1第一行),在大陆地区汉语中,当致使构式所在的小句为状语从句(ClauseType = advbl)(如例25)、复合句(ClauseType = cpd)(如例26)和复杂句主句(ClauseType = main)(如例27)时,其概率较高(70%左右);在补足语从句(ClauseType = compl)与关系从句(ClauseType = rltv)中,其概率较低(40%左右)。在新加坡汉语中,“使”出现在状语从句和复杂句主句中的概率最高,但仍不到40%;在补足语从句和简单句(ClauseType = simple)中“使”的概率较低,仅为20%左右。在台湾地区汉语中,“使”出现概率最高的小句类型同样是复杂句主句,但也仅占30%左右。“使”在关系从句中的概率最低(不足5%)。

(25) 为了使新闻奖的评选更加科学,从2004年起国际新闻奖将纳入中国新闻奖系列。(Variety = ml, ClauseType = advbl)

(26) 来京旅客入境手续均在北京西站出入境联检大厅办理,不仅方便了旅客出行,也使列车单程运行时间压缩了1小时18分。(Variety = ml, ClauseType = cpd)

(27) 由于对项目规模效益估计不足,使企业背上了包袱。(Variety = ml, ClauseType = main)

就“让”的使用情况而言(图1第二行),在大陆地区汉语中,当致使构式位于补足语从句中时,其概率最高(60%左右);在状语从句、复合句和复杂句主句中其概率较低(约30%)。在新加坡汉语中,“让”同样是出现在补足语从句中的概率最高(约80%)(如例28);与大陆地区汉语不同,新加坡汉语中“让”出现在关系从句中的概率最低(50%左右)。在台湾地区汉语中,“让”的概率总体较高(80%左右)且受小句类型影响较小,仅在关系从句中降低到60%左右。

(28) 培训课程的内容主要是让他们了解不同的罪案形式。(Variety = sg, ClauseType = compl)

而“令”的跨语言变异不明显,在三个变体中其概率都较低(图1第三行),仅在致使构式充当关系从句时有显著增长(如例29)。

(29) 她一连好几次为公司缔造了令人刮目的业绩。(Variety = sg, ClauseType = rltv)

总体而言,不同变体中,小句类型对于致使词选择的影响存在较大差异。在大陆地区汉语中,“使”和“让”在不同小句类型中的概率差异更为突出;而在台湾地区汉语和新加坡汉语中,小句类型对于“使”和“让”选择的影响相对较小,且在所有小句类型中“让”的概率均高于“使”。

5. 讨论

5.1 “令”、“让”、“使”的用法差异

从回归模型的结果中可以看出,“令”的结果谓词句法形式倾向于使用形容词,而排斥及物动词。及物动词的特点是需要跟宾语,这使得致使构式整体相对较长,而形容词通常是一个双音节词。“令”的这一倾向验证了宛新政(2004: 77)的观点,即“‘令’后面常跟一个光杆词语,组成一个四字短语”。此外,“令”构式中的被使者倾向于充当感事,后跟表示情绪、感

受的谓词,这一发现符合张丽丽(2005)关于“描述性致使”特点的讨论。当被使者为非定指时,“令”的概率显著提高,这可能是因为“令”更倾向于与非定指的“人”共现(牛顺心 2007; Tian & Zhang 2022)。此外,在三个变体中,“令”构式用于关系从句的概率明显高于其他小句类型(参见图1),这也证实了其具有较高的词汇化趋势(宛新政 2004; 牛顺心 2007)。

当被使者为受事时,“使”的概率显著提高;而“让”更偏好充当施事的被使者。“使”和“让”倾向于搭配不同的结果谓词(杨江锋 2016),也可能是由于被使者语义角色的影响:施事常开展一个运动或动作,而受事总是被改变。此外,使令词到致使词语法化过程的一个重要表现就是其宾语(即被使者)自主性的减弱(参见张丽丽 2005; 曹晋 2011),相较于“使”和“令”,“让”语法化开始时间更晚(张丽丽 2005),还较多地保留着使令词的用法(牛顺心, 2007),这可能也是被使者为施事时说话人倾向于用“让”的原因。

5.2 关于变体变异的讨论

在随机森林模型中,语言变体(Variety)被评估为对“使”、“令”、“让”的选择最重要的变量(参见田笑语 2020)。本研究的结果显示,“语言变体”和“小句类型”之间具有显著的交互效应,小句类型在不同变体中对于“使”、“让”概率的影响幅度不同(参见图1)。而在相同的小句类型中,大陆地区汉语选择“使”的概率总是高于中国台湾地区汉语和新加坡汉语,后两者更偏好用“让”。

前人研究表明,“使”更多地用于书面语,而在口语中说话人倾向于用“让”(宛新政, 2004; Miyake 2005; 杨江锋 2016: 67)。我们的研究对象仅限于新闻语体,因此所发现的变体变异有可能是由三地新闻语体书面化程度的差异所导致,即大陆地区新闻语体与口语存在较大差别,而中国台湾地区、新加坡汉语的新闻语体更接近其口语表达。另一种可能的解释是,“使”和“让”的用法分工出现了差异:大陆地区汉语“使”和“让”在不同小句类型中的分工更为明确;而台湾地区和新加坡汉语中“使”的用法趋于衰落,正逐渐被“让”所取代。要检验哪一种阐释符合语言的实际情况,还需要进一步的研究。

6. 结语

本研究在认知社会语言学的框架下,基于跨变体语料库,使用多分类逻辑斯蒂回归考察了影响致使词“使”、“令”、“让”选择的句法、语义、语言变体等因素。一方面,本研究在更多数据和参项的基础上为前人的一些结论提供了支持,如“令”主要用于描述性致使,“令”构式常用于关系从句、具有词汇化倾向等。另一方面,本研究揭示了被使者的语义角色和定指性对“使”、“令”、“让”选择的显著影响,以及在不同汉语变体中小句类型对“使”、“让”影响的差异。

在理论方面,本研究展示了认知社会语言学将句法、语义等语言内部因素和语言变体等外部因素整合在同一框架内进行分析的优势。在研究方法方面,多因素分析的方法能够在考虑多个参项的基础上找出影响语言变异的显著因素,从而为理论提供实证支持。

当然,本研究也存在一些不足。首先,我们仅使用了新闻语料,未考察其他语体(如口语)中致使词的交替现象。其次,尽管标注了27个参项,但仍有一些可能对致使词的选择有重要影响的参项未被纳入研究之中,如结果谓词和致使词的搭配强度。最后,本研究仅考查

了离线的语料库证据,而关于统计模型所推断出的概率能否反映语言使用者的实际偏好,还有待心理语言学实验的验证。

注释:

- ①注释“变体”(lect)泛指语言的各种变体形式,包括语言变体(language variety)、区域变体(regional dialects)、社会变体(sociolects)、个体变体(idiolects)、语域(register)等(Kristiansen & Dirven 2008: 47)。
- ②本文例句均摘自所用语料库,如为繁体字语料则将其转换为简体字。
- ③我们参考国内学者的一般做法,当“使、令、让”等词表达“派遣、命令、允许”等具体的行为意义时,称其为“使令词”;当这些词不表示行为意义,而只表示单纯的致使义时,称其为“致使词”(参见刘永耕 2000;牛顺心 2007;2008 等)。
- ④语料库网址: <https://catalog.ldc.upenn.edu/LDC2009T14>。在此感谢 Linguistic Data Consortium 委员会所提供的奖学金,使我们能使用该语料库开展本研究。
- ⑤由于篇幅有限,无法在此列出所有自变量和语例,详细描述请参见田笑语(2020)或 <https://osf.io/kx3hy/>。
- ⑥一般认为 C 值(最大值为 1)大于 0.8 代表模型具有很强的预测力(Hosmer & Lemeshow 2000: 162)。
- ⑦该表格仅列举了不涉及交互作用的主效应。由于“语言变体”(Variety)和“小句类型”(ClauseType)之间存在显著的交互作用,其主效应未列于此表,我们将在 4.2.2 节中对此交互效应进行阐释。模型的详细结果(所有效应的回归系数、p 值等)请参见 <https://osf.io/kx3hy/>。

参考文献:

- [1] Arppe A. 2008. *Univariate Bivariate and Multivariate Methods in Corpus-based Lexicography – A Study of Synonymy* [D]. Helsinki: University of Helsinki.
- [2] Arppe A. 2009. Linguistic choices vs. probabilities: How much and what can linguistic theory explain? [A]. In S. Featherston & S. Winkler (eds.). *The Fruits of Empirical Linguistics. Volume 1: Process* [C]. Berlin / New York: Mouton de Gruyter.
- [3] Barlow M. & S. Kemmer (eds.). 2000. *Usage-Based Models Of Language* [C]. Stanford: CSLI Publications.
- [4] Bernaisch T., S. Gries & J. Mukherjee. 2014. The dative alternation in South Asian English(es): Modelling predictors and predicting prototypes [J]. *English World-Wide* (1): 7–31.
- [5] Bod R., J. Hay & S. Hannedy. 2003. *Probabilistic Linguistics* [M]. Massachusetts: The MIT Press.
- [6] Breiman L. 2001. Random forests [J]. *Machine Learning* (1): 5–32.
- [7] Comrie B. 1989. *Language Universals and Linguistic Typology* (2nd Edition) [M]. Chicago: The University of Chicago Press.
- [8] Coleman T. 2010. Llectal variation in constructional semantics “Benefactive” ditransitives in Dutch [A]. In D. Geeraerts, G. Kristiansen & Y. Peirsman (eds.). *Advances in Cognitive Sociolinguistics* [C]. Berlin / New York: Mouton de Gruyter.
- [9] Dixon R. 2000. A typology of causatives form, syntax and meaning [A]. In R. Dixon & A. Aikhenvald (eds.). *Changing Valency: Case Studies in Transitivity* [C]. Cambridge: CUP.
- [10] Fox J. 2008. *Applied Regression Analysis and Generalized Linear Models* (2nd Edition) [M]. Thousand

- Oaks: Sage Publications.
- [11] Geeraerts ,D. , S. Grondelaers & P. Bakema. 1994. *The Structure of Lexical Variation: Meaning , Naming and Context* [M]. Berlin/New York: Mouton de Gruyter.
- [12] Geeraerts ,D. , G. Kristiansen & Y. Peirsman (eds.) . 2010. *Advances in Cognitive Sociolinguistics* [C]. Berlin/New York: Mouton de Gruyter.
- [13] Haiman ,J. 1985. *Natural Syntax: Iconicity and Erosion* [M]. Cambridge: CUP.
- [14] Hosmer ,D. & S. Lemeshow . 2000. *Applied Logistic Regression* [M]. New York: Wiley.
- [15] Huang ,C. 2009. *Tagged Chinese Gigaword Version 2. 0 LDC2009T14. Web Download* [Z]. Philadelphia: Linguistic Data Consortium.
- [16] Kristiansen ,G. & R. Dirven (eds.) . 2008. *Cognitive Sociolinguistics: Language Variation , Cultural Models , Social Systems* [C]. Berlin/New York: Mouton de Gruyter.
- [17] Langacker ,R. 1987. *Foundations of Cognitive Grammar. Vol. 1: Theoretical Prerequisites* [M]. Stanford: Stanford University Press.
- [18] Levshina ,N. 2011. *Doe Wat Je Niet Laten Kan: A Usage-based Analysis of Dutch Causative Constructions* [D]. Leuven: Catholic University of Leuven.
- [19] Levshina ,N. , D. Geeraerts & D. Speelman. 2013a. Towards a 3D-grammar: Interaction of linguistic and extralinguistic factors in the use of Dutch causative constructions [J]. *Journal of Pragmatics* (52) : 34 – 48.
- [20] Levshina ,N. , D. Geeraerts & D. Speelman. 2013b. Mapping constructional spaces: A contrastive analysis of English and Dutch analytic causatives [J]. *Linguistics* (4) : 825 – 854.
- [21] McFadden ,D. 1973. Conditional logit analysis of qualitative choice behaviour [A]. In P. Zarembka (ed.) . *Frontiers in Econometrics* [C]. New York: Academic Press.
- [22] McFadden ,D. 1974. The measurement of urban travel demand [J]. *Journal of Public Economics* (4) : 303 – 328.
- [23] Miyake ,T. 2005. A usage-based analysis of the causative verb *shi* in Mandarin Chinese [A]. In T. Takagaki , S. Zaiman , Y. Tsuruga , F. Moreno-Fernandez & Y. Kawaguchi (eds.) . *Corpus-based Approaches to Sentence Structures* [C]. Amsterdam/Philadelphia: John Benjamins.
- [24] Reif ,M. , J. Robinson & M. Pütz (eds.) . 2013. *Variation in Language and Language Use: Sociolinguistic , Socio-cultural and Cognitive Perspectives* [C]. Frankfurt: Peter Lang.
- [25] Ripley ,B. 1996. *Pattern Recognition and Neural Networks* [M]. Cambridge: CUP.
- [26] Röthlisberger ,M. , J. Grafmiller & B. Szmrecsanyi. 2017. Cognitive indigenization effects in the English dative alternation [J]. *Cognitive Linguistics* (4) : 673 – 710.
- [27] Sampson ,G. 2001. From central embedding to corpus linguistics [A]. In J. Thomas & M. Short (eds.) . *Using Corpora for Language Research* [C]. Beijing: Foreign Language Teaching and Research Press.
- [28] Sampson ,G. 2005. Quantifying the shift towards empirical methods [J]. *International Journal of Corpus Linguistics* (1) : 15 – 36.
- [29] Speelman ,D. & D. Geeraerts. 2009. Causes for causatives: The case of Dutch *doen* and *laten* [A]. In T. Sanders & E. Sweetser (eds.) . *Linguistics of Causality* [C]. Berlin/New York: Mouton de Gruyter.
- [30] Stefanowitsch ,A. 2001. *Constructing Causation: A Construction Grammar Approach to Analytic Causatives* [D]. Houston: Rice University.
- [31] Stefanowitsch ,A. & S. Th. Gries. 2003. Collostructions: Investigating the interaction of words and construc-



- tions [J]. *International Journal of Corpus Linguistics* (2): 209 – 243.
- [32] Strobl, C., J. Malley & G. Tutz. 2009. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests [J]. *Psychological Methods*, (4): 323 – 348.
- [33] Stukker, N. 2005. *Causality Marking Across Levels of Language Structure* [D]. Utrecht: University of Utrecht.
- [34] Tian, X. & W. Zhang. 2022. Lectal variation in Chinese analytic causative construction: Conditional inference trees and random forests [A]. In D. Tay & X. Pan (eds.). *Data Analytics for Cognitive Linguistics: Methods and Insights* [C]. Berlin /New York: Mouton de Gruyter.
- [35] Train, K. 2009. *Discrete Choice Methods with Simulation* [M]. Cambridge: CUP.
- [36] Venables, W. & B. Ripley. 2002. *Modern Applied Statistics with S* (4th Edition) [M]. Berlin: Springer.
- [37] Verhagen, A. & S. Kemmer. 1997. Interaction and causation: Causative constructions in modern standard Dutch [J]. *Journal of Pragmatics* (27): 61 – 82.
- [38] Zhang, W., D. Speelman & D. Geeraerts. 2011. Variation in the (non) metonymic capital names in Mainland Chinese and Taiwan Chinese [J]. *Metaphor and the Social World* (1): 90 – 112.
- [39] 曹晋. 2011. “使令句”从上古汉语到中古汉语的变化 [J]. *语言科学* (6): 601 – 617.
- [40] 范晓. 2000. 论致使结构 [A]. 中国语文杂志社编. *语法研究和探索(十)* [C]. 北京: 商务印书馆.
- [41] 桂诗春. 2004. 以概率为基础的语言研究 [J]. *外语教学与研究* (1): 3 – 9.
- [42] 何元建, 王玲玲. 2002. 论汉语使役句 [J]. *汉语学习* (4): 1 – 9.
- [43] 刘永耕. 2000. 使令类动词和致使词 [J]. *新疆大学学报(社科版)* (1): 93 – 96.
- [44] 牛顺心. 2007. 普通话中致使词的三个语法化阶段 [J]. *社会科学家* (3): 206 – 209.
- [45] 牛顺心. 2008. 从类型学参项看普通话中分析型致使结构的句法类型和语义表现 [J]. *语言研究* (1): 60 – 68.
- [46] 田笑语. 2020. 汉语分析型致使构式中“使”、“令”、“让”的交替——跨变体的视角 [D]. 上海: 上海外国语大学.
- [47] 宛新政. 2004. 现代汉语致使句研究 [D]. 上海: 复旦大学.
- [48] 项开喜. 2002. 汉语的双施力构式 [J]. *语言研究* (2): 70 – 77.
- [49] 熊仲儒. 2004. 现代汉语中的致使句式 [M]. 合肥: 安徽大学出版社.
- [50] 徐丹. 2003. “使”字句的演变——兼谈“使”的语法化 [A]. 吴福祥编. *语法化和语法研究(一)* [C]. 北京: 商务印书馆.
- [51] 杨江锋. 2016. 汉语迂回致使结构的多维度研究 [D]. 杭州: 浙江大学.
- [52] 张懂, 许家金. 2019. 英汉与格交替现象的多因素研究 [J]. *外国语* (2): 24 – 33.
- [53] 张丽丽. 2005. 从使役到致使 [J]. *台大文史哲学报* (62): 119 – 152.
- [54] 张炜炜, 刘念. 2016. 认知语言学定量研究的几种新方法 [J]. *外国语* (1): 71 – 79.
- [55] 周红. 2004. 现代汉语致使范畴研究 [D]. 上海: 华东师范大学.

ABSTRACTS

Multifactorial Contextual Co-selection: The Recent Development of Corpus Linguistics ,p.

1. *XU Jiajin*

Probabilistically-motivated multifactorial contextual co-selection has figured prominently in the field of corpus linguistics lately. The approach aims to address form-meaning correspondence of language namely polysemy/semasiology and synonymy/onomasiology. Lexicography and sociolinguistics have been two initial sites of multifactorial analysis, which has recently been extended to cognitive linguistics, focusing on the study of the distinction of multiple senses of a word, delineation of synonymous words and constructions, the optionality of grammatical markers and word order. Multifactorial analysis is a language description technique on the one hand, and characterized by its strong explanatory power on the other.

Key Words: multifactorial analysis; contextual co-selection; form-meaning correspondence; multivariate statistics; corpus linguistics

Direct/Indirect Causation of English Analytic Permissives: A Multinomial Logistic Regression and Multiple Correspondence Analysis ,p. 11. *HUANG Ying & REN Wei*

Due to the scarcity of multifactorial analysis of direct/indirect causation on the part of the analytic permissives, the present study aims to explore direct/indirect causation and conceptual space of three analytic permissives *let*, *allow* and *permit* using a multinomial logistic regression model and multiple correspondence analysis respectively. The results show that, contrasted with the typical view that analytic constructions tend to encode indirect causation, all the three analytic permissives can denote direct or indirect causation, building up “causative”, “no intervention”, “permissive” and “enabling” conceptual spaces. The results also indicate that there exist multiple correspondences between direct or indirect causation and mapped conceptual spaces. The permissives are subject to statistical preemption under the contextual constraints of multiple factors. The study sheds some light on the correspondence between direct/indirect causation of the analytic permissives and the more fine-grained conceptual space thereof. It enables English learners to grasp the semantic nuances of the permissives as well.

Key Words: analytic permissives; direct/indirect causation; conceptual space; statistical preemption

Chinese Analytic Causative Constructions and Their Lectal Variation: A Multinomial Logistic Regression ,p. 22. *TIAN Xiaoyu & ZHANG Weiwei*

Lectal variation in Chinese analytic causative constructions has not yet been investigated in detail. In the framework of Cognitive Sociolinguistics, this paper employs a multinomial logistic regression model to examine the syntactic, semantic and lectal factors that may affect the alternation of Chinese analytic causative constructions with *shi*, *rang* and *ling* based on regiolectally-balanced corpora. Our analyses show that: 1) the causee of *ling*-constructions tends to be an experiencer, followed by adjectives, and *ling*-constructions are significantly more likely to appear in a relative clause. 2) Compared with *ling*, both *shi* and *rang* are more favored when the effected predicate is a transitive verb and more disfavored than *ling* when the causee is an experiencer. 3) The probability of *shi* significantly increases when the causee is definite or serves as a patient, whereas *rang* favors the causee whose semantic role is an agent; The probability of *rang* significantly decreases when the causee is a theme. 4) There is a significant interaction between Variety and ClauseType, i. e. the influence of ClauseType is strongest in Mainland Chinese, with *shi* and *rang* preferring different clause types, whereas in Taiwan Chinese and Singapore Chinese, *rang* is always favored. The usage of *ling* shows no significant cross-variety variation. This study



not only shows the benefit of Cognitive Sociolinguistics for synthetically analyzing language internal and external factors but also contributes to the research on Chinese analytic causative constructions and variation in Chinese varieties by analyzing large-scale data and incorporating considerable predictors.

Key Words: analytic causative constructions; construction alternation; lectal variation; Cognitive Sociolinguistics; multinomial logistic regression

A Study of Relativizer Omission by Chinese EFL Learners: Triangulating Corpus and Experimental Approaches p. 34. *FANG Yinjie & LIANG Maocheng*

Though the predictability of relativizer omission has been born out in a variety of corpus studies scant attention in verifying its psychological reality has been paid to the combination of corpus and experimental approaches. Targeting Chinese EFL learners' relativizer omission this article attempts to triangulate the findings of corpus and acceptability judgment with binary logistic regression and linear mixed-effects regression applied respectively. The study reveals that , through the comparison of two regression models a) corpus and acceptability approaches both display a converging prediction of relativizer omission; b) Chinese EFL learners have acquired prototype knowledge with regard to the use of relativizer omission; c) discrepancy between the two models may be contingent on the different factors restricting language production/comprehension process. The above findings reinforce the applicability of usage-based constructionist approach in SLA studies.

Key Words: relativizer omission; corpus; acceptability judgment; regression analysis

A Multifactorial Analysis of Subject Anaphor Realization in Translated Chinese: The Conditional Inference Tree Approach p. 44. *XU Xiuling*

Textual features of translated texts are rarely dealt with in corpus-based translation studies. The present study attempts to examine the mechanism of subject anaphor realization (zero vs. non-zero) in translated Chinese. A sizable number of anaphoric instances were collected from two corpora of native and translated Chinese and were annotated for multiple contextual factors. The annotated data were then subjected to multifactorial analysis using the conditional inference tree approach. The results show that the influence of potential interference and sentential referential distance on subject anaphor realization is quite similar in the two corpora. However ,clausal referential distance and the syntactic position of the antecedent exert varying effects ,which can be reflected by the more frequent use of non-zero anaphors in different types of topic chains in translated Chinese ,indicating intralingual cohesive explicitness. The high incidence of non-zero subject anaphors in translated Chinese can be plausibly explained by "SL shining through" and translators' risk aversion. The present study is an endeavor towards fine-grained and contextualized analysis in corpus-based translation studies.

Key Words: subject anaphor; cohesive explicitness; translated Chinese; multifactorial analysis; conditional inference tree

The Use of V-N Collocations in L2 Oral Production p. 54. *WANG Wenyu & ZHOU Dandan*

Collocations are believed to be conducive to the fluency and idiomaticity of the L2 production. In order to further understand the features and the role of the collocation use in Chinese L2 learners' oral production this study analyzed the use of V-N collocations by 73 fourth-year English majors in a speaking task. Data analysis has yielded the following findings: 1) Among the various types of V-N collocations ,the most frequently used was the VO structure followed by the VOPO and VPO structures; among the ten most frequently-used verbs that occurred in the V-N collocations ,a majority were delexicalized verbs. 2) Nearly 30% of all the V-N collocations were deviant and such deviations lay far more often in verbs and nouns than prepositions and deter-