

构式抽象层与具体层的对应性研究

——以关系代词取舍为例

房印杰

(北京邮电大学人文学院,北京 100876)

摘要:构式抽象层与具体层的对应性是构式语法的核心假设之一,但对其开展的实证检验不足。以英语本族语者的关系代词取舍为研究对象,通过二元逻辑斯蒂回归分析,从抽象层与具体层两方面对该构式的原型性、概率性、抽象度进行交叉验证。研究发现,关系从句长度、从句主语词性以及先行词在主句中所充当的句法成分是关系代词取舍构式的原型性特征;隐性关系代词从句呈现迥异的概率性分布,显性关系代词从句的预测效果弱于隐性关系代词从句;两种关系从句在抽象度层面存在差异,隐性关系代词从句更加靠近具体层面。

关键词:关系代词取舍;对应性;二元逻辑斯蒂回归;原型性

中图分类号:H314.3

文献标志码:A

文章编号:1674-6414(2025)04-0025-13

0 引言

构式抽象层—具体层的对应性是构式语法(Goldberg, 1995, 2006)的核心假设之一。在构式语法体系下,构式是语言学习的基本单位,语言结构从具体的语言使用中浮现而出;语言知识由一系列处于不同抽象程度的构式组成,语言的认知处于不断的动态变化之中。构式抽象层—具体层的对应性阐述表明,学习语言的过程涉及掌握一系列处于不同抽象度层面的构式,这包括抽象的构式图式(schema)与实际语言使用中的具体构式样例(exemplar)(Deshors, 2014:246)。基于“累积效应”(董保华等, 2025:37),具体构式样例逐步被抽象与概括,从而向抽象构式图式演变。该假设要求基于构式语法的实证研究应当着力论证具体构式向抽象构式的抽象、概括、范畴化过程。然而,前人研究多以抽象构式图式或具体构式样例为立足点,分析构式属性,对二者的交叉验证并不多见。鉴于此,本文尝试以英语本族语者的关系代词取舍(例1)为研究对象,基于二元逻辑斯蒂回归模型与模型实际预测值,从抽象特征层与具体样例层对其开展构式原型性、概率性、抽象度的交叉验证,

收稿日期:2024-11-25

基金项目:北京市社会科学基金项目“基于统计模型的中国学习者英文体特征研究”(20YYB006)的阶段性成果

作者简介:房印杰,男,北京邮电大学人文学院教授,博士,硕士生导师,主要从事语料库语言学与二语习得研究。

引用格式:房印杰. 构式抽象层与具体层的对应性研究——以关系代词取舍为例[J]. 外国语文, 2025(4):25-37.

进而论证具体构式样例与抽象构式图式在该构式属性中的对应性。

(1) a. The moral *dilemmas it causes* are contentious issues. (LOCNESS)

b. Every *law which the people have not ratified* is null. (NESSIE)

1 文献回顾

1.1 构式的抽象层与具体层

以阿黛尔·戈德堡(Adele Goldberg)(1995, 2006)为代表的构式语法提出,语言学习的本质在于学习大量处于不同复杂度、抽象度的构式。作为语言学习的基本单位,构式包含形式—意义的共现信息,该信息在一定程度上表征语言单位在学习者头脑中的认知固化程度(entrenchment)。戈德堡(2006)直言,语言学习是一个对语言输入不断归纳与抽象,最终获取语言结构的适度能产性的过程。换言之,构式网络中存在一个具体构式—半抽象构式—抽象构式的连续统(Ellis, 2003),语言学习是基于具体构式(item-based construction)向复杂、抽象构式的不断过渡(Tomasello, 2003)。概言之,构式连续统中存在抽象层与具体层的对应性,该对应性在语言学习与使用中发挥关键作用。

虽然构式语法对构式抽象层—具体层的对应性关系做了相应阐述(Ellis, 2003; Deshors, 2014),但仍然缺乏实证研究的支撑。纵观基于构式语法的实证研究不难发现,该类研究主要分为两大类:探究构式的原型性抽象特征,检验具体构式与其内部成分兼容性。在抽象特征层面,斯蒂芬·格莱斯(Stefan Gries)(2003)以线性分类器(linear discriminant analysis)模型提取了英文双宾语构式的原型性特征,托比亚斯·贝尔奈施等(Tobias Bernaisch et al.)(2014)以决策树和随机森林方法分析了南亚英文中GIVE一词所引导的双宾语构式与介词宾语构式的使用特征,并基于该复杂统计模型提取出了两种构式各自的原型性抽象特征集合。在具体构式层面,构式语法学家主要借助搭配构式分析法,确定具体构式的潜在原型(Gilquin, 2015; Schönefeld, 2015)。上述研究各自从抽象层或具体层出发探究构式的原型性特征,但忽视了对两个层面开展同步分析,从而对构式的原型性进行交叉验证。针对上述不足,达格玛·迪维亚克(Dagmar Divjak)等(2013)与桑德拉·德肖尔(Sandra Deshors)(2014)分别尝试对构式抽象层与具体样例层开展共时同步分析,前者以芬兰语中的THINK类近义词与俄语中的TRY类近义词为研究对象,尝试基于具体构式样例论证其向抽象构式的抽象、概括、范畴化过程;后者考察二语学习者的四种情态动词构式(may, can, must, will)在抽象层—具体层的对应性。上述两个研究以复杂统计模型对构式抽象层与具体层的同步考察是对二者理论对应性的探索性实证检验,但它们均以词为研究对象,并未涉及更为复杂的词组与句法构式。而句法构式的复杂度与抽象度显著高于近义词和情态动词所代表的词语构式(Goldberg, 1995: 5),它们所触发的认知加工负载也更

大。同时,迪维亚克等(2013:269)指出,目前只是通过复杂统计模型论证了词语构式中存在抽象层与具体层的对应性,后续研究应着力论证该对应性在更为复杂、抽象构式中的普适性。因此,有必要以句法构式为研究对象,探究该构式中抽象层与具体层的潜在对应性。

鉴于此,本文选择关系代词取舍这一高度复杂的句法现象为研究对象,尝试基于大规模语料,以二元逻辑斯蒂回归模型从抽象层与具体层对构式的原型性、概率性、抽象度开展交叉验证。

1.2 关系代词取舍研究

关系代词取舍研究受到学界的持续关注,在理论层面,针对关系代词取舍的制约机制,学界先后提出了四种理论假设:紧凑句假设(Jespersen, 1933: 359-361)、从句扩展假设(Diessel, 2004)、单句性倾向假设(Fox et al., 2007)和关系从句可预测性假设(Wasow et al., 2011)。在研究焦点层面,前人研究以单个因素分析为主,如依存成分的长度(Gibson, 1998)、不同句法成分在认知层面的可及性(Ariel, 1990)、语义模糊(Temperley, 2003)、先行名词和关系从句主语的有生命或无生命(Mak et al., 2002)等。伴随大规模语料和复杂统计模型的应用,部分研究开始同步考察多个因素对关系代词取舍的潜在影响(Diessel, 2004; Wasow et al., 2011),但这些研究所分析的语言学特征数量相对有限,由此建立的多因素统计模型无法全面反映关系代词取舍的使用机制。

前人研究对关系代词取舍的制约机制做了相关探讨,但仍存在四点不足:(1)研究多聚焦某一个或几个因素,对其他潜在影响因素未能考虑,而语言现象受到多种因素的共同作用(Gries, 2013),仅仅考察单个因素所得出的研究结论解释力不足;(2)在关系代词取舍研究中先后提出的四种理论假设仅仅考察某几个因素,涵盖范围有限;(3)基于构式语法的相关研究仍不充分,虽然芭芭拉·福克斯(Barbara Fox)等(2007)、托马斯·瓦索(Thomas Wasow)等(2011)和丹尼尔·维克曼(Daniel Wiechmann)(2015)均借鉴构式语法框架,对关系代词取舍的原型性做了相应探讨,但多聚焦先行词或从句主语等部分因素;(4)缺乏从抽象层与具体层对关系代词取舍构式的交叉验证,房印杰(2017)从具体样例层面抽取了隐性关系代词从句的潜在原型,瓦索等(2011)聚焦关系代词取舍的抽象特征,但还未见到从抽象层与具体层对关系代词取舍构式属性的交叉验证。针对上述不足,本文以英语本族语者的关系代词取舍为研究对象,通过二元逻辑斯蒂回归模型开展多因素分析,从抽象特征层与具体样例层两个方面探究该构式的原型性、概率性与抽象度。

2 研究设计

2.1 研究问题

通过对关系代词取舍构式的抽象层与具体层的同步分析,本文尝试回答三个问题:

(1) 如何从抽象层与具体层测定关系代词取舍构式的原型性? (2) 如何确定关系代词取舍构式的概率性分布? (3) 如何界定隐性与显性关系代词从句的抽象度?

2.2 语料抽取与标注

本文以英语本族语者为研究对象,选取 LOCNESS(形符数:326 092),NESSIE^①(形符数:291 509)两个本族语者议论文语料库,与 ICNALE^②语料库中本族语者的议论文部分(形符数:91 013)共同组成英语本族语者语料集,合计库容 708 614 个形符。在获得三个语料库后,以 PowerGREP 对其进行文本整理,使用 Stanford Parser 进行句法剖析,然后采用 Tregex 树库检索软件,编写树型正则表达式(NP<SBAR)抽取非主语型限定性关系从句,再经过人工判别,最终抽取出 672 句合格语句。表 1 显示,隐性关系代词从句的使用频次显著高于显性关系代词从句的使用频次。这种基于原始频次的对比初步显示,隐性与显性关系代词从句的使用并非随机,而是呈现出特征性分布,隐性关系代词从句数量均高于显性关系代词从句。

表 1 非主语型限定性关系从句在语料库中的分布

语料库	显性关系代词	隐性关系代词	合计
LOCNESS	133	253	386
NESSIE	86	123	209
ICNALE. ns	36	41	77
合计	255	417	672

在抽取出 672 个非主语型限定性关系从句语料后,将全部语料按照表 2 中设定的特征集进行人工标注。该特征集在参照前人研究发现的基础上归纳而成。其中,先行词与主语的词性参照可及性假设(Sperber et al., 1986)和词语信息已知性等级(Gundel et al., 1993)。生命性特征参照马修·特拉克斯勒(Matthew Traxler)等(2005),具体性参照约瑟夫·塔格利希特(Josef Taglicht)(1972),语义丰富度参照福克斯等(2007),确定性参照瓦索等(2011),先行词在主句与从句中的句法成分分类参照霍尔格·迪塞尔(Holger Diessel)(2004),主语与从句的时态参照伦道夫·夸克(Randolph Quirk)等(1985),语态参照维克曼(2015),句子的肯定或否定形式参照戈德堡(2006);句子长度参照珍妮弗·阿诺德(Jennifer Arnold)等(2000),以词为计算单位;谓语类型与长度参照福克斯等(2007),以词为计算单位。

① NESSIE 语料库由北京外国语大学许家金教授主持建立。
② ICNALE 语料库由日本神户大学 Dr. Shin Ishikawa 主持建立。

表2 特征标注集

变量类型	变量
先行词	生命性、具体性、语义丰富度、确定性、修饰语数量、先行词在主句中成分、先行词在从句中成分
从句主语	生命性、具体性、语义丰富度、确定性、词性
主句	时态、语态、谓语、否定、句子长度、谓语动词长度
从句	时态、语态、谓语、否定、句子长度、谓语动词长度

2.3 统计分析

完成语料抽取、标注后,对其开展宏观与微观两个层面的统计分析,两个层面分别对应构式的抽象层与具体层,从而论证其潜在对应性。在宏观层面,关系从句标注特征集将被导入二元逻辑斯蒂回归模型,检验一系列抽象特征对隐性与显性关系代词从句的潜在预测权重,从而提取显著预测因素,根据贝尔奈施等(2014),最终获得的显著预测因素共同构成该构式的原型性属性;在微观层面,我们将依据最终回归模型计算其对每个关系从句的预测值,将该预测值与隐性和显性关系从句的实际选择值进行对比,从而确定回归模型对整体语料的正确预测率,同时依据正确预测值由高到低排序,测定模型分别对隐性和显性关系从句的预测趋势,并提取出具有高度预测性(预测值>0.9)的隐性和显性关系代词从句开展微观分析。通过对比该类关系从句所包含的语言学特征集和与模型反馈的原型性属性,从而确定二者的潜在对应性,进而论证抽象构式由具体构式概括、抽象而出的论断(Goldberg,2006)。

通过上述研究路径来探究构式的抽象层与具体层的对应性,这样做的理据在于:(1)在理论层面,二元逻辑斯蒂回归模型的构建、拟合过程与构式语法所宣称的构式抽象度连续统论断高度一致。该统计模型是基于每一个具体关系从句的语言学特征集来构建、拟合,并最终确定对隐性和显性关系从句具有显著预测倾向的语言学特征集,模型构建、拟合过程一定程度上反映了由具体构式样例向抽象构式图式的概括、范畴化过程。统计模型会对每一个具体关系从句赋予一个预测值,基于正确的预测值由高到低,可以对所有关系从句排序,该序列表征具体构式样例的原型性特征在逐步减弱。(2)在研究路径层面,前人研究已初步论证了其可行性(Divjak et al.,2013;Deshors,2014),根据蒂莫西·弗尔贝门(Timothy Verbeemen)等(2007),预测值最高的关系从句代表该构式的具体原型样例,该类语句自身的语言学特征表征具体原型的特征属性,通过对比具体原型样例的特征属性与统计模型所归纳的抽象构式原型属性,可以较为客观地分析二者的对应性。

3 研究发现

3.1 构式抽象层

基于语言学特征集的回归模型显示,该模型能够较好地预测隐性和显性关系代词从

句。回归模型 R^2 为 0.314,卡方值 176.92,自由度 11, p 值 <0.0001 ,对隐性和显性关系代词从句的理论预测值(C)为 0.786。回归模型正确预测隐性关系代词从句 357 句,显性关系代词从句 139 句,对语料的实际正确预测率为 0.738^①,该预测值显著优于随机预测($p<0.0001$)。

表 3 回归模型中显著变量

变量	自由度(df)	对数似然值检验(LRT)	显著性($p^{\text{②}}$)
RC_length	1	74.400	$<2.2\text{e-}16^{***}$
Head_ani	1	9.302	$0.002\,289\,6^{***}$
Head_cont	1	4.058	$0.043\,966\,4^*$
Head_def	1	9.664	$0.001\,878\,8^{***}$
RC_subj_conc	1	3.173	$0.074\,859\,5^{\text{ms}}$
RC_subj_POS	1	41.348	$1.274\text{e-}10^{***}$
Head_ext	4	20.382	$0.000\,419\,8^{***}$
RC_negation	1	5.714	$0.016\,831\,4^*$

回归模型最终抽取出八个对隐性和显性关系代词从句具有显著影响的因素(表3):关系从句长度(RC_length)、先行词有生命或无生命(Head_ani)、先行词语义丰富度(Head_cont)、先行词确定性(Head_def)、从句主语具体性(RC_subj_conc)、从句主语词性(RC_subj_POS)、先行词在主句中充当的句法成分(Head_ext)和从句肯定或否定形式(RC_negation)。上述抽象特征在隐性和显性关系代词从句预测中发挥显著权重,这说明,隐性和显性关系代词从句在抽象层面具有不同分布特征。具体而言,对显性关系代词从句具有强烈预测倾向的语言学因素包括:关系从句长度较大,先行词为有生命词语、较为确定、具体,先行词在主句中充当介词宾语、主语,从句主语为名词,关系从句为否定结构。对隐性关系代词从句具有强烈预测倾向的语言学因素则包括:关系从句长度较短,先行词为无生命词语、不确定、不具体,如 something;先行词在主句中充当系词补语,从句主语为代词,关系从句为肯定结构。上述抽象特征具有不同的预测权重,从句长度、从句主语词性与先行词在主句中充当的句法成分三个特征高度显著。

回归模型的高度预测性充分说明,隐性和显性关系代词从句在抽象层面具有迥异的分布特征;根据贝尔奈施等(2014)研究,上述具有显著预测性的语言学因素(表3)共同构成了该构式的原型性属性。需要注意的是,构成原型性属性的多个语言学因素具有不同的预测权重。

① 该预测值计算流程:(357+139)/672=0.738。
② *代表 $p<0.05$,**代表 $p<0.01$;***代表 $p<0.001$;ms表示显著性处于临界值。

3.2 构式具体层

在获取对隐性和显性关系代词从句具有显著预测权重的抽象特征后,本文尝试基于统计模型对各个关系从句的预测值,从具体层面探究两种关系代词从句的内在差异。根据回归模型对两种关系从句的正确预测值,分别生成其预测曲线(图1)。

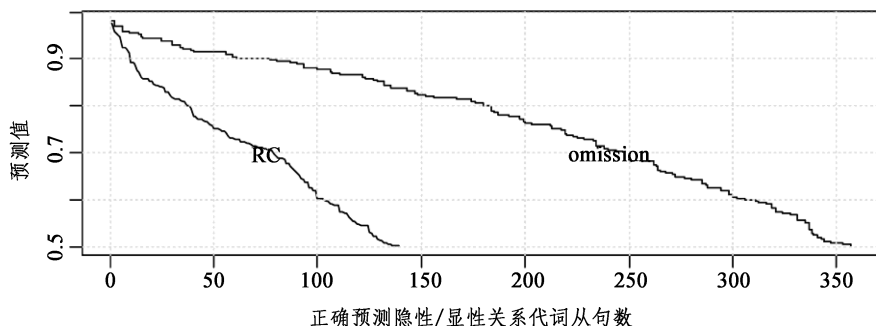


图1 隐性关系代词从句预测趋势^①

图1显示,(1)回归模型对隐性和显性关系代词从句的预测趋势截然不同(表4)。对隐性关系代词从句的预测趋势更加平缓,说明在相同预测区间内(如0.9~1.0)存在更大数量关系从句,相邻从句间的预测值差异较小,相对易于从中归纳出隐性关系代词从句的语言学特征;显性关系代词从句的预测曲线更加陡峭,说明相邻从句之间的预测值差异较大,难以从中归纳该类从句的抽象特征。(2)回归模型对两种关系从句的正确预测值均可达到0.9以上,这种高度准确的预测性指向该类型从句的原型性特征(Bernaisch et al., 2014:25)。

表4 各预测值区间内隐性关系代词从句分布

	0.5~0.6	0.6~0.7	0.7~0.8	0.8~0.9	>0.9	合计
RC	37	24	40	29	9	139
omission	49	62	67	103	76	357

3.2.1 隐性关系代词从句

基于统计模型提取具有高度预测倾向(预测值>0.9)的隐性关系代词从句,根据贝尔奈施等(2014),该类语句可以视为隐性关系代词从句的原型性样例。分析上述语句的语言学特征集(表5),可以发现该构式的原型性属性部分映射到了原型性样例中。具体而言,首先,关系从句长度对隐性关系代词从句具有高度预测性,在预测值>0.9的76个语句中,关系从句长度为3的从句有25个,关系从句长度为4的从句有5个,关系从句长度为2的从句则有46个(例1),在预测值前20句语料中,关系从句长度为2的语句更是占19句。关系从句长度较短与隐性关系代词从句的高度相关性符合回归模型(表3)中关系从句长

^① RC表示显性关系代词从句,“omission”表示隐性关系代词从句。

度的高度显著性。其次,从句主语词性与隐性关系代词从句高度相关,在 76 句语料中,从句主语词性位置全部为代词充当,代词进一步分为人称代词(例 2)和非人称代词(1)。最后,先行词在主句中充当的句法成分较为多样,包括主语(例 1)、宾语(例 2 与例 4)、系词补语(例 3)、介词宾语,但没有孤立名词。这三个语言学因素在构式原型性属性中最具预测性,即代表性。

表 5 具有高度预测倾向的隐性关系代词从句的语言学特征组合^①

从句长度	先行词生命度	先行词语义丰富度	先行词确定性	从句主语具体性	从句主语词性	先行词的句法成分	从句肯定/否定	预测值
2	无生命	高	确定	抽象	代词	主语	肯定	0.98
2	无生命	低	确定	具体	代词	宾语	肯定	0.97
2	无生命	低	确定	具体	代词	宾语	肯定	0.97
2	无生命	低	确定	具体	代词	宾语	肯定	0.97
2	无生命	低	确定	具体	代词	宾语	肯定	0.97
2	无生命	高	确定	具体	代词	系词补语	肯定	0.96
2	无生命	高	确定	具体	代词	系词补语	肯定	0.96
2	无生命	高	确定	具体	代词	系词补语	肯定	0.96
2	无生命	低	确定	具体	代词	介词宾语	肯定	0.96

(2) Some people rely on sport to provide the *exercise they need* but some sports carry a risk of injury. (NESSIE)

(3) Giving governments more control and owners less control is the main *concern I have*. (ICNALE)

(4) They get everything they need. (LOCNESS)

3.2.2 显性关系代词从句

基于统计模型提取具有高度预测倾向(预测值>0.9)的显性关系代词从句,分析其语言学特征的共现组合,可以发现对显性关系代词从句具有高度预测性的语言学特征包括较大的关系从句长度、由名词充当的从句主语、先行词在主句中主要充当孤立名词(例 5)和主语(例 6),它们共同构成了原型性样例的特征属性,这与统计模型所归纳的三个最显著的原型性特征高度一致。

① 由于显性关系代词从句中只有九句语料的预测值>0.9,因此表 5 与表 6 只列举预测值排名前九位的语句。

表 6 具有高度预测倾向的显性关系代词从句的语言学特征组合

从句长度	先行词 生命度	先行词语义 丰富度	先行词 确定性	从句主语 具体性	从句主语 词性	先行词的 句法成分	从句肯定/ 否定	预测值
9	无生命	低	不确定	具体	名词	孤立名词	肯定	0.98
9	无生命	低	确定	具体	名词	孤立名词	肯定	0.96
14	无生命	高	不确定	具体	名词	主语	肯定	0.96
13	无生命	高	不确定	具体	名词	主语	肯定	0.95
9	有生命	低	不确定	具体	名词	主语	肯定	0.95
10	无生命	高	不确定	具体	名词	主语	肯定	0.92
10	无生命	高	不确定	具体	名词	主语	肯定	0.92
6	无生命	高	不确定	具体	名词	主语	否定	0.92
7	无生命	低	不确定	具体	名词	主语	否定	0.91

(5) Secondly, due to the *confessions which* Hugo make at the end of the play...we are forced to ask whether this was purely an act on the part of Hugo in trying to validate himself, to give himself weight and to re define his project in life. (LOCNESS)

(6) A main *claim that the supporters of prayer* in public schools point to is the statistic that, on a regular basis, polls show that 75% or more of Americans favor the return of prayer to schools. (LOCNESS)

两种关系从句在三个语言学特征上的分布特征截然相反,这使得我们得以从具体样例层面将二者客观分离。需要注意的是,具体构式样例所包含的语言学特征集(从句长度、从句主语词性、先行词在主句中充当的句法成分)表征该构式抽象属性中最具原型性的部分。对抽象层面的其他属性,则缺乏足够的对应性。这一发现与弗尔贝门等(2007)对构式抽象层与具体层属性关系的论断一致。

4 讨论

本节基于统计分析结果,尝试从抽象特征层与具体样例层论证关系代词取舍构式的原型性、概率性分布和抽象度。

4.1 关系代词取舍构式的原型性

尼克·埃利斯(Nick Ellis)和费尔南多·费雷拉-儒尼奥尔(Fernando Ferreira-Junior)(2009:190)将构式原型(construction prototype)界定为某一构式中的最典型具体构式,在该实例周围聚集众多非典型性具体构式。典型性与非典型性构式的产生则源自多种语言学因素在构式产出中的竞争。依据竞争原则(MacWhinney,2008),语言产出过程涉及一系列因素的相互竞争,并最终决定语言产出中所采用的具体语言形式。已有研究(Ellis et al.,

2014a)显示,竞争因素涉及语言输入频次、语言复杂度、原型性等。通过基于复杂统计模型的多因素分析,可以抽取出其中的显著因素并确定其对语言形式的客观预测倾向。对某一语言形式具有较高预测信度和效度的显著因素集则被界定为该语言形式的原型性(Bernaisch et al., 2014:25),该原型性所高度预测的具体语言形式则被称为具体层构式原型(instantiation of prototypical construction)。

针对隐性和显性关系代词从句的统计分析表明,从句长度、从句主语词性和先行词在主句中充当的句法成分可以被认定为隐性和显性关系代词从句的原型性因素。具体而言,从句长度较长、名词充当从句主语、先行词在主句中充当孤立名词和主语,这三种语言学特征对显性关系代词从句具有显著预测倾向,且该倾向在具体构式分析(表6)中得到验证,因此显性关系代词构式的原型性可以概括为:

(1)先行词^[孤立名词/主语]+RC+从句主语^[名词]+较长剩余部分

与显性关系代词从句的预测倾向截然不同,隐性关系代词从句的显著预测因素为,关系从句较短(2~4个单词),代词充当从句主语,先行词在主句中充当宾语、系词补语、介词补语、主语,因此隐性关系代词从句的原型可以概括为:

(2)先行词^{[名词][宾语/系词补语/介词宾语/主语]}+omission+从句主语^[代词]+谓语动词

(3)先行词^{[不定代词][宾语/系词补语]}+omission+从句主语^[代词]+谓语动词

该原型基于先行词的词性进一步分为两种:当名词充当先行词时,其能够充当多种句法成分,而当不定代词充当先行词时,其主要充当宾语和系词补语两种句法成分,形成“*That is all I want*”从句结构。对上述两种隐性不定代词从句原型性的探究拓展了将“不定代词+omission+从句主语^[代词]”作为该结构唯一原型(房印杰,2017)的不足。

4.2 关系代词取舍构式的概率性分布特征

关系代词取舍构式的概率性特征与其原型性紧密相关,埃利斯等(2014b)提出,构式学习实质上是一种范畴学习(category learning),构式学习过程中伴随对具体语言范例的不断概括与抽象(Goldberg, 2006),语言学习的过程在本质上是掌握抽象语言结构的适度能产性(Ambridge, 2013:508-509)。基于上述观点,我们可以假定,构式的原型性自身便已经暗示了由原型性构式与非原型性构式所组成的构式概率分布连续统。在该连续统中,构式伴随原型性的不断降低,呈现概率性分布,或称为齐波夫分布(Zipfan distribution)。对关系代词取舍构式的统计分析为该假设提供了较为客观的证据。具体而言,统计模型对关系代词取舍构式的预测值呈现逐步下降趋势(图1),而且对该构式具体样例层的共现语言学特征分析显示,在高度显著的预测区间内,语言学特征较为稳定,伴随预测值逐步降低,语言学特征开始发生变化,如关系从句长度伴随隐性关系代词从句预测值下降,逐步增大。此外,在抽象特征层面,回归模型所提取的显著制约因素(关系从句长度、从句主语词性、先行

词在主句中充当的句法成分)也对关系代词取舍构式做出概率性预测。概言之,关系代词取舍构式的概率性分布与其原型性特征共同存在,二者均可以从抽象特征层与具体样例层加以佐证。

4.3 关系代词取舍构式的抽象度

虽然隐性与显性关系代词从句均具有抽象特征与具体语料的特征分布,但二者是否具有相同的抽象度?构式语法提出,抽象语言知识源自对具体语言范例的不断概括,从而形成了一个由具体构式—半抽象构式—抽象构式组成的构式连续统(Ellis, 2003)。换言之,隐性关系代词从句均存在一个抽象度连续统,而二者在各自连续统中处于何种位置?针对这一问题的讨论,并不多见。对两类关系从句的抽象层与具体样例层的统计分析显示,在抽象度连续统中,隐性关系代词从句应低于显性关系代词从句。这是由于隐性关系代词从句的长度较短,所对应的认知加工负载较小(Rohdenburg, 1996),该从句类型数较少,能产性较低(Barðdal, 2008),隐性关系代词从句在先行词与从句主语位置上的词语类型显著低于显性关系代词从句(房印杰, 2017)。这种词语类型的低能产性易于形成语言学习中的认知固化(entrenchment),从中概括抽象语言规律,同时不定代词与人称代词均属于封闭词类;词语类型有限,认知加工负载远低于名词(Ariel, 1990),从而使得语言使用者易于抽取该类型从句的抽象语言特征。概言之,隐性关系代词从句更加靠近构式抽象度连续统中的词语层面,易于总结规律。

5 结语

构式抽象层—具体层的对应性在构式句法中占据核心地位,本文以英语本族语者的关系代词取舍为研究对象,通过基于二元逻辑斯蒂回归模型的多因素分析,尝试从抽象层与具体层同步论证关系代词取舍构式的原型性、概率性、抽象度。研究发现支持戈德堡(2006)的构式抽象层—具体层对应性假设,进一步验证了迪维亚克和阿尔佩(2013)和德肖尔(2014)基于词语构式所得出的研究发现。后续研究可以纳入二语学习者语料,开展对比分析,也可以尝试对语言使用者的语言产出与理解开展交叉验证,进一步确认构式的原型性,还可以汉语构式为研究对象,开展抽象层—具体层的对应性分析。

参考文献:

- Ambridge, B. 2013. How Do Children Restrict Their Linguistic Generalizations? An (Un-) Grammaticality Judgement Study[J]. *Cognitive Science: A Multidisciplinary Journal* (3): 508-543.
- Arnold, J., Losongco, A., Wasow, T. & R. Ginstrom. 2000. Heaviness vs. Newness: The Effects of Structural Complexity and Discourse Status on Constituent Ordering[J]. *Language* (1): 28-55.
- Ariel, M. 1990. *Accessing Noun-Phrase Antecedents*[M]. New York: Routledge.

- Barðdal, J. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic* [M]. Amsterdam: John Benjamins.
- Bernaisch, T., Gries, S. & J. Mukherjee. 2014. The Dative Alternation in South Asian English(es): Modelling Predictors and Predicting Prototypes[J]. *English World-Wide* (1): 7-31.
- Deshors, S. 2014. Towards an Identification of Prototypical Non-native Modal Constructions in ESL: A Corpus-Based Approach [J]. *Cognitive Linguistics and Linguistic Theory* (2): 243-272.
- Diessel, H. 2004. *The Acquisition of Complex Sentences* [M]. Cambridge: Cambridge University Press.
- Divjak, D. & A. Arppe. 2013. Extracting Prototypes from Exemplars: What Can Corpus Data Tell Us About Concept Representation? [J] *Cognitive Linguistics* (2): 221-274.
- Ellis, N. 2003. Constructions, Chunking and Connectionism: The Emergence of Second Language Structure [G]//C. Doughty & M. Long. *Handbook of Second Language Acquisition*. Oxford: Blackwell, 63-103.
- Ellis, N. & F. Ferreira-Junior. 2009. Constructions and Their Acquisitions: Islands and the Distinctiveness of Their Occupancy [J]. *Annual Review of Cognitive Linguistics* (7): 187-220.
- Ellis, N., O'Donnell, M. & U. Römer. 2014a. The Processing of Verb-Argument Constructions Is Sensitive to Form, Function, Frequency, Contingency, and Prototypicality [J]. *Cognitive Linguistics* (1): 55-98.
- Ellis, N. & M. O'Donnell. 2014b. Construction Learning as Category Learning: A Cognitive Analysis [G]//T. Herbst, H-J. Schmid & S. Schueller. *Constructions Collocations Patterns*. Berlin: Mouton de Gruyter, 63-89.
- Fox, B. & S. Thompson. 2007. Relative Clauses in English Conversation: Relativizers, Frequency and the Notion of Construction [J]. *Studies in Language* (2): 293-326.
- Gibson, E. 1998. Linguistic Complexity: Locality of Syntactic Dependencies [J]. *Cognition* (1): 1-76.
- Gilquin, G. 2015. Contrastive Collostructional Analysis: Causative Constructions in English and French [J]. *Zeitschrift für Anglistik und Amerikanistik* (3): 253-272.
- Goldberg, A. 1995. *Constructions: A Construction Grammar Approach to Argument Structure* [M]. Chicago: University of Chicago Press.
- Goldberg, A. 2006. *Constructions at Work: The Nature of Generalization in Language* [M]. Oxford: Oxford University Press.
- Gries, S. 2003. Towards a Corpus-Based Identification of Prototypical Instances of Constructions [J]. *Annual Review of Cognitive Linguistics* (1): 1-27.
- Gries, S. 2013. *Statistics for Linguistics with R*. (2nd revised & extended ed.) [M]. Berlin: Mouton de Gruyter.
- Gundel, J., Hederg, H. & R. Zacarski. 1993. Cognitive Status and the Form of Referring Expressions in Discourse [J]. *Language* (2): 274-307.
- Jespersen, O. 1933. *Essentials of English Grammar* [M]. London: Allen and Unwin.
- MacWhinney, B. 2008. A Unified Model [G]//P. Robinson & N. Ellis. *Handbook of Cognitive Linguistics and Second Language Acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates, 168-196.
- Mak, W., Vonk, W. & H. Schriefers. 2002. The Influence of Animacy on Relative Clause Processing [J]. *Journal of Memory and Language* (1): 50-68.
- Quirk, R., Greenbaum, S., Leech, G. et al. 1985. *A Comprehensive Grammar of the English Language* [M]. London: Longman.
- Rohdenburg, G. 1996. Cognitive Complexity and Increased Grammatical Explicitness in English [J]. *Cognitive Linguistics* (2): 149-182.

- Schönefeld, D. 2015. A Constructional Analysis of English Un-Participle Constructions[J]. *Cognitive Linguistics* (3): 423-466.
- Sperber, D. & D. Wilson. 1986. *Relevance*[M]. Oxford: Blackwell.
- Taglicht, J. 1972. A New Look at English Relative Constructions[J]. *Lingua* (29): 1-22.
- Temperley, D. 2003. Ambiguity Avoidance in English Relative Clauses[J]. *Language* (3): 464-484.
- Tomasello, M. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition* [M]. Harvard: Harvard University Press.
- Traxler, J., Williams, R., Blozis, S. et al. 2005. Working Memory, Animacy, and Verb Class in the Processing of Relative Clauses[J]. *Journal of Memory and Language* (2): 204-224.
- Verbeemen, T., Vanpaemel, W., Pattyn, S. et al. 2007. Beyond Exemplars and Prototypes as Memory Representations of Natural Concepts: A Clustering Approach[J]. *Journal of Memory and Language* (4): 537-554.
- Wasow, T., Jaeger, T. & D. Orr. 2011. Lexical Variation in Relativizer Frequency[G]//H. Simon & H. Wiese. *Expecting the Unexpected: Exceptions in Grammar*. Berlin: Mouton de Gruyter, 175-195.
- Wiechmann, D. 2015. *Understanding Relative Clauses: A Usage-Based View on the Processing of Complex Constructions* [M]. Berlin: Mouton de Gruyter.
- 董保华, 张涛, 陈崇国. 2025. 语言研究的认知功能范式探析[J]. 外国语文 (2): 31-43.
- 房印杰. 2017. 关系代词取舍的共现词分析——中国英语学习者和英语本族语者的对比研究[J]. 外语与外语教学 (3): 100-108.

Schematic and Concrete Correspondence in Constructions: A Multifactorial Analysis of Relativizer Omission

FANG Yinjie

Abstract: The correspondence between the schematic and concrete levels of constructions is one of the core hypotheses in construction grammar; however, empirical testing of this hypothesis has been underexplored. This study carries out a binary logistic regression analysis of relativizer omission by native speakers of English, aiming to cross check its prototypicality, probabilistic tendency and schematic status in schematic and concrete levels. The multifactorial analysis reveals that a) the length of the relative clause, the part of speech of the subject within the clause, and the syntactic role of the antecedent in the main clause can be verified as the prototypical features; b) a distinct probabilistic distribution could be located with explicit relative clauses displaying a weaker predictability; c) regarding the schematic-concrete continuum, a sharp contrast can be identified between implicit and explicit relative clauses with the former more geared towards the concrete end.

Key words: relativizer omission; correspondence; binary logistic regression; prototypicality

责任编辑: 蒋勇军