

翻译汉语主语回指语显隐机制研究^{*}

——条件推断树法

徐秀玲

(北京外国语大学高级翻译学院 北京 100089)

摘 要: 现有的语料库翻译研究对译文语篇特征的关注较少。本研究旨在考察翻译汉语中显性和隐性主语回指语的选择现象。我们从原创汉语和翻译汉语语料库中抽取大规模回指用例,标注多个语境因素,并采用条件推断树法开展多因素分析。结果显示:原创汉语和翻译汉语中的主语回指语显隐,受到潜在指称干扰和句子间隔距离的相似影响;但进一步受到小句间隔距离和先行语语法位置的不同影响,具体表现为翻译汉语比原创汉语在各类话题链中使用更多的显性回指语,呈现出语内衔接显化的特点。翻译汉语多用显性主语回指语,可能与源语透过效应和译者风险规避策略有关。本研究是语料库翻译研究方法走向细颗粒度语境化分析的一项尝试。

关键词: 主语回指语;衔接显化;翻译汉语;多因素分析;条件推断树

DOI:10.13458/j.cnki.flatt.004681

中图分类号: H0 文献标识码: A 文章编号: 1004-6038(2020)03-0044-10

1. 引言

翻译语言作为一种“第三语码”(Frawley, 1984),既不同于源语,又有别于目标语。翻译语言的这种独特性应当在不同语言层面均有体现。然而,现有基于语料库的翻译语言研究,多限于描写译文的词汇和语法特征,对译文语篇特征关注较少(许家金 徐秀玲, 2016: 94)。鉴于回指是一个重要的语篇现象,是实现篇章衔接连贯的重要手段,本研究以原创汉语为参照,考察英译汉翻译汉语中的回指使用,以管窥翻译汉语在语篇层面的特点。

按照回指语的形态,汉语回指大致可分为显性回指(包括名词和代词)和隐性回指(零形式)。前人研究(如 Chen, 1986: 129)发现,汉语回指语大多出现在主语位置,因而本研究将聚焦于翻译汉语中主语回指语的显隐选择。显性和隐性主语回指语分别如例(1)和例(2)所示。

(1) ①那个男孩_i 穿一条褪色的足球短裤, ②他_i 大约九、十岁大。(焱炎 K46A) ①

(2) ①菲尔_i 已经完成了对柴郡和兰开夏郡的视察, ② 0_i 正在回来的路上。(焱炎 R09A)

语言形式的选择并非任意,而是受语义、句法、语篇、认知等多层面语境特征的制约。本研究从原创汉语和翻译汉语语料库中抽取大规模主语回指用例,标注多个语境特征,开展多因素分析,以期细颗粒度地描写和刻画翻译汉语主语回指语显隐的特点。

^{*} 基金项目: 本文系教育部人文社会科学重点研究基地重大项目子课题“大数据视野下的外语及外语学习研究”(项目编号: 17JJD740003) 阶段性成果。

作者简介: 徐秀玲, 博士, 研究方向: 语言对比与翻译, 语篇分析, 语料库语言学。作者邮箱: xuxiuling@bfsu.edu.cn

2. 文献综述

回指是汉语语言学研究的一个经典议题,不少学者从语义、句法、语篇、认知等角度探讨了影响使用零形回指和显性回指的因素。例如,陆振慧(2002)发现生命度制约回指选择,零形回指常用于指称有生命的人或动物,而指称无生命的事物或概念多用名词。蒋平(2004)和许余龙等(2013)考察了先行语语法位置对回指语的影响,发现汉语零形回指的先行语大多位于主语或主语领属语位置。不少学者从语篇的角度讨论了回指形式的选择问题。例如, Li(1985)和 Chen(1986)探讨了先行语和回指语之间的间隔距离,以及前文中潜在的其他名词指称干扰对回指选择的影响。结果发现,当先行语位于前一小句或同句时,倾向于使用零形回指,随着间隔距离的增大,选择代词和名词等显性回指语的可能性也增大。零形式主要用于无潜在指称干扰的语境中,指称干扰越多,使用显性回指语的几率越大。需要说明的是,回指距离和潜在指称干扰也是 Ariel(1990)可及性(accessibility)理论中影响先行语可及性的两个主要因素。间隔距离较短且无指称干扰时,先行语具有很高的可及性,在大脑记忆系统中处于高度激活状态,回指语可使用较为简单的回指编码;反之,则需要使用较为复杂的回指编码。此外, Li 和 Thompson(1979)以及陈平(1987)分别指出,小句间的“连接性”(conjoinability)和指称对象的连续性(continuity)是零形回指使用的条件。也有学者(廖秋忠,1984)认为语体正式程度影响回指选择,即较正式的语体同较非正式的语体相比,零形回指的使用要少。然而,上述研究或基于内省,或为单因素研究,未考察多个因素如何共同制约回指显隐的选择。

语料库翻译学领域对回指现象关注较少,可能与零形回指和名词回指难以自动提取有关。一些研究探讨了翻译汉语中的人称代词使用,主要限于代词的频数统计,很少关注使用语境。黄立波(2008)与本研究较为相关。该研究考察了英汉翻译中的人称代词主语使用,发现无论是文学还是非文学文本,翻译汉语中人称代词主语的使用频率均高于原创汉语。作者指出,翻译汉语比原创汉语使用更多的代词,主要是因为翻译汉语受到英语原文影响而保留主语,造成译文中的主语冗余现象。在语料库翻译学领域,尚无研究系统考察翻译汉语中显性和隐性主语回指语的选择问题。鉴于此,本研究以原创汉语为参照,对翻译汉语中主语回指语显隐选择开展多因素分析,尝试回答以下研究问题:

- (1) 原创汉语和翻译汉语主语回指语显隐的制约因素有何共性和差异?
- (2) 翻译汉语主语回指语显隐特点的产生有何动因?

3. 研究方法

本研究使用的多因素分析方法为条件推断树法(conditional inference tree),旨在考察原创汉语和翻译汉语中主语回指语显隐与一系列语义、句法、语篇语用等语境因素之间的关系。本文所用语料来自 ToRCH2009 现代汉语语料库(以下简称“ToRCH2009”)和焱炎平衡英汉平行语料库(以下简称“焱炎语料库”)。其中 ToRCH2009 所收集语料为原创汉语文本,焱炎语料库则包含英语原文及对应的汉语译文。两者均由北京外国语大学许家金教授主持创建,均采用布朗语料库(Brown Corpus)的建库模式,涵盖新闻、通用、学术、小说

4 大语体,并 15 子类。每个语料库由 500 篇 2 000 词的文本构成,库容约为 100 万词^②。鉴于回指是一个非常普遍的语言现象,我们对两库中的回指用例进行了抽样,即从新闻、通用、学术、小说 4 个语体中各抽取 15 个文本,选取每个文本前 50 个回指用例中的主语回指。最终获得原创汉语主语回指 1841 例,翻译汉语主语回指 1808 例,具体分布见表 1。

表 1 显性和隐性主语回指语的分布

语料库	显性主语回指语	隐性主语回指语	共计
ToRCH2009(原创汉语)	909	932	1841
赅炎语料库(翻译汉语)	1306	502	1808

根据已有文献(如 Chen,1986;Li,1985;蒋平,2004;许余龙,2004;黄立波,2008;廖秋忠,1984;陆振慧,2002),本文对每一个主语回指用例进行多种语境因素的标注。标注的因素及因素水平见表 2。

表 2 标注框架

因素类型	因素		因素水平	
反应变量	主语回指语显隐	AnapType	显性	Non-Zero
			隐性	Zero
	指称对象生命度	Animacy	有生	Anim
			无生	Inanim
	先行语语法位置	AnteSynp	主语	Subj
			主语领属语	SubjPoss
			其他	Other
	回指距离(段落)	RefdPara	0,1,2,.....	
	回指距离(句子)	RefdSent	0,1,2,.....	
	回指距离(小句)	RefdClau	0,1,2,.....	
解释变量	潜在指称干扰	PotlRefe	1 个潜在指称对象	1
			2 个及以上潜在指称对象	2
			5 个小句内无潜在指称对象	3
	语体	TextType	新闻	News
			通用	General
			学术	Academic
			小说	Fiction
	语料库	Corpus	原创汉语	Native
			翻译汉语	Translated

其中关于主语回指语显隐的标注,专有名词、有定名词、光杆名词、代词均标为显性回指语,零形式标为隐性回指语。指称对象若为人或动物,则标为有生,其余标为无生。在先行语语法位置的标注中,除主语和主语领属语之外的语法位置,如动宾、介宾、兼语及各自的领属语,均标为其他。

回指距离指的是回指语和其最近的先行语之间的篇章间隔距离。本研究参照许余龙

(2004)的做法,将段落、句子、小句多个距离参数综合起来使用,以最大程度地降低单一参数的不可靠性。本研究中,句子的划分以句末句号、问号或感叹号为标准(黄伯荣 廖序东, 1981)。小句的划分则主要整合屈承熹(1998/2006)、陈平(1987)、Hu(2008)的观点,即一个汉语小句至少包含一个述谓结构,通常以逗号、冒号或句末标点断开;连动式和兼语式均看作一个小句,内嵌小句(如主语从句、宾语从句、定语从句等)也视为一个小句。我们对每个小句在整个语篇中的位置进行编号,如表3中的第1个小句位于整个语篇的第2段,第4句,第6小句。根据回指语和先行语所在小句的编号,可推算出两者之间的各类篇章间隔距离。例如,第2段,第5句,第8小句中的“徐光宪”为回指语,其最近的先行语为第2段,第4句,第6小句中的零形式,两者之间的回指距离为0个段落,1个句子,2个小句。

表3 段落、句子、小句的划分与编号(例句选自 ToRCH2009 G31)

段落	句子	小句	文本
2	4	6	五十七年前,徐光宪 _i 与 ₀ 妻子高小霞一起远涉重洋从美国回到祖国,
2	4	7	“一定要把所有的知识都奉献给祖国”。
2	5	8	五十七年后的今天,徐光宪 _i 兑现了自己当初的承诺。

潜在指称干扰是指前文中潜在的可竞争作为指称对象的名词短语的数量(Chen, 1986: 176)。相关文献(如Brown, 1983; Chen, 1986)一般将“前文”限定在回指语之前5个小句的范围。可竞争性是判断是否存在潜在指称干扰的重要依据。Givón(1983: 14)认为,只有当前文中的名词短语与回指语所在小句中的谓语具有语义(生命度、施事性等)相容性时,才被视为潜在的指称干扰。Chen(1984: 9)进一步指出,潜在干扰的判定除了需要交际双方的语言学知识外,还需依赖于其世界知识。本研究采用Brown(1983: 319-320)的三分模式进行标注。当前文(即回指语之前5个小句)中仅存在一个潜在的指称对象时,潜在指称干扰标为1,此时若使用零形回指,不会产生歧义。当前文中存在两个或两个以上的潜在指称对象时,标为2,此时若用零形回指,会造成指称歧义。例如,表3第3个小句中的谓语为“兑现了自己当初的承诺”,从其语义相容性及世界知识可推断,前文中有两个潜在的可竞争指称对象,即“徐光宪”和“徐光宪与妻子高小霞”。第3个小句中的主语若用零形式,则会引起指代不明。当前5个小句中搜寻不到潜在的指称对象(即回指距离长于5个小句)时,标为3,此时由于篇章间隔距离较长,使用零形回指易造成歧义。潜在指称干扰的标注由两名标注员合作完成。随机抽取总语料的10%,两名标注员分别标注后计算一致性,结果显示Kappa系数达到0.78,标注一致性良好。

4. 结果与讨论

语料标注完成以后,使用条件推断树的方法开展多因素分析。条件推断树是分类树中的一种,可将反应变量与多个解释变量之间的关系通过树形结构展现,提高统计结果的可读性。具体而言,条件推断树通过递归二元分裂(recursive binary splitting)确定树形结构,包括以下几个步骤(Levshina, 2015: 291)。1)对反应变量与每个解释变量间的关系进行显著性检验^③,根据 p 值大小选择最显著的解释变量;2)对选中的解释变量进行二元分裂,将数据分成两组;3)对每一组数据重复步骤1)和2),直至所有分裂不再显著。上述过程最终可生

成一个自上而下绘制的树形结构图。

条件推断树适用于研究不同语言形式之间的选择问题,在语言学和翻译研究中已经得到一些应用(如 Bernaisch et al. 2014; Kruger 2019; Tagliamonte & Baayen 2012)。本研究使用条件推断树的方法,以考察多种语境因素对原创汉语和翻译汉语中主语回指语显隐选择的影响。

条件推断树模型的构建在 R 软件中完成,使用了{partykit} 程序包中的 ctree 函数。模型的反应变量为 AnapType(主语回指语显隐),解释变量为其他所有标注因素(见表 2)。最终进入模型的显著解释变量有 5 个,即 PotlRefe(潜在指称干扰)、RefdSent(回指距离(句子))、Corpus(语料库)、RefdClau(回指距离(小句))、AnteSynp(先行语语法位置)如图 1 条件推断树所示。该模型的分类正确率为 88.6%,远高于基准分类正确率 60.7%(计算方式为总是选择频率更高的显性回指语: $(909 + 1306) / (1841 + 1808)$),说明模型具有较好的分类和预测能力。

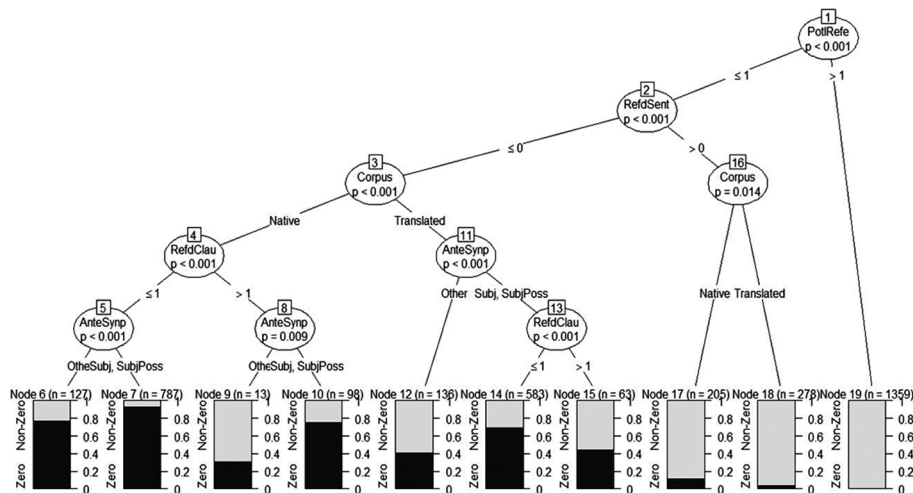


图 1 原创汉语和翻译汉语中主语回指语显隐的条件推断树

图 1 中的椭圆节点代表解释变量,顶部有节点序号。“树枝”基于椭圆节点分裂而出,代表解释变量分裂的条件。不再分裂的节点即为叶节点(位于树形图底部),代表符合每种分裂条件组合的数据集。顶部有节点序号及数据总量。深灰色和浅灰色部分分别代表隐性回指语(Zero)和显性回指语(Non-Zero)的比例。解释变量在树形图中所处的层次越高,意味着其对反应变量的影响越大。图 1 表明, PotlRefe(潜在指称干扰)和 RefdSent(回指距离(句子))对主语回指语显隐的影响最大, Corpus(语料库)影响次之, RefdClau(回指距离(小句))和 AnteSynp(先行语语法位置)影响最小。由于语料库(原创汉语 vs. 翻译汉语)没有处于树状图的顶层,可以得知在原创汉语和翻译汉语中,一些因素对主语回指语显隐选择的影响趋同。此外,两个语料库中的主语回指语显隐选择也受到一些因素的不同影响。下文将分别讨论两者之间的共性和差异。

4.1 原创汉语和翻译汉语中主语回指语显隐选择的共性

从图 1 条件推断树可以看出,潜在指称干扰、句子间隔距离对原创汉语和翻译汉语中主语回指语显隐选择的影响较为相似。根据这两个因素的分裂方式,可分为三种情况讨论,详见表 4。第一种情况:当 PotlRefe > 1,即存在指称干扰时,原创汉语和翻译汉语中的主语回

指语几乎全部为显性表达,整体比例高达 99.6% (叶节点 19)。第二种情况:当 $\text{PolRefe} \leq 1$ 且 $\text{RefdSent} > 0$,即不存在指称干扰,且先行语和回指语跨句时,原创汉语和翻译汉语都非常倾向于选择使用代词或名词等显性主语回指语,整体比例为 93.0% (叶节点 17 和 18)。第三种情况: $\text{PolRefe} \leq 1$ 且 $\text{RefdSent} \leq 0$,即不存在指称干扰,且先行语和回指语位于同一个句子时,原创汉语和翻译汉语都倾向于选择隐性的零形回指,整体比例为 77.1% (叶节点 6、7、9、10、12、14、15)。

表 4 原创汉语和翻译汉语中主语回指语显隐选择在三种语境下的共性

分裂条件	叶节点	显性回指语	隐性回指语
$\text{PolRefe} > 1$	19	1353 (99.6%)	6 (0.4%)
$\text{PolRefe} \leq 1$ 且 $\text{RefdSent} > 0$	17, 18	449 (93.0%)	34 (7.0%)
$\text{PolRefe} \leq 1$ 且 $\text{RefdSent} \leq 0$	6, 7, 9, 10, 12, 14, 15	413 (22.9%)	1394 (77.1%)

上述显隐性回指语的分布支持前人研究发现,即零形回指主要用于无指称干扰且先行语在同一句内的环境中,当有指称干扰或先行语与回指语跨句时,非常倾向选用指称更为明确的代词或名词 (Chen, 1986; Li, 1985; Hu, 2008; 蒋平, 2004; 许余龙, 2004)。潜在指称干扰和回指距离是 Ariel (1990) 可及性理论中影响先行语可及性的两个主要因素,与认知有关。潜在指称干扰、句子间隔距离对原创汉语和翻译汉语中主语回指语显隐选择的影响趋同,这说明翻译汉语与原创汉语中的语言使用均受制于人类的认知机制。

4.2 原创汉语和翻译汉语中主语回指语显隐选择的差异

由图 1 条件推断树可观察到,原创汉语和翻译汉语中的主语回指语显隐选择也存在显著差异,且仅出现在没有指称干扰的情况下 ($\text{PolRefe} \leq 1$)。原创汉语和翻译汉语之间的差异体现在两大类语境中:无指称干扰且跨句 ($\text{PolRefe} \leq 1$ 且 $\text{RefdSent} > 0$) 和无指称干扰且同句 ($\text{PolRefe} \leq 1$ 且 $\text{RefdSent} \leq 0$),分别对应于椭圆节点 16 “Corpus” 和椭圆节点 3 “Corpus”。尽管 4.1 节提到原创汉语和翻译汉语在这两类语境下表现总体上较为相似,但仍呈现出细微差异。下面将分别展开讨论。

4.2.1 “无指称干扰且跨句”语境下的差异

叶节点 17 显示,当没有指称干扰且先行语与回指语跨句时,原创汉语中有 11.2% 的主语回指语为零形式,如例 (3) 第 5 小句主语位置即采用了零形回指。叶节点 18 表明,翻译汉语中该比例仅为 4.0%,绝大多数选择显性回指语,如例 (4) 第 5 小句主语位置的代词。这说明翻译汉语中的零形回指使用对跨句间隔距离很敏感,不如原创汉语灵活。当然,该差异并不是特别明显,由椭圆节点 16 的 p 值 = 0.014 可以看出。

(3) ①小先义_i 成为住校生, ② \emptyset_i 每星期回家背米到学校蒸饭, ③ \emptyset_i 带一搪瓷缸咸菜下饭, ④咸菜长了毛也舍不得扔掉。⑤ \emptyset_i 初中毕业后因学习成绩不好, ⑥ \emptyset_i 没有考上高中, (ToRCH2009 G47)

(4) ①克里夫的老爸_i 会在水中悬留好一阵, ②然后 \emptyset_i 才像个香槟酒里的气泡一样, ③ \emptyset_i 悠然浮到水面上, ④ \emptyset_i 一身清白地迎接新年。⑤然而在未来的 12 月里, ⑥他_i 还是会像个王八蛋一样开始新一轮的折腾、恶搞和胡闹(焱炎 N02)

4.2.2 “无指称干扰且同句”语境下的差异

由椭圆节点3的 p 值 < 0.001 可知,原创汉语和翻译汉语中主语回指语显隐选择的差异在“无指称干扰且同句”的语境下更为显著。图1显示,在该语境的前提下,原创汉语和翻译汉语中主语回指语的显隐均进一步受到 RefdClau(回指距离(小句))、AnteSynp(先行语语法位置)的影响,但影响有所不同。根据这两个因素的分裂方式,可分为三种情况探讨原创汉语和翻译汉语的差异,如表5所示。

表5 原创汉语和翻译汉语中主语回指语显隐在三种“无指称干扰且同句”语境下的差异

分裂条件	原创汉语			翻译汉语		
	叶节点	显性 回指语	隐性 回指语	叶节点	显性 回指语	隐性 回指语
RefdClau ≤ 1 且 AnteSynp = Subj SubjPoss	7	56 (7.1%)	731 (92.9%)	14	179 (30.7%)	404 (69.3%)
RefdClau > 1 且 AnteSynp = Subj SubjPoss	10	24 (24.5%)	74 (75.5%)	15	35 (55.6%)	28 (44.4%)
AnteSynp = Other	69	38 (27.1%)	102 (72.9%)	12	81 (59.6%)	55 (40.4%)

(a) 先行语位于前一小句的主语或主语领属语位置

第一种情况为 RefdClau ≤ 1 且 AnteSynp = Subj SubjPoss,即先行语位于前一小句的主语或主语领属语位置。该语境下,原创汉语主语回指语使用零形式的比例为92.9%(叶节点7),而翻译汉语中该比例仅为69.3%(叶节点14)。这证实了前人研究的发现,即当先行语在前一小句,或位于主语、主语领属语位置,其可及性很高,回指语很有可能选择零形式(Chen,1986; Li,1985; 蒋平,2004; 许余龙等,2013)。实际上,这种语境常构成一个“话题链”(Tsao,1979),即多个小句共享一个话题。在汉语话题链中,话题往往在第一个小句显性出现,在之后的小句中采用零形回指(王文斌,何清强,2016:661),这是汉语典型的篇章组织方式。如原创汉语例(5)是针对话题“几个少年”展开的描述,第2和第3个小句均使用零形回指。然而,翻译汉语中有更多的话题链以显性回指的方式呈现,如例(7)和例(8)的话题链中分别使用了两个代词回指和一个名词回指。

(5) ①几个少年_i,原本正懒懒地斜倚在河边微湿的坡地上,② \emptyset_i 支起三五杆简陋的钓竿,③ \emptyset_i 望着水波里起伏不定的浮子发呆。(ToRCH2009 N10)

(6) ①这一天,他_i径直来到小屋门前,②他_i仔细地查看着,③几个小时,他_i都在摆弄着门上的铰链、把手和锁子。(焱炎 N12A)

(7) ①辉瑞公司发言人证实②该药物_i是 Geodon,③该药物_i已经被批准用于成人精神分裂症和双重人格。(焱炎 A23A)

(b) 先行语位于非相邻小句的主语或主语领属语位置

第二种情况为 RefdClau > 1 且 AnteSynp = Subj SubjPoss,即先行语位于非相邻小句的主语或主语领属语位置。该语境下,原创汉语主语回指语采用零形式的比例为75.5%(叶节点10),而翻译汉语中这一数字仅为44.4%(叶节点15)。我们观察语料发现,先行语

和回指语之间间隔的小句常为内嵌小句,常无标点符号,很多时候是作为话题链某一小句的某个语法成分。这种情况下,原创汉语的主语回指语有较大可能为零形式,如例(8)话题链第3小句采用零形式;而翻译汉语使用零形式的可能性要低一些,如例(9)话题链第3小句选择了代词。

(8) ①剧中,吴琼_i一改②[黄梅戏的调子往下走的]套路,③ \emptyset_i 在4个核心唱段加入高音。(ToRCH2009 C03A)

(9) ①他们_i将自己培养成为②[深入欺骗谷的]记录人,③他们_i讲述自己的故事,(蔡炎 B14)

(c) 先行语位于其他语法位置

第三种情况为 AnteSynp = Other,即先行语位于其他语法位置。此语境下,原创汉语主语回指语选择隐性表达的比例为72.9%(叶节点6和9合并),而翻译汉语中这一比例仅为40.4%(叶节点12)。总体来看,当先行语位于其他语法位置时,选择零形回指的概率要低于主语或主语领属语位置(参见第一、第二种情况),这与前人研究结论一致(蒋平,2004;许余龙等,2013)。然而,当先行语位于存现句宾语或及物动词宾语位置时,有较强的启后性(陈平,1987),可能成为下一小句的期待主题(许余龙,2004),因而选择零形回指的可能性较大。Li(2005)认为,前一小句动宾作为下一小句话题的情况也是一种话题链。我们单独摘出先行语位于前一小句动宾位置的情况,获得原创汉语68例,翻译汉语59例。原创汉语中有52例(76.5%)使用零形回指,如例(10)话题链第2小句的零形主语。然而,翻译汉语仅有27例(45.8%)选择零形回指,超过一半的情况选择显性回指语,如例(11)话题链第3小句的代词主语。

(10) ①针对这种情况,Taylor提出了一项折中方案,② \emptyset_i 由两条原则组成:(ToRCH2009 J34)

(11) ①构造形状最好的方法是②在花瓶中放入一个十字网格,③它_i可以保持展示的
形状和整体结构。(蔡炎 E23C)

4.3 翻译汉语主语回指语显隐特点产生的动因

由上述分析可知,在无指称干扰的语境下,翻译汉语比原创汉语使用更多的显性主语回指语,是语内衔接显化的体现。这其中可能有英汉差异的动因。刘宓庆(2005: 87-88)指出,英语主语具有不可或缺性,是谓语动词形态变化的依据,主语与谓语动词结合,对全句起提挈构架作用。而汉语则不然,汉语主语与谓语动词的关系松散,不决定动词的形态,也没有提挈全句的作用,不是不可或缺。王力(2015: 39)认为,“汉语的主语以不用为常”。翻译汉语中的显性主语回指语较多,很可能是受到英语原文的影响,即“源语透过效应”(SL shining through, Teich 2003: 22)。据我们统计,在无指称干扰语境中使用的显性主语回指语,76.7%(431/562)是由英语原文中的显性表达翻译而来。如例(12)汉语译文中的两个代词主语就是受英语原文表述的影响。

(12) Synthetic insulated jackets are quite a bit heavier at the same thickness and warmth, and they don't compress and pack down as well but they keep much of their loft when wet.

①合成棉做的夹克_i 在同等厚度和保暖性下会重很多 ②它们_i 也不易被压缩打包 ③但它们_i 被打湿后还保持大部分蓬松度 (蔡炎 E22)

另一方面,翻译汉语中主语回指语的显化也可能是因为译者采取了风险规避策略 (Pym ,2015) ,即译者使用更明晰的语言表达,以降低翻译风险。如例(13) 汉语译文中添加两处代词主语回指语“他”,即采用了更明确的指称手段。

(13) Antheil had arrived in Paris in 1923 via Berlin and had soon endeared himself to its Left Bank intelligentsia ,moving with his future wife a Hungarian woman named Boski into...

①安泰尔_i 于 1923 年从柏林来到巴黎 ②很快他_i 就赢得巴黎左岸知识分子们的喜爱 , ③然后他_i 和匈牙利的未婚妻波尔基 搬到……(蔡炎 C10B)

然而, Li(1985: 171) 指出,相较于代词、名词等显性回指手段,零形回指能将小句更为紧密地衔接起来。汉语译文中使用过多的显性主语回指语,表面看上去可以加强语篇衔接,但可能会破坏小句间的紧密联系,影响译文的流畅性和地道性。

5. 结语

本文采用多因素分析中的条件推断树法,对比考察了原创汉语和翻译汉语中显性和隐性主语回指语的选择机制。研究发现,原创汉语和翻译汉语中的主语回指语显隐受到多个语境因素的制约,但两者呈现出一定的相似性和差异性。有指称干扰时,原创汉语和翻译汉语中的主语回指语显隐表现一致。无指称干扰时,两者存在细微差异:(1) 在无指称干扰且跨句的语境下,翻译汉语比原创汉语多用显性主语回指语;(2) 在无指称干扰且同句的语境下,进一步受到小句间隔距离和先行语语法位置的不同影响,具体表现为翻译汉语比原创汉语在各类话题链中使用更多的显性回指语。翻译汉语呈现出语内衔接显化的特点,可能与源语透过效应和译者风险规避策略有关。

本研究标注多个语境特征,采用多因素分析方法,对翻译汉语中的主语回指语显隐进行深入考察,是语料库翻译研究方法走向细颗粒度语境化分析的一项尝试。然而,本研究还有一些局限和拓展空间。首先,本研究仅聚焦于主语回指语的显与隐,今后可进一步区分显化形式中的名词和代词。此外,本研究未考虑不同作者和译者对主语回指语显隐的影响,今后可采用混合效应逻辑斯蒂回归(mixed-effects logistic regression) 考察作者和译者个体差异。

注释:

①本文第 3 节将介绍蔡炎语料库。

②有些语体(如新闻) 的文本一般较短,因而部分 2 000 词文本由多个较短文本组合而成。ToRCH2009 语料库库容为 100 万词,蔡炎语料库则包括 100 万词的英语原文及对应的汉语译文。

③显著性检验需要预先设定显著性水平。本研究选取较为严格的 $p = 0.001$ 的显著性水平,以尽可能降低解释变量错误分裂的可能性。

参考文献:

- [1] Ariel, M. 1990. *Accessing Noun Phrase Antecedents* [M]. London: Routledge.
[2] 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.
- [3] Brown J. 1983. Topic continuity in written English narrative [A]. In T. Givón (ed.). *Topic Continuity in Discourse: A Quantitative Cross-language Study* [C]. Amsterdam: John Benjamins.
- [4] Chen P. 1984. *A Discourse Analysis of Third Person Zero Anaphora in Chinese* [M]. Bloomington: IULC.
- [5] Chen P. 1986. *Referent Introducing and Tracking in Chinese Narratives* [D]. Los Angeles: University of California.
- [6] Frawley W. 1984. Prolegomenon to a theory of translation [A]. In W. Frawley (ed.). *Translation: Literary, Linguistic and Philosophical Perspectives* [M]. London: Associated University Press.
- [7] Givón T. 1983. Topic continuity in discourse: An introduction [A]. In T. Givón (ed.). *Topic Continuity in Discourse: A Quantitative Cross-language Study* [C]. Amsterdam: John Benjamins.
- [8] Hu Q. 2008. *A Corpus-based Study on Zero Anaphora Resolution in Chinese Discourse* [D]. Hong Kong: City University of Hong Kong.
- [9] Kruger H. 2019. That again: A multivariate analysis of the factors conditioning syntactic explicitness in translated English [J]. *Across Languages and Cultures* (1): 1-33.
- [10] Levshina N. 2015. *How to Do Linguistics with R: Data Exploration and Statistical Analysis* [M]. Amsterdam: John Benjamins.
- [11] Li J. 1985. *Participant Anaphora in Mandarin Chinese* [D]. Gainesville: University of Florida.
- [12] Li J. & S. Thompson. 1979. Third-person pronouns and zero anaphora in Chinese discourse [A]. In T. Givón (ed.). *Syntax and Semantics: Discourse and Syntax* [C]. New York: Academic Press.
- [13] Li W. 2005. *Topic Chains in Chinese: A Discourse Analysis and Applications in Language Teaching* [M]. Munich: Lincom Europa.
- [14] Pym A. 2015. Translating as risk management [J]. *Journal of Pragmatics* (85): 67-80.
- [15] Tagliamonte A. & R. Baayen. 2012. Models forests and trees of York English: Was/were variation as a case study for statistical practice [J]. *Language Variation and Change* (2): 135-178.
- [16] Teich E. 2003. *Cross-linguistic Variation in System and Text: A Methodology for the Investigation of Translations and Comparable Texts* [M]. Berlin: Mouton de Gruyter.
- [17] Tsao F. 1979. *A Functional Study of Topic in Chinese: The First Step Towards Discourse Analysis* [M]. Taipei: Student Book Co.
- [18] 陈平. 1987. 汉语零形回指的话语分析 [J]. *中国语文* (5): 363-378.
- [19] 黄立波. 2008. 英汉翻译中人称代词主语的显化 [J]. *外语教学与研究* (6): 454-459.
- [20] 黄伯荣 廖序东. 1981. *现代汉语* [M]. 兰州: 甘肃人民出版社.
- [21] 蒋平. 2004. 零形回指的句法和语篇特征研究——关于汉语零形回指先行语的识别 [D]. 上海: 上海外国语大学.
- [22] 廖秋忠. 1984. 现代汉语中动词的支配成分的省略 [J]. *中国语文* (4): 241-247.
- [23] 刘宓庆. 2005. *新编汉英对比研究与翻译* [M]. 北京: 中国对外翻译出版公司.
- [24] 陆振慧. 2002. 英汉语篇中指同表达的对比研究 [J]. *外语教学与研究* (5): 324-331.
- [25] 屈承熹. 1998/2006. 潘文国等译. *汉语篇章语法 (A Discourse Grammar of Mandarin Chinese)* [M]. 北京: 北京语言大学出版社.
- [26] 王力. 2015. *中国语法理论* [M]. 北京: 中华书局.
- [27] 王文斌 何清强. 2016. 汉英篇章结构的时空性差异——基于对汉语话题链的回指及其英译的分析 [J]. *外语教学与研究* (5): 657-668.
- [28] 许家金 徐秀玲. 2016. 基于可比语料库的翻译英语衔接显化研究 [J]. *外语与外语教学* (6): 94-102.
- [29] 许余龙. 2004. *篇章回指的功能语用探索* [M]. 上海: 上海外语教育出版社.
- [30] 许余龙 孙珊珊 段嫚娟. 2013. 名词短语可及性与篇章回指——以汉语主语属格语为例 [J]. *现代外语* (1): 1-9.

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-