

Distribution of Zero Anaphora in Chinese Texts of Different Genres

Ping Jiang, Yun Cheng

School of Foreign Languages, Nanchang University, Nanchang
Email: peacejp@sina.com, chy70@126.com

Received: Sep. 10th, 2014; revised: Oct. 11th, 2014; accepted: Oct. 20th, 2014

Copyright © 2014 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Texts of three genre types were selected to study and compare the amount and the distribution of zero anaphora in Chinese. It is found that journal and academic texts use much less zero anaphora than narratives. The positional distribution in journal texts is the simplest while the one in narrative texts is the most complicated. The commonalities are that the subject position is predominant for both the zero anaphors and their antecedents and that S-S, T-S, O-S positional relations rank as the top three for all the three genre types in Chinese.

Keywords

Zero Anaphora, Distribution, Multi-Genre Analysis

零形回指在汉语不同体裁语篇中的分布

蒋平, 承云

南昌大学外国语学院, 南昌

Email: peacejp@sina.com, chy70@126.com

收稿日期: 2014年9月10日; 修回日期: 2014年10月11日; 录用日期: 2014年10月20日

摘要

研究选取了三种汉语常用体裁的语篇, 对其中的零形回指数量与分布进行分析和比较。从数量看, 新闻

语篇和学术语篇中的零形回指偏少,叙事语篇偏多。从位置分布看,新闻语篇最简单,叙事语篇最复杂。在共性方面,无论是零形回指的位置还是先行语位置,三种体裁都以主语为主,而且,排在前三位的都是S-S、T-S、O-S位置关系的零形回指。其他位置关系的零形回指很少。这些共性对于进行汉语零形回指规则描写和计算机应用均具有意义。

关键词

零形回指, 分布, 多体裁分析

1. 引言

零形回指作为汉语语篇连贯的最常见手段已经得到公认。由于有关现象的复杂性、在语言学理论研究以及近些年在计算机语言应用研究中的重要性,零形回指一直受到学者们的研究和关注[1]-[7],近年来还有不少针对零形回指的计算机处理进行的研究[8]-[13]。零形回指的研究目的包括零形回指的生成机制和零形回指先行语的确认和计算机解析,两方面的研究都具有重要的现实意义。然而,有关研究在语料的数量和范围方面,除了少数针对口语中的零形回指和少数针对公文和新闻语篇进行的研究之外,大多数以叙事体裁的语篇为语料,较少涉及其他体裁。对于不同体裁的语篇中零形回指的使用频率和分布状态均缺乏比较和认识,不利于对汉语零形回指规则的发现和充分描写,因而也不利于进行计算机解析与实验。本研究选取了三种汉语常用体裁的语篇,根据 Ariel [14] [15]等认知功能语言学研究者的做法,对这些语篇中零形回指的数量与分布情况进行分析、研究和比较。

2. 研究准备

本研究选取汉语的学术语篇、新闻语篇、叙事语篇若干,各约6万汉字,共18万字。语篇的年代主要从上个世纪的30年代(新闻与学术语篇从50年代开始)到本世纪初,按照改革开放前的语篇和改革开放后的语篇分为早期和近期两个时段。通过观察语料及结合认知功能语言学已有的分析模式,本研究对三种体裁中的零形回指位置规律进行分析和研究。位置指的是句法位置,包括零形回指的先行语的位置、零代词本身的位置、零形回指环境中其他成分的位置。位置的类型分为7种,根据句法功能而定,包括:1)主题:用T表示;2)主语属格语:用SM表示;3)主语:S表示;4)同位语:AP表示;5)宾语的属格语:用OM表示;6)宾语:用O表示;7)旁语:用OBL表示。语料中出现零形回指的地方,即,零代词所出现的位置,均采用(*)表示。对于语料中所有的名词性成分(包括零代词在内),本研究采用括弧标注并在其前采用符号标注其句法位置。示例如下:

(1) T(它)S(眼睛里)只有 O(光), AP(一种很柔和很绵软的带着茸毛的亮光), S(*)朝 OBL(屠夫)忽闪几下, S(*)转身走开。

(2) SM(边防部队)的 S(医生和卫生员)还背着 O(药包)S(*)在 OBL(村庄里)为 OBL(居民)进行 O(最后一次治疗)。

(3) 那时 S(上学)对 OBL(这些穷孩子)来说只不过是 O(梦想), S(*)即使勉强能上学, S(*)又会遭到 OM(反动政权和头人)的 O(迫害)。

在以上第一例中,“它”是主题,在其前用T标注,“眼睛里”是主语,用S标注,“光”是宾语,用O标注,此外还有一个同位语,用AP标注,一个旁语,用OBL标注,该语句中有两处用了零形回指(*),均出现在主语位置,用S标注。在第二例中,“边防部队”是主语的属格语,用SM标注,零代词只有一个,是第二个小句的动词的主语。在第三例中,有两个零代词,都出现在主语位置,而它们所指

代的先行语却是前面旁语位置上的“这些穷孩子”用 OBL 标注，该例中有一个宾语的属格语“反动政权和头人”，用 OM 标注。

通过对语料进行分析，我们不仅发现了三种体裁的汉语语篇在零形回指数量上的差别，在位置上所呈现的异同，也发现了先行语和零形回指之间位置关系的共同规律。

3. 数据的提取与分析

从三种体裁的数据来看，零形回指的使用数量存在重要的差别，零形回指的位置则有共同规律，也有一定的差异。零形回指与先行语之间的位置关系有简单的，也有繁杂的，存在共性和个性。首先看零形回指在不同体裁中的数量。

3.1. 零形回指使用数量的分体裁考察

从三种体裁各 6 万字的语料中，本研究共获得了 6401 条零回指的使用情况，各类语篇的零回指使用数量如表 1。

表 1 显示，在总共 6401 例零形回指中，学术语篇中的零形回指最少，只有 1398 例，其次是新闻语篇，共出现 1986 例。叙事语篇中的零形回指最多，共有 3017 例。因此，从总体数量来看，新闻语篇和学术语篇中的零形回指数量偏少，叙事语篇偏多，叙事语篇中的零形回指接近前两种体裁零形回指数量之和。本研究还对各体裁的零形回指数量进行了分时段统计，结果发现，无论是早期文本，还是近期文本，零回指在学术语篇中的使用数量均为最少，新闻语篇次之，在叙事语篇中的使用数量均为最多。这说明，无论是改革开放前的语篇还是改革开放后的语篇，三种体裁在零形回指的使用数量上没有显著变化，也就说明，零代词的使用数量差异主要与语篇的体裁有关，没有时段上的差异。

3.2. 零形回指位置与位置关系的分体裁考察

零形回指的位置关系是指它与它的先行语之间的位置关系和分布规律。以往研究对于叙事语篇中的零形回指的分布有过一些分析[1][9]但对于其他语篇缺乏分析和比较。本研究首先对新闻语篇中的零形回指的位置及位置关系进行了统计和分析。

表 2 显示，在新闻语篇中，零形回指出现在三种位置，而且主要集中在主语位置，共有 1870 例，其次是宾语位置(65 例)和主题位置(51 例)。在位置关系上，又以先行语和零代词都出现在主语位置(即：主语 - 主语)为最多(1345 例)，其次是主题 - 主语(285 例)、宾语 - 主语(224 例)、宾语 - 宾语(49 例)位置关系，其他位置的零形回指很少。新闻语篇中没有主题 - 旁语、主语的属格语 - 主语的属格语、主语 - 旁语、宾语 - 旁语等位置的零形回指。

学术语篇中的零形回指，上文已经提到，其使用数量在三种体裁中最少。从位置上看(表 3)，其零形回指更为典型地集中在主语位置，共有 1332 例，其次是宾语位置(41 例)和主题位置(23 例)，还有 2 例主语的属格语位置上的零形回指。从先行语与零代词的位置关系看，主要集中在主语 - 主语(共 1003 例)、主题 - 主语(186 例)、宾语 - 主语(122 例)位置关系，其他位置关系的零形回指数量很少。没有主题 - 旁语、主语的属格语 - 主题、主语 - 旁语、宾语 - 旁语、宾语的属格语 - 宾语、旁语 - 主题、旁语 - 宾语等位置关系的零形回指。

叙事语篇的零形回指，前文已经提到，其使用数量在三种体裁中最多，占了本次统计近一半的数量。本研究的统计与分析发现，与前两种体裁相比，叙事语篇的零形回指关系更为复杂多样，且具有语篇作者个性差异。有几位小说家的零回指使用方式特别丰富而出格。从位置的总体情况看，叙事语篇的零形回指位置更为丰富，虽然主要数量也多集中在主语 - 主语、主题 - 主语和宾语 - 主语位置关系。

Table 1. Zero anaphors in texts of different genres

表 1. 零形回指在不同体裁中的使用数量

语篇类型	零回指数量	语篇类型	零回指数量	语篇类型	零回指数量
新闻早期	1036	学术早期	708	叙事早期	1551
新闻近期	950	学术近期	690	叙事近期	1466
小计	1986		1398		3017

Table 2. Positional distribution of zero anaphora in news reports

表 2. 新闻语篇的零形回指位置关系分布

先行语位置	零回指位置	有中间干扰成分		无中间干扰成分		合计
		数量	数量	数量	数量	
	T	10	8			18
T	S	85	200			285
	O	4				4
	T	1				1
SM	S	8				8
	T	21	6			27
S	S	1065	280			1345
	O	8				8
	T		4			4
O	S	18	206			224
	O	49				49
	S	1				1
OM	O	1				1
	T		1			1
OBL	S	2	5			7
	O	3				3
总计		1276	710			1986

表 4 显示, 在叙事语篇中, 零形回指出现在主语、主题、主语的属格语、宾语、旁语五种位置。出现在主语位置最多, 共有 2714 例, 其次是宾语位置(168 例)、主题位置(124 例)、主语的属格语位置(8 例)和旁语位置(3 例)。在位置关系上, 也是先行语出现在主语, 零形回指出现在主语的位置关系数量最多(2128 例), 其次是宾语 - 主语(299 例)、主题 - 主语(251 例)、宾语 - 宾语(107 例)、主语 - 主题(86 例)、主题 - 宾语(34 例)、主题 - 主题(32 例)、主语的属格语 - 主语(30 例)位置关系, 其他位置关系的零形回指很少。叙事语篇的零形回指位置关系已明显较前两类语篇的零形回指更为复杂。

4. 讨论

根据以上数据, 从三种体裁中的零形回指数量看, 叙事语篇的零形回指最多, 学术语篇的零形回指最少。从零形回指的位置看, 新闻语篇最简单, 只有三种位置, 其次是学术语篇, 有四种位置, 叙事语篇的零形回指最复杂, 有五种位置。从位置关系上看也是如此。

Table 3. Positional distribution of zero anaphora in academic texts

表 3. 学术语篇的零形回指位置关系分布

先行语位置	零回指位置	有中间干扰成分	无中间干扰成分	合计
		数量	数量	
	T	3	1	4
T	S	44	142	186
	O	13		13
SM	SM	2		2
	S	14		14
	T	14	3	17
S	S	854	149	1003
	O	4		4
	T		2	2
O	S	16	106	122
	O	24		24
OM	S	1		1
OBL	S	5	1	6
总计		994	404	1398

在个性特征方面,新闻语篇以 T-S 和 O-S 位置关系的零形回指为其特色,因为,与回指总数高出很多的叙事语篇相比,新闻语篇 T-S 位置关系的零形回指数量(285 例)甚至高出了叙事语篇(251 例),其 O-S 位置关系的零形回指数量(224 例)也较为接近叙事语篇(299 例),而与回指总数差不多的学术语篇相比,其 O-S 位置关系的零形回指又比学术语篇的(122 例)高出很多。另外,虽然 S-S 位置关系的零形回指是三种语篇的共性,但内部也有一些区别。在新闻和学术语篇中,带中间干扰成分的 S-S 位置关系的零形回指数量(分别为 1064 和 854 例)远远高于不带中间干扰成分的 S-S 回指数量(分别为 281 和 149 例)。而在叙事语篇中,则没有那么大的差距,不带中间干扰成分的 S-S 位置关系的零形回指(718 例)超过了带干扰成分的 S-S 回指数量(1410 例)的二分之一,虽然总数仍然是后者更多。

还有一些特殊情况,在新闻语篇中,动名词短语做先行语的情况较多见,如:“中西部快速增长”。另外,新闻语篇中的无指和泛指较多。领导指示、规章、政策、昭告等报道中多出现无指或泛指。还有一个特点是,与其余两种体裁相比,新闻语篇中的零回指句子结构最简单。其中略有时段上的区别。在时段上,改革开放前的新闻语篇中的零回指表达极为规整,改革开放后的零形回指在结构及位置关系上更为丰富,显示零回指的活跃性与思想开放有一定的关系,可以作为一个重要的观察点考察新闻语篇乃至其他语篇中的零形回指,虽然零形回指的使用数量在时段上并没有产生变化。

学术语篇的一个特点是,在空洞的探讨中多出现无指。另外,小句或者从句作为先行语的回指较多,“使动”结构关系的零形回指多,抽象回指多,排比式的零形回指多,介词短语和动词短语做先行语的也多。另外,语料分析还发现,在 50 年代的学术语篇中,有类似于教师札记类的文章刊登在杂志上,其中尽是零代词的无指用法,多出现在主语位置。学术语篇的语料分析还发现,回指使用的方式与个人的思维方式与思想复杂度有关。个别学者的学术语篇中出现数例极为特殊的零回指用法,含动词短语做先行语、“把 + NP”做先行语等,以及常常将 VP 转类为 NP 并使用零形回指。再次显示,零回指使用的

Table 4. Positional distribution of zero anaphora in narratives

表 4. 叙事语篇的零形回指位置关系分布

先行语位置	零回指位置	有中间干扰成分	无中间干扰成分	合计
		数量	数量	
T	T	30	2	32
	S	88	163	251
	O	34		34
	OBL	1		1
SM	T	1		1
	SM	8		8
	S	29	1	30
S	T	51	35	86
	S	1410	718	2128
	O	24		24
	OBL	1		1
	T		5	5
O	S	27	272	299
	O	107		107
	OBL	1		1
OM	S	3		3
OBL	S	2	1	3
	O	3		3
总计		1820	1197	3017

复杂度似与思想活跃、思维的严谨性和开放度有关。例如，季羨林的零回指使用极少。

叙事语篇中，主语属格语位置上的先行语大于前两类语篇主语属格语之和。与体裁风格有关系，因为叙事语篇总以人物的思想活动与行动为主述对象，人物除了在主语位置之外，也常常在主语的属格语位置出现，因而获得更多的零形回指。分析还发现，语言使用灵活性大和表达的随意性强的叙事作品中多出现特殊回指，且与作者的文化高低及个人风格有关。

在先行语的位置方面，主题位置的先行语在学术语篇中最少，新闻语篇的主题位置先行语与叙事语篇的很接近，虽然新闻语篇零形回指总数比叙事语篇的少很多。这可能与新闻语篇表达要求突出主题性这一特点有关。学术语篇中主题位置的先行语较少，也应与学术语篇强调客观性有一定的关系。

在共性方面，三种体裁的零形回指位置都以主语为主。新闻语篇的零形回指总数为 1986 例，主语位置的零形回指占了 94.2%。学术语篇中主语位置的零代词比例更高，达到 95% 以上。叙事语篇虽然在主题位置和宾语位置上还有一定数量的零形回指，但是，主语位置的零形回指仍然遥遥领先，在总数 3017 例中，约占 90%。其他位置上的零形回指在三种体裁中，总数不超过 500 例。其中，主题位置上共有 198 例。宾语位置上略多一些，共有 274 例。主语的属格语位置上很少，只有 10 例，多出现在叙事语篇。宾语的属格语位置上没有零形回指。旁语位置上有 3 例零形回指用法，都出现在叙事语篇。

在先行语的位置方面，三种体裁也都是主语占据绝对数量，共有 4643 例。其次是宾语位置的先行语，

共有 837 例，其次是主题位置，共 828 例。其他三个位置上的先行语都不多，主语属格语位置上的先行语共有 64 例，旁语 23 例，多出现在新闻语篇，宾语属格语位置上的先行语最少，只有 6 例。

如果说零形回指的先行语除了最主要的主语位置之外，还有一定数量的主题、主语属格语和宾语，约占总数的 28%，那么，零代词的位置则更为单一地集中在主语，其余位置所占的比例总共不到 8%。这是一个很重要的现象。

最后，在回指的位置关系上，三种体裁都以 S-S 位置关系的零形回指最多，而且，排在前三位的都是 S-S, T-S, O-S，排在第四位的都是 O-O 位置关系的零形回指。其他三个位置关系的零形回指很少，总共 378 例，约为 6%。

5. 小结

以上研究不仅挖掘出汉语零形回指在三种不同体裁的语篇中的个性，以引起学界的关注，更重要的是总结出了其中的共性。这些共性对于全面了解汉语语篇中零形回指的分布以及进行规则描写和计算机应用研究具有一定的指导意义。

基金项目

全国社会科学基金项目：基于数据库的零形回指解析方式实证研究，09BYY001。

参考文献 (References)

- [1] Jiang, P. (2004) Syntactic and discourse features of zero anaphora: With specific reference to its resolution in Chinese. Ph.D. Thesis, Shanghai International Studies University, Shanghai.
- [2] 陈平 (1987) 汉语零形回指的话语分析. *中国语文*, 5, 363-378.
- [3] 蒋平 (2004) 零形回指现象考察. *汉语学习*, 3, 23-28.
- [4] 蒋平 (2011) 指向宾语的零形回指. *南昌大学学报*, 5, 122-126.
- [5] 徐纠缠 (2003) 现代汉语篇章回指研究. 中国社会科学出版社, 北京.
- [6] 杨宁 (2010) 汉语零形回指: 基于话题的分析. *华南师范大学学报*, 6, 34-40.
- [7] 曾小荣 (2011) 汉语自然会话与剧作会话中第三人称零形回指现象对比研究. *解放军外国语学院学报*, 2, 18-22.
- [8] Duan, M. (2006) Parameter setting in centering theory and its effects on Chinese anaphora resolution: An empirical study. Ph.D. Thesis, Shanghai International Studies University, Shanghai.
- [9] Hu, Q. (2008) A corpus-based study on zero anaphora resolution in Chinese discourse. Ph.D. Thesis, City University of Hongkong, Hongkong.
- [10] Yeh, C.-L. and Chen, Y.-C. (2001) An empirical study of zero anaphora resolution in Chinese based on centering theory. *Proceedings of ROCLING XIV*, Tainan.
- [11] 段嫚娟, 许余龙, 付相君 (2009) 前瞻中心的排序对指代消解的影响——一项向心理论参数化实证研究. *外国语*, 3, 20-27.
- [12] 侯敏, 孙建军 (2005) 汉语中的零形回指及其在机器翻译中的处理对策. *中文信息学报*, 1, 14-20.
- [13] 王德亮 (2004) 汉语零形回指解析: 基于向心理论的研究. *现代外语*, 4, 350-359.
- [14] Ariel, M. (1988) Referring and accessibility. *Journal of Linguistics*, 24, 65-87.
- [15] Ariel, M. (1990) Accessing noun-phrase antecedents. Routledge, London and New York.