

○语言的社会维度

编者按: 社会属性是语言的标志性特征之一。本期刊发的两篇文章昭示我们两个有价值的课题: (1) 政府工作报告: 极其严谨、规范的语料, 其中蕴含着丰富的社会特点及其对语言的制约机制; (2) 语言在地域和社会形态上的变异。其实, 语言的社会性、群体性甚至变异都是社会语言学的重要研究领域。

基于语料库的中国政府工作报告核心主题词研究

狄艳华 杨忠

(东北师范大学, 长春 130024)

提 要: 语言与社会存在必然联系。一方面, 社会的发展变化会影响语言; 另一方面, 语言的发展体现社会变化。语言变化最敏感的是词汇。本研究基于语料库工具技术, 历时分析中国政府工作报告(1978-2010)的核心主题词, 从词汇入手剖析语言使用与社会发展的共变关系及规律, 结合社会表征等相关理论对其进行阐释。

关键词: 语料库; 政府工作报告; 核心主题词

中图分类号: H0

文献标识码: A

文章编号: 1000-0100(2010)06-0069-4

A Corpus-based Analysis of Central Key Words in Chinese Government Work Reports

Di Yan-hua Yang Zhong

(Northeast Normal University, Changchun 130024, China)

Social changes and development influence language, and in the mean time the development of language reveals the changes of society. Lexical changes are more obvious and direct representations of social changes. This essay analyzes the central key words in the Chinese Government Work Reports from 1978 to 2010, in order to reveal how language and society co-change based on the corpora techniques, and then gives further explanation in light of the theory of language as social representation.

Key words: corpora; government work reports; central key words

30多年来的改革开放,使汉语的词汇变化空前。《中国政府工作报告》(以下简称《报告》)作为典型的政治语篇,对中国社会的发展具有重要作用,而且体现中国社会的巨大变化。文章通过分析《报告》中的核心主题词的使用特点、分布及规律,揭示其表征的社会变化,在社会发展中揭示语言的动态本质。

1 语言与社会表征

社会表征是一种含有其本身的逻辑和语言的认知系统,是集体成员共享的观念、意向、知识及实践系统,这种思想的共识形态由社会产生,并由社会的沟通而形成共同意识的一部分(Moscovici 2000: xii)。这种社会共识性的知识体系,主要源自于人们的经验基础,同时也源自于人们通过传统、教育和社会交流接收和传递的信息、知识和思维模式。

语言是表征社会的一种形式。人们通过语言得以便捷沟通,了解社会,并形成对社会的集体认同。有一定代表性的语言载体能充分表征社会的现状、发展及变化。《报告》作为典型的政治语篇,是我们党和政府工作的航标,具有政策性、权威性和纲领性。无论从《报告》的内容、表现形式以及使用语言都十分规范。其职责传达中央政府工作的政策及实施路径,使各级政府更好地贯彻各项方针政策。《报告》直接与中国社会的经济、政治等各方面发展息息相关。追溯中国改革开放以来的历年政府工作报告,我们能看到中国政治文明所取得的巨大进步。《报告》语言作为一种典型的社会表征形式,直接体现出中国社会的发展变化。无论是新理念、新举措都通过报告的形式得以体现,被了解、认同并深入人心。这种表征使社会群体达到共识,合力共促中国之发展。

2 《报告》及其主题词、核心主题词提取

根据中国《宪法》,中央政府每5年要换届,同时召开5年1届的政协、人大会议,简称“两会”。因此,这时的《政府工作报告》可谓承前启后,既要对过去5年的工作进行总结,又要对未来的工作进行规划。《报告》及两会是中国政治生活中一件重要大事,对中国经济和社会发展有着极其深远的影响。对《报告》进行历时分析,可以看出中国政府针对国家社会发展的思路和社会变化的趋势。改革开放30多年来,中国政治、经济和社会等诸多方面都发生了翻天覆地的变化,这些变化的轨迹可以从30年来历届国务院总理所作《政府工作报告》的词汇变迁中管窥(刘小彪2008)。

《政府工作报告》作为规范、正式的文体,措辞严谨、凝练,语言朴实,信息容量大。为了更好地研究《报告》中词汇的使用与中国社会共变的关系,本研究采用了语料库技术,用科学的方法收集、整理出数据,进一步量化和统计,以实证的手段对文本的特征进行定量描述和定性分析。此分析主要集中于《报告》中的核心主题词,一是由于语料庞大(近70万字);另外根据语料库工具,对于语词意义的分析重点应放在“主题词”、“核心主题词”上,因为它们具有文本所凸显的重要意义。《报告》的“主题词”蕴含着一定的政治理念,反映出社会发展动态。本研究对《报告》中的主题词、核心主题词及其词群、搭配进行分析,并结合社会背景、《报告》内容进行进一步阐释,因为研究语言的规律,需要充分考虑到其社会文化因素。

语料库分析工具首先可以提取文本词汇的频率分布,在此基础上可以提取“主题词”并验证“主题性”(关键性)在主题词中提取“核心主题词”。这些基本信息是进一步分析语料的基础。为了清楚主题词和核心主题词的分布特点和使用规律,还要看它们在语篇中与什么词同现,词群分布怎样。

本研究采用了Mike Scott的语料库文本分析软件Wordsmith Tools 5.0程序对中国改革开放33年的政府工作报告(近70万字)进行主题词关键性分析。首先,参照LCMC(兰卡斯特现代汉语语料库,100万词次)平衡语料库,得出按关键性高低排列的词表,排在最前的意味着起用情况较大。一般情况下,在一个由相同主题文本构成的语料库中,与主题表述相关的词语具有超常的使用频率,通过统计方法把这些词提取出来并生成一个主题词表。主题词提取的主要条件是:(1)把相同主题的语料库作为观察语料库,语料的主题必须相同或类似。(2)必须具备另一个较大的语料库作为参照语料库。(3)构成语料的文本数要足够大,每篇文本的长度基本相等(卫乃兴2005:166)。如果观察语料库由多个独立文本组成,而每一个独立文本都可以生成自己的主题词表,再对这些主题词表进行对比统计,就会发现某些主题词出现的篇次

要高于其它主题词。如果设定主题词的复现篇次,就会得到另一个词表。该词表中的词由于是从各个主题词表中按篇次统计抽取的,所以与文本的主题表达具有更密切的关系,称之为“核心主题词”。核心主题词实际上要满足两个设定值,一个是该主题词在单篇文本中的频数,另一个是该主题词复现的篇次(参见WordSmith工具使用)。该核心主题词表中的词是观察语料库中所有文本共享的主题词。通过统计和分析主题词、核心主题词以及与之密切相关的词的分布和内在关系,能够发现以主题为中心的词语群。

根据公式可以计算出某一个词项的“主题性”(X²值)(参见WordSmith工具使用)。主题词的显著性也就是该词在两个词表的差异显著性,可以计算P值(表示其显著水平),P值越低,显著性水平越高。一般情况下P值小于0.05就有显著性意义,在本研究中,设定P<0.000001来判断该词为主题词,以确保主题词表具有较高的准确性。由于篇幅有限,这里仅截取了《报告》前20个主题词的词表(见表1)以说明问题。

表1 1978-2010年政府工作报告文本的部分主题词表

序号	主题词	频数	观察文本 比率%	频数	参照语料库 比率%	值 (主题性)	P值 (P<0.000001)
1	发展	3840	1.12	1434	0.17	4413.1	0.000000
2	经济	3005	0.87	1162	0.14	3373.7	0.000000
3	建设	2754	0.80	523	0.06	4370.3	0.000000
4	改革	2347	0.68	503	0.06	3556.8	0.000000
5	企业	1995	0.58	786	0.09	2209.6	0.000000
6	加强	1912	0.56	322	0.04	3158.5	0.000000
7	工作	1816	0.53	1309	0.15	1175.6	0.000000
8	我们	1727	0.50	1967	0.23	540.5	0.000000
9	社会	1689	0.49	1055	0.12	1279.9	0.000000
10	人民	1582	0.46	537	0.06	1920.3	0.000000
11	国家	1446	0.42	864	0.10	1145.0	0.000000
12	提高	1400	0.41	426	0.05	1804.9	0.000000
13	继续	1329	0.39	271	0.03	2050.9	0.000000
14	我国	1192	0.35	614	0.07	1079.5	0.000000
15	增长	1114	0.32	282	0.03	1571.0	0.000000
16	社会主义	1082	0.31	393	0.05	1260.5	0.000000
17	生产	1063	0.31	579	0.07	917.0	0.000000
18	积极	1042	0.30	225	0.03	1572.1	0.000000
19	增加	1025	0.30	289	0.03	1373.4	0.000000
20	管理	1005	0.29	567	0.07	839.7	0.000000

在上表中,“主题词”是通过与参照语料库对比提取的文本中频率超常的词语,第一个频数即该词在观察文本中出现的频数,第一个比率%即在观察文本中该词频数除以文本的总词量所得的百分比;第二组“频数”和“参

照语料库比率”是该词语在参照语料库中的频数以及与总量的百分比。“X 2 值”即主题性值,该值越大,表示该词语的主题性越强。与之相反,“P 值”表示两个样本差异的显著性,该值越小,表示差异的显著水平越高(李文中 2003: 286)。此表中居于频率榜首的词大都是语料中的关键词,也是核心主题词提取的重要基础。可见,词频和词的关键性有着必然的、正比例的关系,是文本主题的核心之所在。

为了进一步呈现《报告》的规律性特点,用语料库工具提取出从 1978 年改革开放以来 5 年 1 届的节点年《报告》“主题词”(截取主题性居于前 10 的词,以例说明,见表 2)。

表 2 政府工作报告
五年一届的节点年的部分“主题词”

序号	1978	1983	1988	1993	1998	2003	2008
1	四人帮	建设	改革	建设	改革	建设	建设
2	人民	人民	建设	改革	发展	加强	加强
3	毛主席	改革	发展	发展	加强	改革	推进
4	革命	经济	经济	经济	企业	发展	制度
5	斗争	提高	逐步	加强	建设	坚持	改革
6	社会主义	加强	加强	要	经济	增加	发展
7	我们	国家	人民	提高	社会	支持	完善
8	我国	发展	积极	人民	要	积极	保障
9	建设	必须	体制	我国	工作	实施	农村
10	发展	继续	政府	国家	继续	农村	体系

统计的第二步是计算出文本的“核心主题词”。核心主题词具有两个含义,一个与参照语料库相比,该词语在观察语料库各个单篇文本中具有超常的出现频率;二是该词语作为主题词在观察语料库中一系列相关文本的主题词表中具有较高的复现率。核心主题词的提取过程可参见语料库工具使用说明(卫乃兴 李文中等 2005: 169)。表 3 中呈现的是用语料库工具统计出的《报告》的前 10 个核心主题词,以说明问题。

表 3 1978 - 2010 年政府工作报告
文本部分核心主题词表

序号	关键主题词	复现篇数(33)	百分比%
1	发展	33	100
2	经济	33	100
3	建设	31	93
4	企业	30	90
5	改革	29	87
6	工作	27	81
7	加强	26	78
8	社会	23	69
9	继续	21	63
10	人民	17	51

根据参照标准,这里仅列出核心主题词中出现篇数 $> / = 17$ 的核心主题词。该表通过统计主题词表而提取,“复现篇数”表示该词在不同主题词中的复现数。

3 主题表达分析

Michael McCarthy 认为,我们在运用词语表达思想或对现实世界的认识和知识时,存在一种词语的谱系关系(Michael McCarthy 1990)。在这种等级结构关系中,处于最上层的是“语义空间”,“语义空间”由各种“语义场”所组成,后者表达对真实世界的知识或概念,实现为各种“词语场”。“词语场”由一些具体的词语或词组构成,表达“语义场”中的概念。同一主题能触发具有共同语用特征和意义联系的词群。从主题出发的有关主题词以及核心主题词共同构建了词语网络,通过词语选择实现语义表达,围绕主题形成独特的语义场。在这一语义场中,词与词之间具有复杂的语义关系,这些相互关联的词在语篇层面也呈现出其特点。

通过对主题词和核心主题词的统计,根据词表对《报告》主题进行语义映射,判断并划分出不同类别的词语场,以便更好地观察词群之间的关系。主题词及核心主题词是在表达某一主题时经常使用的共同词汇或核心词群。可以认定,在一个含有核心主题词的主题词表中,主题词与核心主题词,以及主题词与主题词之间构成了围绕某一主题表达而触发的复杂的词语网络(卫乃兴 2005: 169)。由于篇幅有限,本文只考察围绕以“社会”为中心的核心主题词、主题词网。围绕“社会”这一的第一层面的核心主题词有“经济”、“改革”、“发展”、“建设”4 个大方面,以这 4 个词为中心又有不同的词群搭配。其中围绕“改革”呈现的词群主要表达以下几方面的论题:(1) 改革的范围,也可称为改革涉及的面,包括宏观和微观方面,如:体制、经济、制度、企业、机构、社会、政府、教育等;(2) 行为词类,如调整、推进、深化、建设;(3) 表达进程类的词,如完善、加强、发展、提高、继续、加快、进一步等;(4) 表示“性质”的词,如社会主义改革等。围绕“经济”呈现的词群主要表达以下几方面的论题:(1) 表达性质类的词,如:宏观经济、非公有制经济;(2) 表达经济发展覆盖面的词:我国、社会、体制、结构、生活等;(3) 表达行为类的词,如坚持、调整;(4) 表达进程的词,如平稳、发展、推进、加强、增加、加快等。以“建设”为中心的词群分为:(1) 表达性质类的词,如社会主义、小康;(2) 表达范围类的词,如社会、文化、经济、体系、农村、重点、全面;(3) 表达程度和进程的词,如加强、保障、推进、加快、提高等。围绕“发展”呈现的词群主要表达以下几方面的论题:(1) 表达性质类的词,如科学、和平;(2) 表达发展范围的词,如社会、经济、制度、市场、农业、产业、地区、教育、

全面、重点; (3) 表达行为类的词,如稳定、协调、坚持、建设、改革; (4) 表达程度和进程的词,如推进、促进、加快、大力等。从语料中可以看出,行为的执行者大多数情况下是隐现的,一是因为不言自明,另一方面也是为了凸显行为本身。

从核心主题词提取的结果可以看出(见表3),中国改革开放以来的《政府工作报告》语篇中,“发展”和“经济”的复现篇数为100%,即:从1978-2010年改革开放的33篇《报告》中每一篇都出现过。足以看出,中国政府一直把“经济”和“发展”放到了重中之重的位置。其次是,“建设”、“企业”、“改革”、“工作”、“加强”、“社会”、“继续”的复现篇数在20以上,说明在经济发展的主题下,强调行动力,重在“建设”,这是各项工作的核心。同时,强调“企业”的发展,重在改革。“加强”强调的是加大力度,进一步强化当下需要不断完善的各项“工作”。“继续”强调的是持续性,不间断的延伸性行为。“社会”一词的复现率高说明中国政府对社会的整体关注是不变的宗旨。当然,社会的发展体现在诸多层面,包括经济的不断发展、社会体制的不断完善、健全的社会法律体系及切实可行的社会福利事业能满足广大人民群众所需等一系列相关内容,都不同程度地融入了“社会”这个大概概念中。在不同的历史发展阶段,中国政府从不同的发展重心对社会的全面发展给予了不同的关注。另外,“人民”、“国家”、“我国”、“提高”、“生产”、“必须”、“我们”这几个词的复现篇数在10以上,可以将其分为三类:“人民”、“国家”、“我国”、“我们”为所属性名词,“提高”和“生产”为行为动词,“必须”为意愿性动词。从这几个核心主题词的复现来看,中国政府的发展建设立足于中国特色的发展道路,归属感较明显。首先,从国家的大局出发,强调“我国”的发展道路和特点,“人民”为社会发展的首要关注。“生产”和“提高”是经济不断发展的必要条件。“必须”强调的是一种使命感,是我们肩负的责任。由此看出,核心主题词的复现篇数的多少可以看出《报告》语篇所关注的核心内容和侧重点。结合每一年政府工作报告的核心主题词提取进一步分析,便可以得出更加具体的结果,在改革开放这33年的总体发展目标下,各个阶段(见表2)及每一年要解决的重点问题。

4 结束语

语言与社会必然联系在一起,语言具有表征社会的

作用。词汇作为语言的一个重要组成部分,对经济、社会的发展变化反应最敏感、最直接,具有灵活性和动态性。基于语料库分析工具,对庞大的语料进行数据处理,为我们更科学地研究提供了重要的依据。中国政府工作报告是改革开放三十多年来中国发展变化的集中体现,基于语料库工具对其文本主题词及其特点、规律进行分析,能更好地了解其主题意义,尤其是核心主题词集中体现政府工作的重点和亮点。从词汇管窥《报告》语言呈现的规律性特点,能进一步了解语言、社会及其共变,对相关学科的研究也会有一定启示作用。

参考文献

- 管健. 社会表征理论的起源与发展[J]. 社会学研究, 2009(4).
- 李文中. 基于英语学习者语料库的主题词研究[J]. 现代外语, 2003(3).
- 刘小彪. 政府工作报告30年词汇变迁看中国社会发展[N]. 新京报, 2008.
- 卫乃兴 李文中 濮建忠. 语料库应用研究[M]. 上海: 上海外语教育出版社, 2005.
- 殷晓蓉. 话语与社会变迁[M]. 北京: 华夏出版社, 2003.
- 张曙光. 社会表征理论述评: 一种旨在整合心理与社会的理论视角[J]. 国外社会科学, 2008(1).
- Abric, J. C. Central System, Peripheral System: Their Functions and Roles in the Dynamics of Social Representations[J]. *Social Representations*, 1993(2).
- Burton, C. *Political and Social Change in China Since 1978* [M]. New York: Greenwood Press, 1990.
- Fairclough, N. *Discourse and Social Change* [M]. Cambridge: Polity Press, 1992.
- McCarthy, M. *Vocabulary* [M]. Oxford: Oxford University Press, 1990.
- Moscovici, S. *Social Representations: Explorations in Social Psychology* [M]. Cambridge: Polity Press, 2000.
- Scott, M. *WordSmith Tools Help Manual Version 4* [Z]. Oxford: Oxford University Press, 2004.
- Wagner, W. Theory and Methods of Social Representations [J]. *Asian Journal of Social Psychology*, 1999(2).