

# 基于人民日报语料的中央一号文件词频历时分析

黄水清<sup>1,2</sup>, 王东波<sup>1,2</sup>

(1. 南京农业大学信息科学技术学院, 南京 210095; 2. 南京农业大学领域知识关联研究中心, 南京 210095)

**摘要:** [目的 / 意义]揭示中央一号文件重要词语的时代特征。[方法 / 过程]以经过人工分词处理的人民日报语料为数据对象, 统计 2020 年中央一号文件中各部分的若干重要词语在 2015 年以来每年 1 月份《人民日报》中出现的频次, 从历时性的角度对比不同时期《人民日报》中频次的变化。[结果 / 结论]首先, 宏观性的顶层政策以及普遍性的社会现象涉及的词语在《人民日报》中有比较高的频次; 其次, 与“三农”相关的词汇在《人民日报》中的频次总体都是不很高, 特别是部分与具体工作相关的词汇。最后, 以人民日报语料为基础对政策文件的词频做历时性分析是可行的, 能够得出有益的结论。

**关键词:** 人民日报语料; NEPD; 中央一号文件; 词频

**中图分类号:** G353

**文献标识码:** A

**文章编号:** 1002-1248 (2020) 03-0004-06

**引用本文:** 黄水清, 王东波. 基于人民日报语料的中央一号文件词频历时分析[J]. 农业图书情报学报, 2020, 32(3): 04-09.

## Diachronic Analysis of Word Frequency in the First Document of Chinese Party Central Committee Based on People's Daily Corpus

HUANG Shuiqing<sup>1,2</sup>, Wang Dongbo<sup>1,2</sup>

(1. College of Information Science and Technology, Nanjing Agricultural University, Nanjing 210095;

2. Research Center for Correlation of Domain Knowledge, Nanjing Agricultural University, Nanjing 210095)

**Abstract:** [Purpose / Significance] The aim of this study is to reveal the characteristics of our times of important words in the No. 1 Central Document, the first document issued by the Central Committee of the Communist Party of China every year. [Method / Process] Taking the People's Daily corpus processed by manual word segmentation as the research object, the frequency of several important words in the No.1 Central Document of 2020 which have appeared in People's Daily corpus in January since 2015 is calculated. From a diachronic perspective, this paper compares the changes of word frequency in People's Daily in different periods and analyzes the reasons. [Results / Conclusions] First, words about macro top-level decision making policies and common social phenomena are found to have higher frequency in People's Daily corpus. Second, words related to three rural issues, i.e. agriculture,

收稿日期: 2020-02-06

作者简介: 黄水清 (ORCID: 0000-0002-1646-9300), 教授, 博士生导师, 南京农业大学信息科学技术学院。王东波 (ORCID: 0000-0002-9894-9550), 教授, 博士生导师, 南京农业大学信息科学技术学院。

rural area and farmers, do not have higher frequency, especially those related to specific jobs. Finally, it is feasible to use People's Daily corpus as a basis to conduct the diachronic analysis of word frequency of policy making documents.

**Keywords:** People's Daily corpus; NEPD; First Document of Chinese Party Central Committee; word frequency

## 1 引言

一号文件,即中央一号文件,原指中共中央每年发布的第一份文件。改革开放以来,由于中央一号文件多次以“三农”为主题,中央一号文件已成为中共中央重视农村问题的符号,具有强烈的政治象征意义,成了专有名词。2020年2月5日,中央一号文件由新华社授权正式发布,这是新世纪以来连续第17个以“三农”工作为主题的一号文件<sup>[1]</sup>。文件对2020年的“三农”工作作出了全面部署,明确了工作重点。

一号文件是指导当年“三农”工作的纲领性文件,其主要内容将以不同形式呈现于同时期的中央媒体。《人民日报》是中共中央机关报,是党和政府的喉舌。《人民日报》宣传党和政府的政策主张,记录中国社会的变化,报道中国正在发生的变革。包括中央一号文件在内的全国“三农”方面的政策和变革也会在同时期的《人民日报》有充分的体现。

中央一号文件一般在每年的1月份发布,也有少数年份在2月发布。即使是2月份发布的年份,中央通过的日期也在1月。如,今年的一号文件新华社授权发布日期为2月5日,但文件标明中央通过文件的日期为1月2日。一号文件通过之前,中央要召开农村工作会议,一号文件的基本内容和思路在中央农村工作会议上基本定型,中央农村工作会议的精神很快见诸于各大媒体。因此,各年度1月份的《人民日报》都将充分宣传报道与当年一号文件有关的内容。借助文献计量学中常见的词频分析的方法,对比2020年中央一号文件各部分内容涉及的关键词语在近几年1月份《人民日报》中出现的频次,可以勾画分析一号文件关键词语的时代特征。

基于2015年以来的《人民日报》文章,此前笔者

构建完成了新时代人民日报语料(New Era People's Daily Segmented Corpus,简称NEPD),并已经在网站<http://corpus.njau.edu.cn>上公布<sup>[2]</sup>。NEPD收录的《人民日报》文章全部都经过了多遍人工分词和校对,NEPD是经过人工分词标注的精语料。今年的中央一号文件确定了2020年“三农”工作的两大重点任务,共分5个部分。从一号文件各部分内容中选取涉及的关键词语,在NEPD中分别检索各年度的1月份的语料,可以很方便地得到词语出现的频次,分析词语的时代特征。

## 2 与全面建成小康社会及脱贫攻坚相关的词汇频次及分析

一号文件开篇即明确,2020年“三农”工作的两大重点是:对标对表全面建成小康社会目标,强化举措、狠抓落实,集中力量完成打赢脱贫攻坚战和补上全面小康“三农”领域突出短板<sup>[3-4]</sup>。同时,文件第1部分部署了打赢脱贫攻坚战的相关工作,包括全面完成脱贫任务、巩固脱贫成果防止返贫、做好考核验收和宣传工作、保持脱贫攻坚政策总体稳定、研究继续推进减贫工作等。

与全面建成小康社会、打赢脱贫攻坚战相关的部分重要词汇在2015年至2020年各年度1月份的《人民日报》中出现的频次见表1。

“小康”的概念在改革开放初由邓小平同志提出。党的十八大首次提出全面建成小康社会的目标。从表1可见,“小康”一词在近几年1月份的《人民日报》出现频次非常高,而且2016年较上年有大幅度增加。2020年是全面建成小康社会目标达成之年,今年1月的频次因此达到643次。

2015年11月27日—28日,习总书记在中央扶贫开发工作会议上强调,消除贫困、改善民生、逐步实

表 1 与建成小康社会及脱贫攻坚相关的词汇在《人民日报》出现的频次

	2015年(次)	2016年(次)	2017年(次)	2018年(次)	2019年(次)	2020年(次)
小康	185	396	188	334	226	653
脱贫	38	447	655	1 067	898	1 138
脱贫质量	0	0	1	25	15	14
两不愁	0	2	0	9	16	21
三保障	1	2	1	11	18	19
返贫人口	0	0	0	2	0	6
相对贫困	1	1	0	3	1	4

现共同富裕，是社会主义的本质要求，是中国共产党的重要使命，向全党吹响了脱贫攻坚的冲锋号<sup>[5]</sup>。2016年“脱贫”一词在1月份《人民日报》中出现的频次几乎是2015年1月的12倍。以后的几年，“脱贫”一词出现的频次仍然基本上保持高速增长态势，2020年1月达到创记录的1138次。在2017年召开的中国共产党第十九次全国代表大会上，习总书记作出“脱贫攻坚取得决定性进展”的重要论断，进一步明确要“脱真贫、真脱贫”<sup>[6]</sup>。“脱贫质量”一词从2018年起频次激增。此外，与之相关的“两不愁”“三保障”等词汇的频次从2018年起均明显增高。

打赢脱贫攻坚战，还需要巩固脱贫成果，防止返贫<sup>[1]</sup>。“返贫人口”在前3年的频次均为0，2017年出现了2次，今年1月份频次达到6。

十九届四中全会首次提出要“建立解决相对贫困的长效机制”<sup>[7]</sup>。这是十八大以来中央全会首次提及“相对贫困”，解决“相对贫困”将成为未来扶贫工作的重要核心内容。今年一号文件对脱贫攻坚任务完成后的工作进行了部署，要求将扶贫工作重心转向解决相对贫困，要研究建立解决相对贫困的长效机制，加强解决相对贫困问题顶层设计<sup>[1]</sup>。“相对贫困”一词除了在2018年1月出现了3次外，前几年的频次均为0或1，今年则达到了4次。

### 3 与补上农村基础设施和公共服务短板相关的词汇频次及分析

全面建成小康社会，最突出的短板在“三农”<sup>[8]</sup>。今年的中央一号文件共分5个部分，两个部分论述了

补短板。其中，第2部分为对标全面建成小康社会加快补上农村基础设施和公共服务短板，第5部分为强化农村补短板的具体措施。文件对农村公共基础设施、农村供水、农村人居环境、农村教育、农村基层医疗卫生、农村社会保障、乡村公共文化服务、农村生态环境等8个方面补齐短板提出了具体要求<sup>[1]</sup>。与8个方面补短板相关的部分重要词汇在2015年—2020年各年度1月份的《人民日报》中出现的频次见表2。

总体来看，除了“农村基础设施”和“农村人居环境”两个词频次略高外，与补短板相关的词语在2015年—2020年1月的《人民日报》都只有零星出现，频次均不高。即便是在2020年1月的《人民日报》中情况也没有太大变化。如“数字乡村”一词，2020年之前的各年度的频次都是0。2019年5月中共中央、国务院发布了《数字乡村发展战略纲要》，2020年该词出现了1次。这些情况从一个侧面说明农村的各项短板在全社会重视程度不够，今年中央一号文件提出补短板任务，要求加快补上农村基础设施和公共服务短板，让亿万农民有更多获得感，有极大的迫切性。特别是在当前新型冠状病毒感染的肺炎疫情防控中，加强农村薄弱环节建设尤为重要<sup>[8]</sup>。

### 4 与保障农产品供给和促进农民增收相关的词汇频次及分析

保障重要农产品有效供给是“三农”工作的头等大事，实现农民收入到2020年比2010年翻一番是全面建成小康社会的硬任务<sup>[9]</sup>。今年一号文件的第3部分布置了保供给、促增收的5方面任务，分别是稳定粮

表 2 与补短板相关的词汇在《人民日报》出现的频次

	2015年(次)	2016年(次)	2017年(次)	2018年(次)	2019年(次)	2020年(次)
农村基础设施	1	13	5	13	8	4
数字乡村	0	0	0	0	0	1
农村供水	0	1	0	2	0	4
城乡供水	1	0	1	0	0	0
农村厕所革命	0	0	0	0	2	0
农村人居环境	2	3	3	10	29	10
农村教育	0	8	0	3	0	0
农村基层医疗	0	0	1	0	0	1
农村社会保障	1	0	0	0	0	0
乡村公共文化	0	0	0	1	0	0
农村公共文化	0	3	1	1	0	0
农村生态环境	0	2	0	1	0	1

食生产、加快恢复生猪生产、加强现代农业设施建设、发展富民乡村产业、稳定农民工就业。与保供给、促增收 5 个方面任务相关的部分重要词汇在 2015 年—2020 年各年度 1 月份的《人民日报》中出现的频次见表 3。

根据表 3 中的词汇频次分布情况,表 3 中的词汇大致可分为 3 种类型。

第 1 种类型,词汇频次对应的各列没有取 0 值的,且极大值与极小值的比率小于 10。这类词汇包括“粮食产量”“菜篮子”“农民增收”“农民工就业”

“农业工工资”等。这种类型的词汇所代表的工作在各年度均有一定的关注度,并且,虽然每年的关注度略有起伏但总体变化幅度不是特别激烈,反映了这些工作是近年“三农”持续关注和努力的方向。

第 2 种类型,词汇的频次同样没有取 0 值的列,但极大值与极小值的比率大于 10。这类词汇包括“高标准农田”“水利建设”“返乡创业”等。这种类型的词汇所代表的工作在各年度均有一定的关注度,但起伏很大,某年或某几年可能受各种因素影响关注度陡增,反映了这些工作在近年“三农”中得到持续关

表 3 与保供给促增收相关的词汇在《人民日报》出现的频次

	2015年(次)	2016年(次)	2017年(次)	2018年(次)	2019年(次)	2020年(次)
粮食产量	18	6	5	8	5	5
生猪生产	0	0	1	0	0	5
菜篮子	5	3	3	14	13	17
现代农业设施	0	0	0	0	0	0
高标准农田	5	19	23	15	7	2
水利建设	10	16	3	14	1	1
农产品仓储	0	1	0	0	0	0
农产品冷链物流	1	1	0	1	0	0
农民增收	30	42	27	39	12	7
富民乡村产业	0	0	0	0	0	1
农民工就业	1	3	1	1	9	2
农民工工资	33	50	61	37	37	167
返乡创业	2	32	15	27	12	8
下乡创业	0	0	0	25	2	0

注并且在某年或某几年得到特别关注。

第3种类型，词汇在大多数年份频次为0，甚至6年的频次全部为0。这类词汇包括“生猪生产”“现代农业设施”“农产品仓储”“农产品冷链物流”“富民乡村产业”“下乡创业”等。其中，“现代农业设施”一词6年的频次全部为0，“农产品仓储”“富民乡村产业”两个词除了2016年、2020年频次分别为1外其它5年的频次均为0。这种类型的词汇所代表的工作在多数年份关注度很低，个别年份受到一定程度重视。典型的就是“生猪生产”一词。从表3可见，在2020年以前的5年中，“生猪生产”有4年的频次均为0，只有2017年频次为1。2019年，中国发生大面积的非洲猪瘟，生猪生产受到很大影响，因此，一号文件把生猪生产作为一项任务单独列出，2020年1月《人民日报》“生猪生产”一词也出现了5次。这个例子证明，以词汇在《人民日报》中出现的频次为依据，分析中央一号文件中的词汇，是可以得出合理的结论的。

## 5 与加强农村基层治理相关的词汇频次及分析

农村基层治理是国家治理的基石，也是国家治理的短板。中国共产党历来高度重视农村治理工作，近年来中央制定和出台一系列针对农村基层治理的制度

规定，特别是2019年1月重新制定和印发了《中国共产党农村基层组织工作条例》<sup>[10]</sup>。今年的一号文件把加强农村基层治理作为文件的第4部分，表明了农村基层治理存在的短板以及提高农村治理效能的重要性。文件要求充分发挥党组织领导作用，健全乡村治理工作体系，调处化解乡村矛盾纠纷，深入推进平安乡村建设<sup>[1]</sup>。

与农村基层治理的4个方面任务相关的部分重要词汇在2015年—2020年各年度1月份的《人民日报》中出现的频次见表4。

从表4可以看出，与农村基层治理相关的词汇在各年度《人民日报》中出现的频次总体并不高，大部分词汇在各年度1月份的《人民日报》中都只有零散出现。“农村基层组织”在2019年1月份的频次为115，原因是当月中央印发了《中国共产党农村基层组织工作条例》<sup>[10]</sup>，而其它5年“农村基层组织”的频次或者为0或者为个位数。另外，“农村基层巡察”“县乡村联动”“包村制度”“一村一法律顾问”等词汇，甚至“农村基层治理”本身，在6年中对应的频次都是0。

以上数据说明，农村基层治理涉及的各项工作在一定程度上有被忽视的现象，故而在《人民日报》这样的中央媒体上相关词语频次偏低，农村基层治理也是“三农”工作的短板，今年的一号文件将加强农村基层治理单独作为文件的一个部分，很有必要。

表4 与农村基层治理相关的词汇在《人民日报》出现的频次

	2015年(次)	2016年(次)	2017年(次)	2018年(次)	2019年(次)	2020年(次)
农村基层组织	0	5	2	6	115	0
四议两公开	0	1	1	0	5	1
村务监督	0	5	0	2	9	0
村第一书记	1	18	21	32	28	40
乡村治理	1	2	0	79	29	8
村民自治	4	9	3	21	7	0
枫桥经验	0	1	0	5	21	18
平安乡村	0	0	0	1	3	0
村霸	1	0	2	5	5	0
网格化管理	9	4	5	14	4	4
网格化服务	8	1	0	0	3	5

## 6 结语

从以上对 2020 年中央一号文件各部分内容涉及的重要词语在各年度 1 月份《人民日报》中出现的频次的统计分析可以有以下 3 点结论。

首先,宏观性的顶层政策以及普遍性的社会现象涉及的词语在《人民日报》中有比较高的频次,前者如涉及两大重点任务的“小康”“脱贫”,后者如“农民工工资”。

其次,与“三农”相关的词汇在《人民日报》中的频次总体都不是很高,特别是部分与具体工作相关的词汇。这一方面反映了全社会对“三农”工作可能关注度不足,同时表明“三农”工作中确有部分短板亟待加强。

最后,对一号文件的词频分析表明,以人民日报语料为基础对政策文件的词频做历时性分析是可行的,能够得出有益的结论。

### 参考文献:

- [1] 中共中央,国务院.关于抓好“三农”领域重点工作确保如期实现全面小康的意见[M].北京:人民出版社,2020.
- [2] 黄水清,王东波.新时代人民日报分词语料库构建,性能及应用(一)——语料库构建及测评[J].图书情报工作,2019,3(22):5-12.
- [3] 董峻,侯雪静.2020 年中央一号文件公布 提出两大重点任务[EB/OL].[2020-02-05].[http://www.xinhuanet.com/politics/zywj/2020-02/05/c\\_1125534973.htm](http://www.xinhuanet.com/politics/zywj/2020-02/05/c_1125534973.htm).
- [4] 董峻,侯雪静.抓好“三农”领域重点工作 确保如期实现全面小康——中央农办主任、农业农村部部长韩长赋解读 2020 年中央一号文件 [EB/OL].[2020-02-05].[http://www.xinhuanet.com/politics/zywj/2020-02/05/c\\_1125536076.htm](http://www.xinhuanet.com/politics/zywj/2020-02/05/c_1125536076.htm).
- [5] 新华网.习近平:脱贫攻坚战冲锋号已经吹响 全党全国咬定目标苦干实干 [EB/OL].[2020-02-12].[http://www.xinhuanet.com/politics/2015-11/28/c\\_1117292150.htm](http://www.xinhuanet.com/politics/2015-11/28/c_1117292150.htm).
- [6] 习近平.决胜全面建成小康社会 夺取新时代中国特色社会主义伟大胜利——在中国共产党第十九次全国代表大会上的报告[M].北京:人民出版社,2017.
- [7] 中国共产党中央委员会.中共中央关于坚持和完善中国特色社会主义制度、推进国家治理体系和治理能力现代化若干重大问题的决定[M].北京:人民出版社,2019.
- [8] 胡璐.加快补上农村基础设施和公共服务短板 让亿万农民有更多获得感[EB/OL].[2020-02-08].[http://www.xinhuanet.com/politics/zywj/2020-02/07/c\\_1125543238.htm](http://www.xinhuanet.com/politics/zywj/2020-02/07/c_1125543238.htm).
- [9] 于文静.持续抓好农业稳产保供和农民增收[EB/OL].[2020-02-08].[http://www.xinhuanet.com/politics/zywj/2020-02/08/c\\_1125546517.htm](http://www.xinhuanet.com/politics/zywj/2020-02/08/c_1125546517.htm).
- [10] 中共中央.《中国共产党农村基层组织工作条例》[M].北京:党建读物出版社,2019.