

1997 年中文报纸媒体流通度分析

隋 岩 张 普

北京语言文化大学语言信息处理研究所

摘 要:“流通度”概念的提出为语感的量化找到了一条可能的途径。真实文本的“流通度”基本上可以反映人们的“共同语感”。本文以“流通度”分析方法为基点,对 1997 年国内公开出版发行的报纸媒体进行了实验性处理,得出了一些初步的数据并加以分析,为以后大规模媒体及文本的分析处理做了必要的准备和探索。

关键词: 语感 流通度 加权系数 动态语料库

Analysis of Circulation of Chinese Newspapers in 1997

Sui Yan Zhang Pu

Language Information Processing Center, Beijing Language and Culture University

Abstract The concept of "Circulation" indicates a way to measure the feel for the language. The "Circulation" of real text can reflect the common feel for the language. This paper analyzes the Chinese newspapers in 1997 and has gathered some data for future research on a large scale.

Keywords FEELFORTHETHELANGUAGE CIRCULATION WEIGHTEDCOEFFICIENT DYNAMIC CORPUS

在纷繁复杂的语言现象中,研究者们困惑不解的东西很多,“语感”就是其中最具代表性的一个。语感是一种重要的客观存在,很多新词、新表达法的流行和使用都跟语感有着密不可分的关系。然而,关于语感的研究还不是很深入,“语感到底是什么?”这这个问题还不是十分清楚,关于语感的量化问题,更是没有一个让人满意的方法。

张普提出的“流通度(Circulation)”的概念,“通过测量流通度来对语感加以数学界定、加以量化,使得‘能不能说’、是否已经‘被理解’、‘被认可’、‘被传播’变得可以通过流通度的计算进行判断”,给上述问题提出了一种有前途的解决方案,为语感的量化找到了一种可能的概率统计方法,也为大规模真实文本的研究开启了一扇新的大门,具有一定的理论意义。流通度不是具体语言学的概念,而是普通语言学的概念,它的具体算法和公式应该是跨语言的,超语言的。

要找到文本的“流通度”,首先必须确定承载该文本的媒体的“流通度”。

一、媒体流通度分析的必要性

我们认为：“‘真实文本’的最重要最核心的问题是文本的‘真实流通’。只有在流通度高的文本的基础上计算使用度才是真实的使用度”，“……语言的生命力就在于这种稳定中的变化，这些变化的端倪就隐藏在大规模的真实文本（无论它们是经典的还是非经典的文本）之中，甚至就隐藏在那些非规范的现象里”。（张普：《关于语感与流通度的思考》）

真实文本是动态的，不断变化的，而流通度的计算，正好能以客观量化的手段揭示这些发展变化的规律。可以用真实可见的数学模型对真实文本的流通度加以描述和解释，对大规模语料进行实时处理，甚至对语言未来的趋势走向进行预测。这对语言应用研究和语言规范化工作都会有极大的帮助。

1、媒体流通度与文本流通度的关系

“流通度”这个概念不是人们凭空想象出来的，它是语感的量化，是语言在真实世界而非“无菌状态”中使用的真实反映。要想得到具体文本的流通度的具体数值，首先必须找到承载该语料文本或该语言事实的载体的流通度，这是最基础的部分。媒体流通度分析，是整个流通度研究方法的最基本、最重要的一个步骤。如果失去了媒体流通度的支持，文本流通度以至于更深层次的流通度也就成了“无源之水，无本之木”。

文本流通度和媒体流通度之间是一种继承与被继承的关系。媒体流通度的一切特征，在文本流通度中都应该有所体现。而文本流通度在继承了媒体流通度的全部特征之后，又会生成自己特有的属性。

流通密度是体现媒体流通度和文本流通度的重要指标之一。虽然流通密度在媒体流通度和文本流通度上都有所体现，但是在这两者中却是两个层次的问题。对于媒体流通度来说，它体现为承载语料文本的媒体整体的流通密度；对于文本流通度来说，它体现为文本在媒体内部的分布密度。如果媒体的流通密度高，文本流通度值一定会较高。可见，媒体流通度会直接影响到文本流通度值的高低。

2、确定媒体流通度的标准

既然涉及到了“流通”，媒体的流通度毫无疑问是一个动态的值。不同的媒体，在不同的时期和不同的环境等动态变化的条件影响下，其流通度的数值是各不相同的。我们不可能将媒体流通度死锁在某个固定的刻度之上。在这种情况下，用来测定媒体流通度的标准的制定就显得尤为重要。我们认为，影响和制约媒体流通度的因素应该包括：发行地域、发行内容、发行数量、发行周期、受众对象、阅读率等几个方面。这些因素本身，在一定时期、一定范围之内是稳定的。只要能够科学合理地对这些因素进行分析，就可以制定出确定流通度的标准。本文在后面给出了一种初步的设想，还只是一个粗略的概框，仍需要进一步细化和改进。

3、关于分析的可信度

在流通度的研究中，我们时时刻刻都面对着大规模的真实文本，我们的研究方法和处理模型能否实时有效地处理这些“真实”文本，是一个至关重要的指标。关于文本的真实，张普在有关论述中曾不止一次的强调过。也就是说，“真实”是流通度研究方法的生命所在。鉴于这种考虑，在本文中，我们的分析完全基于国内公开正式出版发行的报刊等出版物，并且是穷尽式的数据分析。大家知道，分析对象的数量在很大程度上决定着分析结果的可信度。样本数目越大，结果的可信程度也就越高。本文计算媒体流通度，就是要最大限度地接近高可信度这一目标，以便为后续的研究打下坚实的基础。

二、确定媒体流通度时采用的媒体范围

1、媒体的范围

“流通度”是一个通用的概念，它可以表示各种媒体在社会上的流程程度。比如报纸、杂志、书籍、广播、电视，在今天的信息化社会中，还有日益壮大起来的 Internet 等等，范围极广，数目巨大。由于篇幅所限，本文在此不可能对上述媒体进行全面的分析，因为那将是一个巨大的系统工程。所以，我们将分析范围限定在报纸和杂志上，以报纸为主，兼顾杂志。所选的报纸为 1997 年中国大陆正式出版发行的所有报纸（2058 种）。杂志方面，由于条件和时间的限制，我们暂时无法搜集到相关详细原始数据，在此主要对报纸进行初步分析。

2、类型划分

随着改革开放的深入，经济的高速发展，人们的文化生活也越来越丰富，国内报刊，特别是报纸的发行数量和种类也突出体现了这种变化。报纸刊期的形式也是多种多样的，归纳起来有以下这些类型：日报、周报、旬刊、半月刊，杂志主要有：月刊、双月刊、季刊，另外，还有一部分专业年刊等。

报刊的刊期（我们称为发行密度）对其流通度的影响也是比较大的。比如日报，每天一刊，365 天几乎没有间断，这种发行密度对读者的阅读必定会有相当程度的影响。

3、加权系数

为了对媒体流通度进行科学的量化数值计算，必须将影响流通度的各种因素考虑进去，在这里我们引进了“加权系数”概念，并首先为流通度的数值设定一个“种子”，记作“M”，我们称之为“绝对媒体流通度”。这个“种子”不是凭空想象出来的，而是通过计算某报纸或刊物的年总发行量与全部报刊年总发行量的比例关系来得到。有了这颗“种子”，再将其

与各加权系数相乘，最后求得该媒体流通度的数值。

目前已经考虑的加权系数有：

①地域系数：地域差异对媒体流通密度的影响很大。在我国，东南部沿海经济发达地区跟西北部经济欠发达地区就存在一定差距，而大城市跟中小城市之间也会有所差别。就中国大陆来说，从总体上看：东部高于西部，南部高于北部。

②版数（字数）系数：一份报纸版数的多少对流通密度也有影响。不过这种影响跟流通密度之间并不是一种正比例关系。版数多的报纸其流通密度并不一定很高。

③范畴系数：尽管目前大多数报刊都带有相当的综合性，但还是有大量专门性报刊深受人们的欢迎。如《足球》、《汽车之友》等。它们在特定人群中有着很高的流通密度，也有很高的影响力。

④类型系数：前面已经谈到，报刊的刊期方式对流通度影响很大。

⑤阅读率系数：一种报刊，不管它的发行量有多大，也不管它的内容多么的重要和有趣，只有被受众接受并阅读，才能最终实现它的价值，它所承载的内容才能对人群的语感产生影响。因此，阅读系数是影响媒体流通度诸因素中相对较重要的一个。不过，阅读率的获取也不是轻而易举的，需要大规模的科学合理的读者调查活动才能完成，本文受时间和条件限制，暂没有考虑阅读率系数。

4、媒体流通度计算公式：

综合以上论述，我们可以得到流通度的标准计算公式如下：

$$C_m = M \times ① \times ② \times ③ \times ④ \times ⑤ / 100$$

从形式上看，这个公式相当简单，然而它却充分考虑到了多种内在和外在于因素对媒体流通度变化发展的作用和影响。例如，某一份报纸的“绝对媒体流通度”可能排在第三位，表面上看似乎很高，但是，经过加权系数处理之后，它可能就排到第二十位去了，而原来处于第二十位的报纸，也可能上升到前几位。这就是“加权系数”控制作用。

三、数据分析

1、1997年国内报纸发行概况

近二十年来，中国大陆报纸的发行量增长速度很快，1978年，全国仅有报纸数百种，到1997年，已经发展到两千多种，而且种类齐全，内容丰富。我们统计了来自国家新闻出版署的数据，以1997年为例，全国共发行报纸2058种。其中全国性报纸206（10%）种，地方性报纸1852（90%）种。如下图所示：

在这十份报纸中，全国性报纸为4种，分别是1、6、7、9。值得注意的是，在地方性报纸中，广东省就占了3种，分别是3、5、10，可见，经济发展的程度对媒体的发行也有直接的影响。

从发行量上来看，名次在前 300 位的报纸就占到了总发行量的三分之二以上，这说明，尽管发行总量很大，但是，高发行量的报纸却相对集中。

另外，东南部沿海地区报纸发行种类明显高于内陆地区。表二是这 300 种报纸中，种类占前十名的省份地域分布情况：

表二 报纸发行种类前十位地域分布

位次	地区	数量
1	全国性	55 种
2	广东省	30 种
3	山东省	23 种
4	贵州省	20 种
5	江苏省	20 种
6	四川省	20 种
7	浙江省	20 种
8	河南省	14 种
9	湖南省	12 种
10	福建省	10 种

以上的统计数字可以说明，经过穷尽性分析，从一定的时间段内（比如一年）抽取出相当数量的报纸，计算它们的媒体流通度，就可以基本上代表了全体的媒体流通度。

2、加权系数的设定

①地域系数：地域系数的确定，主要依据我国各地区经济发展水平的不同。从总体上说，我国经济发展水平是东南沿海高于西北内陆，南方高于北方，这是有目共睹的实事，有关的经济论著很多。因此，我们暂时将全国划分为四个区域（这样做较为粗略，还有待进一步细分）。内蒙、宁夏、四川、贵州以西的地域系数为 1；东北三省的地域系数为 2；山海关以南，长江以北的地域系数为 3；长江以南的地域系数为 4。

②版数系数：各报纸的开本和版数不尽相同，详细数值参见下表：

表三 版数系数

开本	系数	开本	系数
8 开 4 版	1	4 开 4 8 版	9
8 开 8 版	2	4 开 6 4 版	10
8 开 16 版	3	2 开 4 版	3
4 开 2 版	1	2 开 6 版	4
4 开 4 版	2	2 开 8 版	5
4 开 8 版	3	2 开 10 版	6
4 开 12 版	4	2 开 12 版	7
4 开 16 版	5	2 开 16 版	8
4 开 20 版	6	2 开 20 版	8.5
4 开 24 版	7	2 开 24 版	9
4 开 32 版	8		

③范畴系数：目前我们所获得的 2058 报纸，按不同的范畴，分为 16 大类，这种分类标准是否合适，还可以研究，这里只是方案之一。这 16 大类分别是（括号中是范畴系

数): 人文(15)、经济(14)、卫生保健(13)、教育(12)、广播影视(11)、综合新闻(10)、文摘类(9)、体育(8)、省市机关报(8)、法律(7)、政治(6)、科技(5)、农村(4)、工业工人(3)、专业性报纸(2)、其他(1)。

④类型系数: 报纸的刊期直接决定了它的发行数量, 刊期密度大发行量也相对会提高。这2058种报纸的刊期主要有: 日报(每日发行, 9)、周六(每周6天发行, 8)、周五(每周5天发行, 7)、周四(每周4天发行, 6)、周三(每周3天发行, 5)、周二(每周2天发行, 4)、周一(每周1天发行, 3)、旬刊(2)、半月(1), 个别还有双日发行的, 大致跟周三的刊期一致。

参考上述系数制定标准, 对媒体流通度的“种子”“M”进行再处理, 可以得到下表所示的数据:

表四 媒体流通度前十位(未包含阅读率参数)¹

位次	地域	媒体流通度	年总发行量(%)
1	全国性	25.96	2.89
2	广东	19.82	1.55
3	广东	13.61	1.18
4	全国性	12.84	1.07
5	福建	11.25	1.41
6	江苏	9.97	1.56
7	全国性	9.13	0.81
8	全国性	8.91	1.11
9	广东	8.22	0.92
10	江苏	7.06	0.59

这十份报纸的年总发行量的百分比之和为 13.90%, 这个数值跟表一的数值很接近, 这说明, 发行量对媒体流通度的影响是很大的, 可以这么说, 发行量和阅读率是制约媒体流通度的两个最主要的因素。

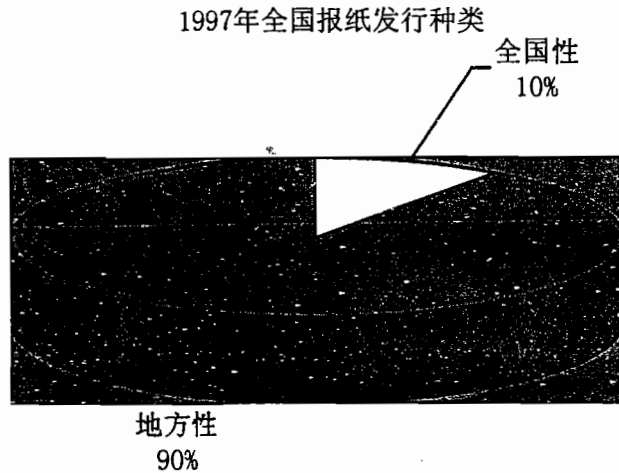
3、媒体流通度的地位

从上面的分析我们不难看出, 媒体流通度是整个“流通度大厦”的基石, 是整个流通度研究的起点。它的可信性和精确性直接关系到整个后续研究的成败, 因此需要认真对待和深入挖掘。本文所提出的计算和分析过程只是为了说明媒体流通度的确定方法, 还很不完备, 肯定存在着诸多漏洞和不妥之处, 我们还将进行系统的研究和论证。

¹ 这里计算出的“媒体流通度”是一个绝对值, 理想的做法是将其换算成0与1之间的某个对应值, 这样便于对比分析, 至于换算的方法和原则, 我们正在探讨, 将有专文论述。

四、流通度分析方法前瞻

1、关于流通度体系



其中日报 298 种，周报 1685 种，旬刊 27 种，半月刊 15 种，双日刊 1 种，其他 29 种。另外，这些报纸中还包括 79 种少数民族文字及英文报纸。

按年总发行量计算，在这 2058 种报纸中，前 10 位的报纸年总发行量之和，占全部报纸总发行量的 13.92%；前 300 位总发行量之和占 69.40%。

表 1 1997 年总发行量前 10 位的报纸 (%)²

位次	百分比
第 1 位	2.89
第 2 位	1.56
第 3 位	1.55
第 4 位	1.41
第 5 位	1.18
第 6 位	1.12
第 7 位	1.11
第 8 位	1.10
第 9 位	1.07
第 10 位	0.92
合计	13.92

流通度是一种普遍存在的现象，它存在于两大类媒体之中：书面文本（报纸、杂志、图书等）和口头文本（广播、电影、电视等），也存在于因特网上。本文我们拟先给出这两种

² 由于众所周知的原因，我们在此隐去了报纸的具体名称。

媒体的流通度层次框架及其相关的标记体系:

C: 流通度

Ct: 书面流通度	Cs: 口头流通度
Ctm: 书面媒体流通度	Csm: 口头媒体流通度
Ctt: 书面文本流通度	Cst: 口头文本流通度
Ctw: 书面词语流通度	Csw: 口头词语流通度
Ctp: 书面短语流通度	Csp: 口头短语流通度
Ctg: 书面语法流通度	Csg: 口头语法流通度
Cts: 书面语义流通度	Css: 口头语义流通度
Ctl: 书面语用流通度	Csl: 口头语用流通度

书面语和口头语之间是一种相互作用、相互影响的关系，它们各自的流通度也就处在这种不断变化和影响之中。

2、关于动态语料库

我们主张建立动态语料库，动态语料库与传统语料库最大的不同之一就在于入库的语料都带有“流通度”属性及其属性值。为了实现这一目标，建立有效的“流通度体系”的语料深加工是非常有必要的。大到文本语篇，小到短语、词语，从语音到语义都应该标注适当的流通度属性和属性值。这些标记应该包括上述体系列出的所有流通度层次，这样就能够应用中全方位把握该种语料在真实语言流通中的轨迹和特点。

3、动态智能辞典的建造

所有的这些努力，最终的目标之一就是建造一个能够即时更新的动态智能辞典。这样的辞典对于大规模真实文本的处理，对于知识的自动挖掘和自动获取都有着极其重要的意义。有了文本各个层次的流通度的信息，我们就可以更准确地掌握大规模真实文本在现实世界中的流通规律，就能够以量化的方法估算出人们的语感，从而对语言运用中的各种变化作出即时的评估，为各种实际应用提供更准确、更完备的支持。

参考文献

- [1]. 张普，关于大规模真实文本语料库的几点理论思考，《语言文字应用》，1999年第1期，北京。
- [2]. 张普，关于语感与流通度的思考，《语言教学与研究》，1999年第2期，北京。
- [3]. 张普，关于网络时代语言规划的思考。
- [4]. 孙维张，《汉语社会语言学》，贵州人民出版社，1994年。
- [5]. 中国新闻年鉴 1997年、1997年，北京。
- [6]. 中国出版年鉴 1997年、1998年，北京。