

# 从句子长度看新闻语体和小说语体

## —— 一个从统计得来的结果

吴云芳

### 一、小引

由于信息传递的目标、接受对象以及谈及内容的差异,使得不同的语体都烙上了“特别的”语言特征(linguistic features)。新闻语体面向公众发布消息,这使得它的行文比较正式、语法比较规范、词汇书面化。小说语体面向个人讲述故事,这使得它的行文比较随意、语法不很规范、词汇口语化。句子长度是语言特征的一个表现。长句信息丰富,表述严谨;短句活泼、明快。本文从句子长度这一角度出发,来考察新闻语体和小说语体的特征。

语体特征的分析需要一个比较的基点,只有在比较中我们才能发现哪些特征是这一语体所特有而别的语体所无的。因此,本文的分析建立在新闻语体和小说语体的对比上。

### 二、从统计来的数据

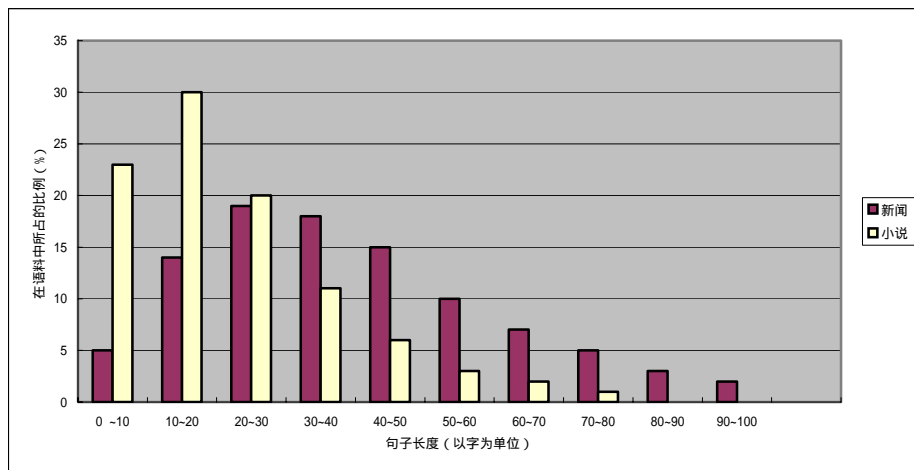
#### § 2.1. 对统计的一点说明

我们统计了新闻和小说两种语体中句子长度的分布状况。新闻语料包括了《人民日报》98年8月所有的内容,共1,980,275个字,48,882个句子。对新闻语料进行了一点预前处理,删除了“编者按”、“(记者xxx报道)”等这样的内容。小说语料是几个作家的作品的“杂交”,这样做是为了不受作家个人风格的影响。小说语料共215,843个字,9,881个句子,来源于:方方《埋伏》的第一节、第四节、第八节;苏童《妻妾成群》第二节、第五节、第六节;韩寒《三重门》第一节、第九节、第十四节;杨绛《洗澡》第一部第一章、第二部第六章、第三部第十二章;王安忆《我爱彼尔》第一节、第八节、第九节;冰心《斯人独憔悴》;王海鸰《牵手》中的部分;高行健《灵山》中的部分。

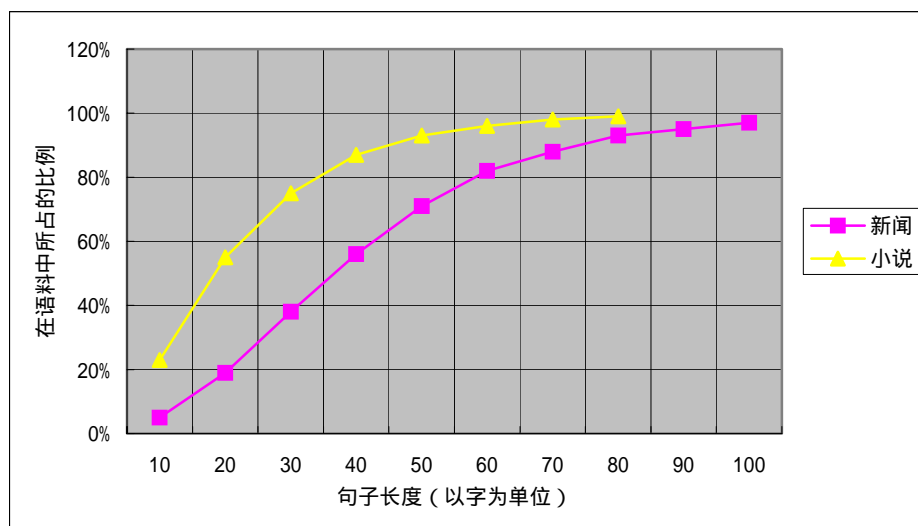
统计中,我们是以“。 ! ? ;”作为一个完整的句子结束的标志。考虑到“:”是一个较大的停顿,并且隔开的两个部分都是完整的句子形式,因此也作为了句子结束的标志。

#### § 2.2. 统计结果

图一:新闻和小说语料中句子长度分布的柱形图



图二：新闻和小说语料中句子长度分布的相对累加频率曲线



### 三、从数据来的结果及对结果的简单分析

#### § 3.1. 小说语料的平均句子长度小于新闻语料。

对于小说这种语体，其句子长度的分布集中在 0~10、10~20 和 20~30 这三个区间内。其中，长度在 10~20 的句子最多，约占三分之一。从标示的中位数可以看出，小说的相对累加频率曲线的中值约为 18，即长度小于 18 字和长度大于 18 字的句子分别占小说语料中句子总数的 50%。句子长度到达 80 字时，已占小说句子总数的 99% 强，基本上已经覆盖了所有的句子。

对于新闻这种语体，其句子长度的分布集中在 20~30、30~40 两个区间内，两个区间的句子数约占总数的 38%；如果加上稍微次之的 40~50 字的区间，三个区间的句子是总数的二分之一强。新闻的相对累加频率曲线的中值约为 40 字，即长度小于 40 字和长度大于 40 字的句子分别占新闻语料中句子总数的 50%。

#### § 3.2. 小说语料句子之所以短的原因

考察小说语料中 0~20 的句子，发现主要有两种类型：

一、由于我们把“：”作为了分割句子的单位，小说语料中经常出现“xxx 说”类，如：

对方赶着又问：

林雨翔望着父亲的背影，自言自语道：

二、对话，如：

“ 没来。”

“ 你以为山里人不会道歉？”

可见，对话的频繁出现是小说语体句子长度短的最直接的原因。这与我们的常识是相吻合的。小说是一种接近口语的叙事文体，人物的言谈是推进事件发展和刻画人物性格的主要手段。人们的对话稍纵即逝，是一种纯粹的“现场”的交流，这种交流不允许有过长

的句子，以及过于复杂的结构。所以，“对话”的频繁出现，最终导致了小说中的句子偏短。

### § 3.3. 新闻语料句子之所以长的原因

新闻语体是一种书面语体，主要是用来报道消息、告知情况。“对话”在其中基本不出现或出现很少。适应这种相对复杂的内容，客观上要求长句出现。因为长句可以蕴涵丰富的信息，表达复杂的逻辑关系。

这儿举一个例子予以说明：

在印尼发生严重侵害华人华侨生命财产事件以后，中国政府多次通过外交途径向印尼政府进行交涉，对华人华侨所受到的不公正待遇表示强烈关注和不安，要求印尼政府彻底查处，并采取有效措施，避免类似不幸事件的发生。

这个句子首先是一个时间状语“在……以后”，然后是一个主谓结构，再然后是一个省略主语的主谓结构，接下来又是一个省略主语的主谓结构，其宾语部分还是一个复杂结构。

§ 3.4. 新闻语体和小说语体句子长度分布上的差异反映了语言的语体变异 (register variation)，这种语体变异无论对教学还是写作，都有可资借鉴之处。

### 参考文献：

- [1] Douglas Biber, Susan Conrad, Randi Reppen. 2000. *Corpus Linguistics*. Foreign Language and Research Press, Cambridge University Press.
- [2] Laura Wright, Jonathan Hope. 2000. *Stylistics: A Practical Coursebook*. Foreign Language and Research Press, Routledge.
- [3] Anthony Woods, Paul Fletcher, Arthur Hughes. *Statistics in Language Studies* (中译本，陈小荷等. 2000. 《语言研究中的统计方法》. 北京语言文化大学出版社)