基于语料库的汉语同义词语体差异定量分析①

张文贤[1]　邱立坤[2]　宋作艳[3]　陈保亚[4]②
([1,2,4]北京大学，北京 100871; [3]北京师范大学，北京 100875)

[摘　要] 通过分析语体差异大的同义词，可以了解语体之间的差异。本文采用定量分析的方法，计算出 1343 对具有显著口语、书面语语体差异的同义词。通过对这些词对的调查分析可知: 词性上，语体差别最大的同义词中动词最多; 重叠、词缀、古汉语遗留词汇在同义词中所占的比重都较小; 若一对同义词有音节上的差异，则口语倾向于为单音节，书面语倾向于为双音节。本文的调查结果对语言教学以及教材编写都有一定的启示。

[关键词] 同义词; 语体差异; 定量分析; 对外汉语教学

[中图分类号] H195.1　　[文献标识码] A　　[文章编号] 1003 - 7365(2012)03 - 0072 - 09

零、引言

语体是学界一直探索的问题。近年来，不断有学者就划分语体的标准进行讨论，如刘大为(1994)、陶红印(1999)、方梅(2007)等。还有一些学者在语体中发现了汉语语法的规律，比如方梅(2000)，张伯江(2007)，王洪君等(2009)，王伟、周卫红(2005)，宋作艳、陶红印(2008)等。而关于口语与书面语在词汇上的差异的研究却相对较少，代表性成果为冯胜利、胡文泽(2005)等。口语、书面语的词汇差异到底有多大？除了方言词以外，哪些常用的词汇有语体差异？目前还缺少这方面的定量研究。

在对外汉语教学中，要使学生能够得体地表达，就要有语体意识。李泉(2004)认为，对外汉语教学的根本目的就是培养学习者准确地把握和正确地使用各种语体的能力。有些教材注明了词汇的语体差异。例如北京大学出版社出版的博雅系列汉语教材中《冲刺 2》在生词注解中就标明了某些词语有语体倾向，比如第 233 页: 揍(口语)，攒(口语); 第 235 页: 渐次(书面语)。《飞翔 1(使用手册)》第 5 页"深邃"常用于书面语，"深刻"口语、书面语都可以用，"提示"多用于书面语，"提醒"口语、书面语都常用。那么，到底哪些同义词有语体上的差异？当我们说某个词口语、书面语都可以用的时候，是不是意味着该词的语体差异不大？在判断一对

① [基金项目] 本文得到教育部人文社会科学研究青年基金项目"基于语篇与语体的连词主观性研究"(项目编号: 11YJC740145)以及国家自然科学基金项目"基于自消歧模式的语法知识自动获取技术研究"(项目编号: 61103089)的资助，谨致谢忱。

② [作者简介] 张文贤，女，北京大学对外汉语教育学院，讲师，博士，研究方向为语言学及应用语言学; 邱立坤，男，北京大学计算语言学研究所博士后，讲师，研究方向为计算语言学; 宋作艳，女，北京师范大学文学院，讲师，博士，研究方向为语言学及应用语言学; 陈保亚，男，北京大学中文系教授，博士研究生导师，北京大学中国语言学中心研究员，研究方向为语言学及应用语言学。

同义词的语体差异时,除了语感以外,还有没有其他依据?

程雨民(2004)指出"语体建立在同义性的基础上"。"语体的实质是在一些使用场合上有区别的同一变体的选择。"这样看来,比较词语的语体差异的最好办法就是看同义词之间的语体差异。本文通过考察同义词在不同语体中的分布,计算出它们的语体差异度,从而对同义词的语体差异进行一个定量化的描述,然后分析这些同义词词对之间的区别特征,最后说明同义词语体差异对教学的启示。

一、语料来源

尽管学者们对语体有不同的认识,划分出来的类别也不统一,本文还是采用了口语、书面语这一说法,因为这种划分最方便,在教学中也是最常用的。本文所使用的口语语料共计 149 万字,包括三部分:第一部分是电视情景剧《我爱我家》、《编辑部的故事》的对白,分别为 55 万字和 13 万字;第二部分是电视访谈节目《实话实说》、《对话》的对白,61 万字;第三部分是完全无准备的自然谈话,包括北京大学自然口语语料库,13 万字,北京语言大学口语语料,7 万字。书面语语料选取的是《人民日报》语料库(1998 年 1 月份的数据),总字数为 186 万字。① 该语料库经过分词和词性标注(人工校对),但是为了使口语语料库和书面语语料库具有一致的分词和词性标注,我们没有使用该语料库的分词和词性标注结果,而是使用同样的工具对口语和书面语语料库进行自动分词和词性标注。

本文使用的同义词词典为《同义词词林(扩展版)》。《同义词词林》由梅家驹等(1984)编纂而成,该词典按照树状的层次结构把所有收录的词条组织到一起,并把词汇分成五层类别,第一层有 12 个类,第二层有 97 个类,第三层有 1428 个类,第四层有 4223 个类,第五层有 17807 个类。哈尔滨工业大学信息检索技术研究中心参照多部电子词典资源,对之进行了大规模的扩展,其最终的词表包含 7 万余条词语,称之为《同义词词林(扩展版)》(以下简称《词林》)。② 本文所取的同义词为第五层,即,如果两个词在《词林》中属于同一个五层类,则我们视之为一对同义词。平均而言,一个同义词簇包含大约 5 个词。需要说明的是,《词林》中的同义词并不一定是严格意义上的同义词,有许多是近义词,在下文中我们会进一步分析这一事实对实验结果的影响。

二、同义词语体差异的计算方法

本文所使用的调查方法流程有如下几个步骤:(1)分别对口语语料和书面语语料进行分词和词性标注;(2)分别统计口语语料和书面语语料的带词性词频;(3)将绝对词频转换为相对词频;(4)遍历同义词词典中的第一对同义词,计算同义词的语体差异显著度;(5)根据所得的语义差异显著度对同义词进行排序,即可得出语体差异显著度比较大的词。下面我们将详细说明上述五个步骤的情况。

2.1 分词和词性标注

① 《人民日报》中也有部分语料不是书面语,但比例较小。本文按照词语的频率计算语体差异,应该能够反映书面语词汇的全貌。即使某一口语词语出现在《人民日报》中,频率也不会太高,不影响本文的结论。

② 《同义词词林(扩展版)》的相关说明请见 http://ir.hit.edu.cn/phpwebsite/index.php?module=pagemaster&PAGE_user_op=view_printable&PAGE_id=162&lay_quiet=1

本文的分析基于词语这一级单位进行,因此首先要对词语进行分词和词性标注的预处理。我们使用中国科学院计算技术研究所的 ICTCLAS 汉语分词系统对语料进行词语切分和词性标注,该软件所使用的词性标注体系与“人民日报”标注语料库相同,便于在词性一级直接对应。下面是一个经过词语切分和词性标注的示例: 迈向/v 充满/v 希望/n 的/u 新/a 世纪/n。词与词之间用两个空格隔开,词与词性之间用“/”隔开,“v、n、a、u”分别表示动词、名词、形容词和助词。

2.2 统计带词性词频

对语料库进行分词之后,我们就可以从语料库中统计出词语的频次。在汉语中存在大量的兼类词,一个词可能被标注多种词性,比如“希望”兼属于动词和名词,不同词性的词在语法意义上差别较大,应该作为不同的单位进行分析。通过词性标注,在一定程度上可以消解这种歧义,从而实现更细粒度的比较分析。

因此,我们在分词和词性标注的基础上统计出带词性的词频表,下面的表 1 和表 2 分别标明了书面语和口语语料库的带词性词频示例。

2.3 将绝对词频转换为相对词频

绝对词频容易受到语料库规模、分布不平衡等影响,一般不能直接用于语料库之间的词汇比较。因此,我们需要将绝对词频转换为相对词频,通过相对词频的差异来比较同义词词对在不同语体中的使用差异。

我们的转换方法是(以书面语语料库为例) : (1) 将书面语语料库带词性词频表按频次降序排列; (2) 将词频表等分为 1000 块,每一块的 ID 分别为由 1 到 1000; (3) 每一块中的所有词语的相对词频为 1001 - ID,即绝对频次最高的一块中相对词频为 1000,绝对频次最低的一块中相对词频为 1。依照相同的方法可以对口语语料库带词性词频进行转换。我们可以得到两个相对词频表(表 1、表 2) 。

表 1 书面语带词性词频示例

词/词性	频 次
的/u	60281
在/p	13078
和/c	12735
了/u	11384
是/v	11009
一/m	7621
有/v	5071
不/d	4942
对/p	4556
中/f	3677
为/p	3664
工作/v	3655
要/v	3638
上/f	3614
这/r	3546

表 2 口语带词性词频示例

词/词性	频 次
的/u	34780
我/r	29752
是/v	25469
不/d	17914
你/r	16095
就/d	13980
了/y	12309
这/r	11026
说/v	9630
有/v	9059
一/m	8629
也/d	7828
了/u	7627
他/r	7557
我们/r	7043

2.4 语体差异显著度

给定一对同义词 wi 和 wj,设两个词在口语语料库中的相对词频分别为 fsi 和 fsj,在书面语语料库中的相对词频分别为 fwi 和 fwj。我们用一个词语在口语和书面语两种语料库中的相对词频比来表示该词语的语体差异显著度,如公式(1)中 S_i 表示 wi 的相对词频比。然后,可以在两个词各自的语体差异显著度基础上计算两个词语之间的语体差异显著度,如公式(2)所示。其中 OS_{ij} 表示同义词 wi 和 wj 之间的语体差异显著度。

公式(1) $S_i = \frac{fs_i}{fw_i}$ 公式(2) $OS_{ij} = \frac{s_i - s_j}{s_i + s_j}$

由于我们所使用的语料库是没有进行词义消歧的,所以《词林》中的多义词在语料库中无法分开。在这种情况下,我们只能使用词性标注来消解一部分歧义。因此,在选择同义词词对时,除了要求在《词林》中处于同一个第五层类之外,还要求两个词的词性相同。

比如“今儿”和“今日”这一对词在《词林》中属于同一个第五层类(类 ID 为 Ca23A03),并且两者具有共同的词性标记“t”(时间词),因此我们可以将它们作为一对同义词来计算其语体差异显著度。“今儿”在口语和书面语语料库中的相对频次分别为 179 和 1,而“今日”的相对频次分别为 5 和 69,因此,我们可以计算出“今儿”和“今日”的语体差异显著度为 0.999。

我们在调查时把词类分为:名词(n)、动词(v)、副词(d)、代词(r)、形容词(a)、连词(c)、方位词(f)、介词(p)、数词(m)、语气词(y)、助词(u)等。与同一个词对应的可能有几个同义词,每一对的语体差异度可能不同。比如“要是”的同义词有“如果、假如、假使、倘若、若”等,在计算语体差异时,我们分别拿“要是”与“如果”,“要是”与“假如”,“要是”与“假使”,“要是”与“倘若”,“要是”与“若”进行对比。

三、实验结果与分析

3.1 实验结果

通过上述方法得到每一对同义词差异的具体数值,数值越大语体差异越大。统计显示,差异度大于 0.900 的同义词共有 2470 对。我们对这 2470 对进行了人工校对,删除不合格的同义词,得到 1343 对口语、书面语语体差异显著的同义词,算法正确率为 54%。

不合格的同义词的产生,主要是由于以下两种情况:第一,我们的语料库没有经过词义消歧,所以无法区分多义词的不同义项,这就使得对多义词的处理效果不好。比如“打”有很多义项,每个义项都有对应的同义词,调查的结果就显示“打_选购、打_发射、打_建造、打_砌”等的差异度都在 0.9 之上。第二,有一些词对只是同类而已,也被当作了同义词。比如量词、助词、表示职称或称呼类的名词等。出现错误的词对有“班_旅、班_组、丝_分、俩_两、呢_哉、啊_也罢、哎_咦、校长_院长、总裁_总统、满族_匈奴、姥姥_老妈妈、家长_县长、旗_市、奥运会_研讨会、年级_班组”,等等。

3.2 语体差异最大的同义词

表 3 语体差异大的前 100 对同义词的词性分布

实 词(54 对)				虚 词(46 对)					
动词 v	名词 n	代词 r	形容词 a	副词 d	连词 c	方位词 f	数词 m	介词 p	语气词 y
30	19	3	2	26	11	3	3	2	1

表 4　1330 对同义词的词性分布差异

实　词				虚　词				
动词 v	名词 n	形容词 a	代词 r	副词 d	连词 c	方位词 f	数词 m	介词 p
658	250	99	8	200	41	35	21	18

我们对语体差异度最高的前 100 对同义词,即差异度在 0.9838 以上的词汇的词性进行了统计,发现实词有 54 对,虚词有 46 对。其中,动词(30 对) 、名词(19 对,包括 4 对时间名词 t) 所占的数量较大。出乎意料的是,形容词居然只有两对,而且是同义词,它们是“不错 - 良好”、“好 - 良好”。虚词方面,同义副词的差别最大(26 对) ,其次为连词,有 11 对,再次为方位词(3 对) ,数词(3 对) 。

如果把统计范围扩大到 1330 对同义词,则数量大的词性的分布结果为上面的表 4,可以看出词性差异的排列次序几乎没有变化,仍然是实词里动词最多,其次为名词。虚词里最多的是副词,其次为连词,然后是方位词、数词、介词。也就是说,前 100 对语体差异大的词基本能够反映词汇语体差异在词性分布上的特点。

3.3 同义词语体差异的类型

我们进一步考察了语体差异大的 1343 对同义词,统计具有口语词汇特征的重叠、有词缀的词和具有书面语特征的古代汉语遗留词汇的出现情况,以及单双音节的对应情况等。

3.3.1 重叠与词缀

重叠与包含词缀“子、儿、头”的词汇为典型的口语词,但是数量并不多。口语为重叠式,书面语为非重叠式的同义词有 9 对,除了“慢慢_日益”①为副词外,其余 8 对均为动词。

表 5　重叠与非重叠式(按差异度排列)

同义词	差异度	同义词	差异度	同义词	差异度
慢慢_日益	0.998	问问_讯问	0.974	问问_咨询	0.974
想想_思索	0.974	听听_收听	0.971	谢谢_感谢	0.969
谈谈_座谈	0.964	看看_察看	0.961	想想_思考	0.918

含有词缀“子、儿、头”的同义词语体差别最大,除了“脑子_脑”、“味儿_滋味”外,其它的同义词全部都在 0.97 以上。“儿”是典型的口语词标记,表中带有“儿”的口语词有 8 个②,带“子”的有 5 个,带“头”的只有 1 个。

表 6　口语词带有词缀(按差异度排列)

同义词	差异度	同义词	差异度	同义词	差异度
事儿_事务	0.999	事儿_事宜	0.999	今儿_今日	0.999
女孩子_女童	0.998	那会儿_其时	0.997	样子_势头	0.997
这会儿_此时	0.996	一块儿_共同	0.996	一点儿_些许	0.991
片子_影片	0.990	外头_外侧	0.971	房子_房屋	0.968
脑子_脑	0.909	味儿_滋味	0.909		

3.3.2 古代汉语遗留词汇

书面语词汇是古代汉语在现代汉语中的遗留词汇。这样的词汇有 20 对,其中包括“比如说_诸如”、“乐意_甘于”、“叫_令”、“挨_倚”、“看_访”、“好像_恍若”、“瞧_觑”、“信_函”、“杀_

① “慢慢_日益”、“慢慢_日趋”、“慢慢_日渐”、“慢慢_渐”分别构成同义词,本表只列了一对。

② “一块儿”与“共同”、“一道”、“一齐”等分别构成同义词,本表只列了一个。

屠”9 对实词,其他 11 对为虚词。①

表 7 书面语词是古代汉语的遗留(按差异度排列)

同义词	差异度	同义词	差异度	同义词	差异度
挺_颇	0.999	跟_与	0.998	还_仍	0.997
比如说_诸如	0.997	就是说_即	0.994	要是_倘若	0.993
哪_焉	0.989	乐意_甘于	0.984	叫_令	0.982
挨_倚	0.979	看_访	0.976	没有_尚未	0.975
还_尚	0.974	好像_恍若	0.966	瞧_觑	0.960
你_汝	0.944	信_函	0.940	把_将	0.934
全_皆	0.901	杀_屠	0.900		

3.3.3 汉语口语与书面语单双音节的对应关系

我们对提取出来的口语词与书面语词差异度高的词对进行了考察,得到口语单音节、书面语双音节的 427 对,口语双音节、书面语单音节的 47 对,二者的比例是 9:1。可见,口语词单音节词的数量远远多于书面语。口语是单音节、书面语是双音节的词汇有两类:

一类是一个口语单音节词与几个书面语双音节词对应,如表 8 所示。

表 8 一个单音节词与多个双音节词对应(按差异度排列)

同义词	差异度	同义词	差异度	同义词	差异度
帮_扶助	0.984	办_设立	0.977	扔_抛弃	0.971
帮_协助	0.971	办_开设	0.964	扔_摈弃	0.971
帮_声援	0.957			扔_废弃	0.971

另一类是口语词是书面语词中的一个构成成分,如表 9 所示。

表 9 单音节词是多音节词的一个成分(按音序排列)

同义词	差异度	同义词	差异度	同义词	差异度
爱_喜爱	0.965	搬_搬迁	0.980	比_比照	0.957
必_必将	0.953	菜_蔬菜	0.971	带_带动	0.971
得_博得	0.959	得_得以	0.977	得_获得	0.974
得_获取	0.976	得_赢得	0.990	登_刊登	0.961
等_等待	0.966	等_等候	0.974	地_土地	0.956
对_针对	0.980	盖_覆盖	0.980	跟_跟随	0.967
好_友好	0.975	家_家园	0.985	交_提交	0.974
交_交纳	0.960	开_开办	0.963	开_开设	0.974
开_召开	0.961	忙_忙碌	0.973	忙_繁忙	0.978
签_签署	0.993	书_图书	0.979	算_计算	0.971
听_听取	0.979	推_推迟	0.966	推_推动	0.978
屋_房屋	0.982	写_编写	0.966	写_书写	0.959
写_撰写	0.957	行_施行	0.979	行_履行	0.996
演_上演	0.968	有_享有	0.969	找_查找	0.972
长_增长	0.985				

口语双音节、书面语单音节的主要是虚词,如下表 10 所示。

① 既是古代汉语遗留词汇,又与口语词有单双音节差异的,列在下文表 10,不在表 7 重复列出。

表 10　口语双音节书面语单音节(按差异度排列)

同义词	差异度	同义词	差异度	同义词	差异度
要是_如	0.999	就是_即	0.984	里面_内	0.980
因为_因	0.967	或者_或	0.965	所以_故	0.961
仍然_仍	0.958	肯定_必	0.936	已经_已	0.926
假如_若	0.909				

四、对对外汉语教学的启示

4.1 词汇选择与语体的关系

正如刘大为(1994) 指出的,语体类型先是由交际需要决定的,交际需要支配着我们选择不同的行为方式,不同的行为方式又会影响语言的形式变异,使交际者在进行某一语体的行为时倾向于选择某些语体特征。所以,口语、书面语选择的词汇体现了其语体特点。吴丽君(2004) 指出,口语体口语使用的是口语体词汇,书面语体书面语使用的是书面语体词汇,通用语体词汇包括书面语体口语词汇与口语体书面语词汇。本文统计出来的语体差异大的同义词应该分别属于口语体口语与书面语体书面语。

在对外汉语教学中,可以有针对性地加入语体知识,增强学生的语体意识。首先,可以从总体上讲明语体与词汇的关系。我们从表 1、表 2 中可以看出口语与书面语在用词上的一些差别,比如口语中人称代词“我、你、他、我们”使用较多; 若同义词有单双音节的差异时,实词在口语中倾向于使用单音节,在书面语中倾向于使用双音节。其次,可以根据本文的结论设计一些语体转换的练习,教会学生表达同一个意思时,书面语用什么词汇,口语用什么词汇。比如,在教学中可以设计这样的练习:

把下面的口语性强的句子改为书面语强的句子:

①你听听新闻就知道了,我们的国家慢慢地强了。

→你收听新闻就知道了,我们国家的实力日益增强了。

②我们叫这座山为“天下第一山”。

→这座山被誉为“天下第一山”。

4.2 单双音节词汇与语体的关系

单双音节的不同也是口语与书面语词汇差异的一个重要类型。在对外汉语教学中,谈到词汇的语体差异时,我们常说“所以”是口语的,“故”是书面语的,这时给留学生的印象是,口语表达倾向于双音节,书面语表达倾向于单音节。但有时却会使留学生得出完全相反的结论,比如我们告诉他“买”是口语的,“购买”是书面语。本文总结出了哪些词汇在口语中是单音节,而在书面语中是双音节,哪些词汇相反。在 1343 对同义词中,口语单音节、书面语双音节的是 427 对,口语双音节、书面语单音节的是 47 对。也就是说,有音节差异的词对共有 474 对,占 35%。除了少数同义词是单音节对单音节之外,大多数同义词还是双音节对双音节。

口语单音节、书面语双音节的实词居多,口语双音节、书面语单音节的虚词居多。这是因为这些虚词是对古汉语的继承,因为文言化而显得书面化。一般认为,汉语词汇的发展经历了一个从古代汉语单音节向现代汉语双音节发展的过程,现代汉语以双音节为主。张国宪(1996) 、汤志祥(2001) 都有统计数据证明这一点。曹炜(2003) 以《现代汉语词典》所收录词语为研究对象,得出的结论是“从音节结构来看,口语词中双音节词占绝对优势,而书面语词

中,单音节词的数量逼近双音节词的数量。”根据现代汉语的实际语料,这一结论需要再验证。也有一些学者观察到单音节在现代汉语中占有相当的地位,如刁晏斌(2006)、蔡长虹(2007)等。什么是口语词?什么是书面语词?口语词与俚语、方言词怎么区分?如果仅仅以词典或词表为依据,不可避免地会将口语词与其它词语纠缠在一起。正如苏新春、顾江萍(2004)所指出的那样,口语词难以确定,一方面是因为与方言词纠缠在一起,另一方面是因为口语词自身所发生的变化。本文从实际语料出发,研究普通话中常用的词汇,说明口语与书面语词汇在音节上确实存在差异,但这种差异的比例并未过半。

4.3 对外汉语教学中的同义词辨析视角

同义词教学是对外汉语教学中的难点,赵新、李英(2001)、杨寄洲(2004)、敖桂华(2008)、张博(2008)、苏英霞(2010)、田惠刚(2010)等都从不同角度论述过同义词辨析问题。目前的同义词辨析主要是从以下角度来进行的:(1)词义的侧重点、轻重、范围;(2)词义的褒贬色彩;(3)与该词时常搭配使用的词;(4)语体差异。在这些常用的辨析方法中,语体特征究竟占据什么样的地位?前人大多认为语体差异在同义词辨析中占有很重要的地位,比如吴丽君(2004)认为“近义词间的差别除了功能和使用范围以外,很重要的一点就是适用语体的不同”。田惠刚(2010)认为“语体是同义词最重要的特征之一”。但是这些结论多来自感性的认识,论证是举例性质的,没有统计方面的证据,更没有指出语体差异的程度到底有多大。

通过本文的调查分析,可以给同义词的语体特征一个更加准确的定位。首先,从整体上看,同义词的语体差异并不一定是同义词最主要的差异。第二,语体差异只是同义词辨析的一个方面,但不一定是最主要的方面,因为只有部分同义词有语体差异。第三,即使一对同义词有语体差异,也不能简单地用音节、词缀等因素来解释。音节、词缀等只能用来解释少部分同义词的差异。第四,虽然语体差异是词汇、语法等综合作用的结果,要培养学生得体地运用语言的能力要考虑多种因素,但是毫无疑问,同义词是一个很好的切入点。在这些词上多做训练,能够增强留学生的语体意识。对于语体差异大的动词、副词、名词和连词,在教学时更要特别给予重视。

五、结语

本文通过计算的方法得到了口语、书面语实际语料中有语体差异的同义词。口语、书面语的同义词差异主要在词性以及音节上。动词、名词、副词、连词为语体差异大的词类。口语中单音节词多于书面语。本文的成果主要在于:首先,得到一张口语、书面语语体差异大的同义词的列表,从词汇的角度考察了汉语的语体问题。其次,描写了口语、书面语在词汇特征上的差异,这使我们对语体的差异有了更深入的认识。第三,本文的结论为对外汉语教学以及教材编写提供了参考依据。

参考文献:

[1] 敖桂华. 对外汉语近义词辨析教学对策[J]. 汉语学习,2008,(3).
[2] 蔡长虹. 当代汉语词汇的单音节化现象考察[J]. 汉语学报,2007,(1).
[3] 曹 炜. 现代汉语口语词和书面语词的差异初探[J]. 语言教学与研究,2003,(6).
[4] 程雨民. 英语语体学[M]. 上海:上海外语教育出版社,2004.
[5] 刁晏斌. 现代汉语词的音节及其发展变化[A]. 南开语言学刊(第 1 期)[C]. 北京:商务印书馆,2006.

[6] 方　梅. 自然口语中弱化连词的话语标记功能 [J]. 中国语文,2000,(5).
[7] 方　梅. 语体动因对句法的塑造 [J]. 修辞学习,2007,(6).
[8] 冯胜利,胡文泽. 对外汉语书面语教学与研究的最新发展 [C]. 北京: 北京语言大学出版社,2005.
[9] 冯胜利. 汉语书面用语初编 [M]. 北京: 北京语言大学出版社,2006.
[10] 李　泉. 面向对外汉语教学的语体研究的范围和内容 [J]. 汉语学习,2004,(1).
[11] 李绍林. 对外汉语教学词义辨析的对象和原则 [J]. 世界汉语教学,2010,(3).
[12] 刘大为. 语体是言语行为的类型 [J]. 修辞学习,1994,(3).
[13] 宋作艳,陶红印. 汉英因果复句顺序的话语分析与比较 [J]. 汉语学报,2008,(4).
[14] 苏新春,顾江萍. 确定"口语词"的难点与对策——对《现汉》取消"口"标注的思考 [J]. 辞书研究,2004,(2).
[15] 苏英霞. 汉语学习者易混淆虚词的辨析视角 [J]. 汉语学习,2010,(2).
[16] 汤志祥. 当代汉语词语的共识状况及其嬗变 [M]. 上海: 复旦大学出版社,2001.
[17] 陶红印. 试论语体分类的语法学意义 [J]. 当代语言学,1999,(3).
[18] 田惠刚. 双语体同义词语的探索及其教学实践意义 [J]. 云南师范大学学报,2010,(1).
[19] 王福生. 对外汉语教学活动中口语和书面语词汇等级的划界问题 [A]. 汉语口语与书面语教学——2002 年国际汉语教学学术研讨会论文集 [C]. 北京: 北京大学出版社,2002.
[20] 王洪君,李榕,乐耀. 何时用"了$_2$"? ——兼论话主显身的主观近距交互式语体 [A]. 语言学论丛(第四十辑 [C]. 北京: 商务印书馆,2009.
[21] 王　伟,周卫红. "然后"一词在现代汉语口语中使用范围的扩大及其机制 [J]. 汉语学习,2005,(4).
[22] 吴丽君. 口语词汇与书面语词汇教学研究 [J]. 云南师范大学学报,2004,(3).
[23] 杨寄洲. 课堂教学中怎么进行近义词语用法对比 [J]. 世界汉语教学,2004,(3).
[24] 张　博. 第二语言学习者汉语中介语易混淆词及其研究方法 [J]. 语言教学与研究,2008,(6).
[25] 张伯江. 语体差异和语法规律 [J]. 修辞学习,2007,(2).
[26] 张国宪. 单双音节形容词的选择性差异 [J]. 汉语学习,1996,(3).
[27] 赵　新,李英. 对外汉语教学中的同义词辨析 [J]. 暨南大学华文学院学报,2001,(2).

Corpus – based Quantitative Analysis on Stylistic Difference of Chinese Synonyms

ZHANG Wen-xian[1] & QIU Li-kun[2] & SONG Zuo-yan[3] & CHEN Bao-ya[4]
([1,2,4]*Peking University, Beijing* 100871; [3]*Beijing Normal University, Beijing* 100875)

Abstract: By analyzing the differences in synonyms, we can understand the stylistic differences. In this paper, certain algorithms are used to calculate 1343 pairs of synonyms which show significant stylistic differences. From quantitative analysis on these pairs, we conclude that: most words in the pairs are verbs; overlapping, affixes, ancient Chinese words are small in proportion; if one pair of synonyms have different syllables, the spoken word tends to be monosyllabic and the written word disyllabic. These conclusions can provide some inspiration for language teaching.

Key words: synonym; stylistic difference; quantitative analysis; teaching Chinese as a second language