

面向语体自动识别的现代汉语副词语体分类



邵田, 荀恩东*

北京语言大学信息科学学院, 北京 100083

*通信作者: 荀恩东, edxun@126.com

摘要: 在自然语言理解中, 一般通过词汇与句式的特点来识别不同的语体。但是, 当前语言学界对于语体的相关研究主要集中于汉语语体体系的构建及其特征, 还未涉及到对具体的词语进行语体类别的划分, 存在理论与应用不对接的问题。在语言学界, 通常从语义的角度出发对副词进行分类, 本文则面向语体的自动识别, 从语体的角度出发对副词进行分类。首先, 将汉语语体看成是一个有层次的系统, 分为正式语体和非正式语体两类, 正式语体包括典雅语体与书面正式语体, 非正式语体包括普通话口语体与方言语体; 并提出了汉语副词语体划分的标准有: 构词特点、韵律、形态、频次, 同时提出了“跨语体副词”的概念并对其进行了介绍。不仅可以为汉语其他词类的语体划分提供参考, 也可以为汉语语体的自动识别提供数据基础, 进一步提高语体自动识别的准确率。

关键词: 副词; 语体分类; 分类标准; 跨语体副词

DOI: [10.57237/j.cll.2022.01.002](https://doi.org/10.57237/j.cll.2022.01.002)

The Stylistic Classification of Modern Chinese Adverbs for Automatic Style Recognition

Tian Shao, Endong Xun*

School of Information Science, Beijing Language and Culture University, Beijing 100083, China

*Corresponding author: Endong Xun, edxun@126.com

Abstract: In natural language understanding, different styles are generally identified by the characteristics of words and sentence patterns. However, the current research on style in linguistics mainly focuses on the construction of the Chinese style system and its characteristics. It has not yet involved the classification of style for specific words, which has the problem of mismatch between theory and application. In the linguistic field, adverbs are generally classified from the perspective of semantics, while this paper is oriented toward the automatic identification of style. Therefore, we classified the adverbs from the perspective of style. Firstly, the modern Chinese style is regarded as a hierarchical system, and we divide it into formal and informal styles, with formal style including elegant style and written formal style, and informal style including spoken mandarin style and dialect style. Secondly, we proposed the criteria for classifying Chinese adverbial style, and they are composition characteristics, rhyme, morphology, and frequency. Finally, we also proposed the concept of "cross-style adverbs" and introduced it. It can not only provide a reference for the corpora classification of other Chinese words but also provide a database for the automatic recognition of Chinese corpora, and further improve the accuracy of the automatic recognition of corpora.

1 引言

- 1) 斯巴达人都特别勇敢善战。—1926
- 2) 特别都是贫苦的女儿，学习积极性非常高。—9
- 3) 皆特别—0
- 4) 特别皆—0

此外,学界关于语体的研究主要集中于语体本身的定义、特征、分类体系、生成机制、语体分类的意义以及不同语体之间的差异等的研究,比如李熙宗对语体定义的几个问题展开了详细的论述[2],方梅论述了语体的特征及句法表现[3],冯胜利论述了语体的分类体系与生成机制[4],陶红印论述了语体分类的语法学意义[5],张伯江详细说明了不同语体之间的差异[6],方梅论述了语体的动因及其对句法的影响[7]。近些年,以冯胜利为代表的学者提出了语体语法的概念并展开了相关研究,比如冯胜利介绍了语体语法及其文学功能[8],并论述了语体语法对语言中“形式-功能”对应律的影响[9],施春宏进一步论述了语体何以作为语法[10]。但是学界缺少对词语的语体分类的相关研究,存在理

本文的内容安排大致如下：首先，介绍语言学界对汉语语体体系划分的不同观点，据此构建本文的汉语语体体系；其次，依据冯胜利提出的汉语语体的鉴定方法并结合相关研究，针对副词这一词类，提出汉语副词的语体分类标准[10]；据此标准，并以 TianShao 所构建的副词词表（下文简称为副词词表）为数据基础，对汉语副词进行语体划分，得到根据语体分类后的副词词表，并对副词语体的分类结果进行统计分析[11]。

2 汉语语体体系的构建

典雅 (古代词语)

随便 (俚语) 通俗白话 正式 (现代汉语书面)

图1 冯胜利(2010)汉语语体体系

但是冯胜利根据交际距离说把语体的定义修改为：“实现人类直接交际中最基元属性（亦即确定彼此之间关系和距离）的语言手段和机制”，并把冯胜利提出的汉语语体的三分系统修改为：俗常体、正式体与庄典体。其中，俗常体与口语体相对应，正式体与书面语体相对应，庄典体与文言语体相对应。可以明显看出，冯胜利对汉语语体观有一定的修改，即取消了俚语这一语体类[1]。

然而，对汉语语体体系的划分并非只有一家之言，《现代汉语词典》（第七版）在收录词语的时候就对一些词语的语体属性进行了标注，虽然没有明确提出汉

语语体的划分体系，但是通过对词典中标注语体的词语的统计分析来看，其语体大致分为：方言、书面和口语[12]。与冯胜利的对语体的分类不同，冯胜利根据时代的不同将书面语体细分为正式语体和典雅语体，而《现代汉语词典》（第七版）则根据地域因素将口语语体细分为方言和口语。造成不同分类结果的原因是二者的分类角度不同，冯胜利是根据交际距离说再糅合时代的因素，即语用因素与历时因素的结合；而《现代汉语词典》（第七版）则是根据使用人群的不同以及地域因素而划分出来的，主要从共时的角度出发[1, 12]。

同时《现代汉语词典》（第七版）把具有这些语体特点的词语标示了出来，通过对《现代汉语词典》（第七版）中所收录的副词的语体进行排查之后，统计结果见表 1。

表 1 《现代汉语词典》（第七版）标示出副词语体的统计

语体	书面语	方言	口语	汇总
数量	161	82	34	277
占比	14.9%	7.6%	3.1%	25%

据统计，《现代汉语词典》（第七版）共收录了 1081 个副词，但是并没有全部对这些副词进行语体的标注，上表分别列出了标明不同语体的副词数量以及占标明总数的百分比，可以看出，对副词语体的标注仅占副词收录总数的 25.6%。

由于二者分的角度与类别不同，有细分、有粗分，本研究综合《现代汉语词典》（第七版）与冯胜利[1]对语体的分类的结果，利用共时与历时结合的方法，加以语用因素和地域因素之后，本文的语体体系分类结果如表 2 所示。

表 2 汉语语体分类体系

汉语语体分类体系			
非正式语体		正式语体	
普通话口语体	方言语体	书面正式语体	典雅语体

此分类体系并没有沿袭冯胜利的三分法，而是对冯胜利语体分类体系进行了一定的调整，认为汉语语体是一个具有层级性的系统，是共时与历时的结合，同时也是语用因素与地域因素的结合[1, 3]。参考《现代汉语词典》（第七版）中把非正式语体划分为普通话口语体和方言语体的观点，并根据冯胜利对语体的论述，语体的划分是根据交际距离的不同[1, 12]。因此，对一个普通人来说，他在面对一个与之来自同一个地区的人时，会说家乡话以示亲近，说明了划分方言语

体的必要性；而在面对来自其他地区的人时，会说普通话口语体，以便交流，如果一个来自广东粤语区的人与一个来自河南晋语区的人用各自的方言沟通，恐怕也是不太可能发生的，此时，双方都会选用一个中间的状态，即普通话口语体进行交流。

3 副词语体分类标准

在明确了汉语语体的分类体系之后，就要进一步明确分类标准，分类标准是否合理直接影响着分类结果的准确性。冯胜利提出：书面语有一套自己的词汇[1]。相应，口语语体也有一套自己的词汇，本文便以口语词典为参考依据，同时以冯胜利提出的鉴定语体的两种方法为理论指导，根据汉语副词的特点，总结出副词语体划分的标准大致有：构词特点、韵律、形态、频次，下文分别对这几类分类标准进行详细的介绍。

3.1 构词特点

构词特点分为两部分：一是词汇本身具有某种语体的特点，二是由具有某种语体特点的语素构成的合成词具有该语素的语体特点。本文在根据语体对副词进行分类时，以施光亨为参考[13]，遗憾的是，其所收录的副词个数寥寥无几，但是通过对其所收录的副词进行分析时受到启发：如果一个合成词是由口语语素构成的话，那么这个词有很大的可能是口语词，同时这一点冯胜利也有提到：词缀也有语体属性[1]。

如果一个词语是由古汉语演化而来的或者是直接借用的古汉语的词汇，当然就具有典雅语体的特点，比如：殊、颇、郅、煞、綦、愈等，这些词除去是从古汉语中演化而来的之外，还多是单音节的。另外一种情况是如果一种词语来源于方言，自然也就属于方言语体，比如：齁、贼、溜等。

另外，一些由具有某种语体特点的语素构成的合成词也具有该语素的语体特点，由“何、愈、为、无、未、岂”等构成的合成词具有典雅语体的特点。由“何”构成的合成词有“何其、何等”等；由“愈”构成的合成词有“愈发、愈为”等；由“为”构成的合成词有“尤为、甚为”等；由“无”构成的合成词有“无缘、无宁”等；由“未”构成的合成词有“未始、未曾”等；由“岂”构成的合成词有“岂非、岂止”等。

由“老”构成的合成词属于非正式语体，比如表示高频的“老是”属于普通话口语体，表示程度深的“老大”属于方

言语体；由“子”构成的合成词也属于非正式语体，比如表示统括性的“一塌刮子”属于方言语体，表示时间的“一下子”属于普通话口语体。还有由后缀“的”构成的副词，一般也属于非正式语体，比如：乍猛的、可可儿的，都属于典型的方言语体词汇，但是，典雅语体“端的”除外，这里的“的”不属于后缀，而是由古时构词法构成的合成词。

3.2 韵律

冯胜利[1]提出的两点语体鉴别方法分别是：交际中的时空鉴定法与语法中的时空鉴定法。交际中的时空鉴定法，即“看哪个句子是谁说的、说给谁的（人）、在什么场合说的（地）、说的是什么内容（事）、听话者的态度如何（意）”。涉及到交际就意味着这种鉴定方法受语用因素的影响较大，在具体的分类过程中的主观性因素就比较大。语法中的时空鉴定法，即“从语法结构标记时空的机制上看（如指示词、量词、单复数、体态标记等），语法系统可以很好地表现‘时空标记’和‘语体属性’之间的对应关系”，并进一步提出：“语体可以通过语法和词义中的时空义素来鉴定。单双音节词汇的不同，不仅是韵律的对立，而且是语体的对立；不仅是语体的不同，而且是语法的不同”。

由此，把语体分类的标准大致概括为“韵律”因素，

即韵律反映了语体，语体反映了语法。

表 3 副词音节数的统计

音节数	单音节	双音节	三音节及以上
数量	359	939	116
占比	25.4%	66.4%	8.2%

表 3 是对副词词表的音节数的统计，可以看出，双音节的副词义项占据绝大多数，单音节次之，三音节及以上的最少。另外，通过对《现代汉语词典》（第七版）对副词的语体标注情况进行相应的考察之后，发现方言语体与普通话口语体中收录的三音节、四音节的副词很多，比如“冷不丁、一塌刮子”等等，而在正式语体中，除了表示揣测的“庶几乎”这一个三音节的副词之外，其余的副词都是单音节或双音节的。

同时，还对《现代汉语词典》（第七版）中标注语体的副词进行了音节状况的统计，并在此基础上，根据每一语体中音节数在其语体类的内部以及外部所占的比例情况进行了统计。其中“内部占比”指的是每一音节数的副词在每一语体内部所占的比例，“外部占比”指的是每一语体类的不同音节数的副词占所收录的这一音节总数的比例。统计结果大致如表 4 所示：

表 4 语体音节数的内部占比及外部占比

音节数	内部占比			外部占比		
	方言	口语	书面	方言	口语	书面
单音节	16.9%	20.7%	59.4%	12.1%	6%	81.9%
双音节	55.4%	38.2%	40%	37.4%	10.6%	52%
三音节	20.5%	38.2%	0.6%	54.8%	41.9%	0.3%
四音节	7.2%	2.9%	0	85.7%	14.3%	0

由表 4 中的“内部占比”可知，不同音节的副词在各个语体中的分布情况还是有一定的差异的，在方言语体内部，双音节的副词占据一半以上；在口语体内部，双音节的副词和三音节的副词所占比例相同，其次是单音节的副词，最低的是四音节的副词；在书面语体内部，主要分布在单音节和双音节这两类中，而三音节和四音节的副词则几乎没有。由此可见，韵律作为语体的分类标准是有事实证明的。

音节数与最省力原则有关，但是由表 4 中的“外部占比”可知，在单音节这一层面，书面语占比高达 81.9%，而口语语体仅占 6%；在双音节这一层面，书面语体仍占据一半以上的比例，其次较高的是方言语体，最低的则是口语语体；而在三音节这一层面，方言语体和

口语语体的总和已经高达 99.7%；同样在四音节这一层面，方言所占比例高达 85.7%，书面语则占 0%。因此便假设视觉信号无噪声，而方言与口语语体在语音传输中，产生的噪声多，需要用多音节提供的信息以抗噪。

据此，作出预测：首先，如果一个副词是四音节的，就不能分到书面语体和典雅语体中，只能分到方言语体和口语语体中，而分到方言语体的可能性又高于分到口语语体的可能性；其次，如果一个副词是三音节的，那么它被分到书面语体和典雅语体的可能性相当小，而被分为方言语体和口语语体的可能性较大；如果一个副词是双音节的，那么较大比例是属于书面语体的，其次是方言语体和口语语体；而如果一个副

词是单音节的,它被划分为书面语的概率要大大高于口语语体和方言语体。

这种设想当然是最理想的状态,但是语言丰富多变,根据这一预测对副词进行语体的划分时发现,还有许多副词,单纯依靠音节上的特点,无法准确地判定它们到底是属于哪一语体的,这就需要找寻其他形式上更加容易判定的标准。但不可否认的是,这一判定标准在判定三音节和四音节的副词的语体时,还是发挥了比较有效的作用,比如:左不过、一塌刮子等都属于非正式语体的方言语体;怪不得、可不是等属于非正式语体的口语语体。

除此之外,据高艳总结的口语词的十三个特点,其中有一点是“轻声”,通过分析副词词表发现,像“老大(表示程度深)、大半(表示限定)、到头来、短不了”等等这些具有轻声的副词,都属于非正式语体[14]。

3.3 形态

本文采用的形态是广义上的形态,包括儿化、重叠等等。在对副词进行语体划分的过程中,韵律标准虽然对双音节以上的词汇起到了有效的指导作用,但是对双音节和单音节副词语体的划分显得有些无能为力。因此针对单双音节的副词就需要更加明确的形式上的判定标准。

从另一个角度看,如果想要找出明确的副词语体划分的标准,那就需要分别找出各个语体的特点,找出每个语体的典型特点之后,再根据这些典型特点对副词的语体进行分类。现代汉语非正式语体的特点最明显也最容易区分,现代汉语非正式语体有两大典型特点,即儿化现象和重叠现象。因此,如果一个副词具有儿化或者重叠特点,那么就可以认为它属于非正式语体。

儿化相对于重叠更容易识别,如果一个词具有儿化的特点,那么就认为这个词属于非正式语体的词汇,经统计,一共有40个具有儿化特点的副词属于非正式语体的词汇,比如:倍儿等。

由于重叠式的副词并不多,经分析,重叠式的副词有:偏偏等,共36个,这些重叠式的副词都属于非正式语体。但是,其中也有一些例外情况,比如:比比、切切、断断、冉冉、垂垂、窃窃、在在、款款则属于正式语体,对于这类从古汉语沿用下来的重叠式副词,本文首先依据构词特点将其划分为正式语体的典雅语体。

3.4 频次

在实际的分类过程中,除了参考词汇、韵律、儿化及重叠这些形式上的标准之外,还有许多副词的语体特点不那么明显,基于这种情况,便根据副词在实际运用中的表现,即在大规模数据库中考察相关的语料,并参考其在各类文体中出现的频次来对副词进行语体划分。

找出副词在不同语体中出现的频次,首先,就要对目前语料库中构建的文体进行语体分类,本文主要参考北京语言大学语料库(BCC)中汉语文体的划分结果:微博、小说、科技、报刊和古汉语这五类[15]。其中,古汉语这一类顾名思义属于典雅语体,无论是在用词上,还是在时间跨度上,都应该把古汉语划分到正式语体的典雅语体这一类中,即如果一个词语属于古汉语词汇,那么就可以认定它属于典雅语体。而科技与报刊则是没有具体的对话场景的,其面向的是大众群体,并没有具体的对话人,同时,在用词上也倾向于正式化,因此,把科技与报刊划分到正式语体的书面正式语体中。微博则是日常生活中经常用到的一种文体,而小说同样也是描述了人们日常生活中的话语,这两种文体的共同特点是都有具体的对话场景,并且都有具体的对话人,因此,把这两种文体归到非正式语体中。

但是在语料的抽取过程中,古汉语这一语体是没有经过分词词性标注的,原因是现在的分词词性标注技术主要针对现代汉语,古汉语的分词技术或许还没有达到一个较高的准确度,因此抽取结果便只有文学、微博、科技和报刊四种。通过对在这几种文体中副词出现的频次进行统计发现,高频部分出现的副词几乎相差无几,其中,在四种文体中频次高居第一位的都是“不”。下文记录了在这四种文体中频次排名前20的副词。

科技:不、也、就、都、还、将、已、又、更、最、必须、很、没有、均、再、已经、进一步、却、未、比较

报刊:不、也、都、就、还、将、更、很、已、又、最、已经、没有、必须、正在、进一步、曾、特别、却、在

文学:不、也、就、都、又、还、很、没有、已经、却、再、没、只、才、更、便、太、最、已、真

微博:不、都、就、偷、太、很、最、好、也、还、又、不要、没、其实、再、更、没有、真、别、终于

通过分析上表发现,在这四个文体排名前二十的副词排序中,有十一个副词是一样的,分别是:不、也、都、就、还、更、很、又、最、没有、再。从一个侧面说明,它们具有多种语体的性质。

最高频段的副词一般相差不大,随着频次降低各个文体的副词逐渐分化,语体分化一般表现在频次排名比较靠后的副词中,因此,便对以上四种文体频次排名后二十的副词进行分析,如下文所示。

科技:保不住、万般、永世、信手、顺序、活活、源源、决然、纵情、好歹、压根儿、拼死、何苦、尽数、毋庸、顺次、莫非、甬、立时、特为

报刊:万般、兴许、忽地、日臻、一窝蜂、何须、来回、拦腰、略微、美美、何苦、必得、好歹、确乎、徒然、但凡、另外、死活、徒手、怪

文学:切切、按期、统共、殆、着意、冉冉、挨个儿、照实、定然、倍加、按理、倏然、既、信手、日见、无怪、断断、大举、终年、偏巧

微博:必得、忽地、不仅、渐次、连声、至多、终年、贸然、冷不防、径直、骤、交互、按理、次第、斗胆、迭、压根儿、下意识、临了、改日

四种文体频次排名后二十的副词语体分化比较明显,同时,在科技、报刊这样的正式语体中,即使是频次最低的副词也很少使用非正式语体的副词,然而,文学、微博等非正式语体的表现却不同,有些副词明显是正式语体的,比如文学中的“殆”,这种情况也反映出一种语言现象,即在非正式语体中,有可能暂时借用正式语体的词语以表达某种语用效果:或者是彰显说话者的文化程度、或者是为了使听话者更好地理解说话者所表达的意思等等;但是在正式语体表达中,借用非正式语体的概率相对来说就较小。

4 跨语体副词

在副词语体的具体划分过程中,有一些副词可以同时具有多种语体的性质,本文称之为“跨语体副词”,跨语体副词可以与多个语体的副词连用。

但是,跨语体副词也分为两种情况:一种是跨小类的语体,即一个副词跨非正式语体的两个小类或者属于跨正式语体的两个小类,它可以与同一大类的不同小类的副词连用,比如“极”是属于典雅体和书面正式体的副词;还有一种是跨正式语体与非正式语体两个大类的副词,它可以与不同大类的副词连用,比如“最”,无论是与正式语体还是在非正式语体的副词都可以连用。

汉语的语体划分就如同词类划分一样,汉语词类有兼类词,汉语的语体也有跨语体的词,虽然二者的叫法不同,但实质却是相同的,都是指一个词可以跨几个类,或者兼几个语体。同时汉语语体与词类相同的一点是,用认知语言学理论来解释即原型范畴,也就是说,在词类中,有典型词语,有非典型词语,而在汉语语体词汇中,有典型语体词汇,也有非典型语体词汇。这与语法的普遍性原则相一致。

5 副词语体分类结果统计与分析

通过以上几种分类标准对副词词表进行语体划分之后,每种语体所占副词词表总个数及相应的比例的情况如图2所示:

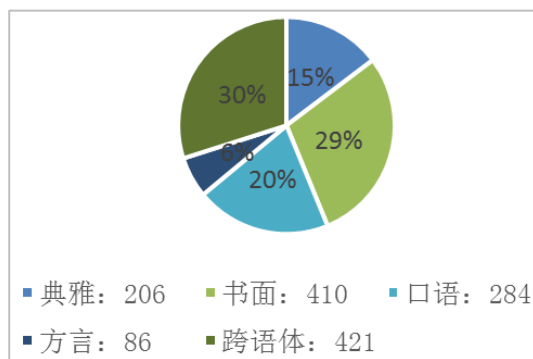


图2 各语体占比

由上图可知:书面语体和跨语体副词所占比例最高,其次是口语语体、典雅语体,方言语体所占比例最低。其中,典雅语体有:煞、颇为等共206个;书面语体有:全然、倍加等共410个;口语语体有:有点儿、刚刚等共284个;方言语体有:猛不丁、可可儿的等共86个;跨语体副词有:不、最等共421个。跨语体副词这一类别所占的比例将近三分之一,从一个侧面反映出跨语体副词使用的情况还是很常见的,同时类比兼类词的使用,除去停用词外,一般情况下使用频率越高的词,是兼类词的可能性越大,兼有的词类越多。

6 结语

判定一个副词的语体归属,一般情况下,可以从三方面进行判断:一是从整体功能上,本文主要探讨副词语体的划分,副词一般情况下都是作状语的,这一点的区分性不大;二是从内部构成上,比如构词特点、韵律、形态等等,另外还可以根据构成副词的词

汇来源，这在一定程度上影响着该副词的语体归属；三是从语用上，即该词运用的典型语境，主要参考该词在不同文体中出现的频次。

本文主要利用后两种判定标准来对副词进行语体分类，但是这几种方法也具有顺序上的优选性，即先以构词特点为首要判定标准，比如“比比、窃窃”，它们首先在构词特点上属于古汉语词汇，因此本文就将其划分为正式语体，而不是根据是否具有重叠特点将其归为非正式语体；其次是是否具有儿化、重叠等特点；最后再参考副词在不同语体中出现的频次，并综合其在语料库中出现的语境进行辅助判定。韵律则主要在用来判定三音节或者多音节的副词的语体时起到了有效的指导作用。

通过对副词进行语体的划分，找出了可以连用与不可以连用的副词词表，弥补了自然语言处理中的数据统计与分析只能找出存在于语言事实中的实例，而不能找出不存在的实例的缺点，为汉语其他词类的语体划分提供一定的参考，同时也为计算机辅助语言学习与汉语篇章语体的自动识别提供一定的参考。

致谢

本文为国家语委一般项目《面向汉语语义计算的意合语法研究与资源建设》(ZDII35-114)与北京语言大学研究生创新基金资助项目(中央高校基本科研业务费专项资金)(22YCX026)的阶段性成果之一。

参考文献

- [1] 冯胜利. 《汉语语体语法概论》[M]. 北京语言大学出版社, 2018.
- [2] 李熙宗. 关于语体的定义问题[J]. 复旦学报(社会科学版), 2005 (03): 176-186+196.
- [3] 方梅. 谈语体特征的句法表现[J]. 当代修辞学, 2013 (02): 9-16. DOI: 10.16027/j.cnki.cn31-2043/h.2013.02.002.
- [4] 冯胜利. 论语体的机制及其语法属性[J]. 中国语文, 2010 (05): 400-412+479.
- [5] 陶红印. 试论语体分类的语法学意义[J]. 当代语言学, 1999 (03): 15-24+61.
- [6] 张伯江. 语体差异和语法规律[J]. 修辞学习, 2007 (02): 1-9.
- [7] 方梅. 语体动因对句法的塑造[J]. 修辞学习, 2007 (06): 1-7. DOI: 10.16027/j.cnki.cn31-2043/h.2007.06.001.
- [8] 冯胜利. 语体语法及其文学功能[J]. 当代修辞学, 2011 (04): 1-13. DOI: 10.16027/j.cnki.cn31-2043/h.2011.04.001.
- [9] 冯胜利. 语体语法: “形式-功能对应律”的语言探索[J]. 当代修辞学, 2012 (06): 3-12. DOI: 10.16027/j.cnki.cn31-2043/h.2012.06.005.
- [10] 施春宏. 语体何以作为语法[J]. 当代修辞学, 2019 (06): 1-20. DOI: 10.16027/j.cnki.cn31-2043/h.2019.06.001.
- [11] Tian Shao, Chengwen Wang, Gaoqi Rao and Endong Xun. The Semantic Change and Distribution of Adjoining Adverbs in Modern Chinese. From Minimal Contrast to Meaning Construct. 2020: 149-164.
- [12] 《现代汉语词典》(第七版)[M]. 北京: 中国社会科学院语言研究所词典编辑室, 2016.
- [13] 施光亨. 《汉语口语词典》[M]. 商务印书馆, 2012.
- [14] 高艳. 现代汉语口语词的主要类型及基本特征[J]. 海外华文教育, 2017 (09): 1188-1199.
- [15] 荀恩东, 饶高琦, 肖晓悦, 臧娇娇. 大数据背景下 BCC 语料库的研制[J]. 语料库语言学, 2016, 3 (01): 93-109+118.