汉语语义倾向语料库的建设 *

杨　江[1]，李　薇[2]，彭石玉[2]

（1. 湖南科技大学，湖南 湘潭 411201；2. 武汉工程大学，湖北 武汉 430205）

摘　要：该文从研究背景、设计思路、标注体系和方法、加工步骤等方面介绍了汉语语义倾向语料库的建设过程。该语料库是一个以研究语言主观性表达为目的的共时、非平衡、单语标注语料库，依据语言主观性多维度描述体系而设计，规模为 100 万字，配备有集检索与统计、结果检查与可视化于一体的专用语料库工具箱系统，具有可用性大、标注质量高、语言学理据强等特点。

关键词：语义倾向；语料库；主观性；建设

The Construction of the Chinese Semantic Orientation Corpus

Jiang Yang[1], Wei Li[2], Shiyu Peng[2],

(1. Hunan University of Science and Technology, Xiangtan, Hunan 411201, China

2. Wuhan Institute of Technology, Wuhan, Hubei 430205, China)

Abstract: This paper introduces the construction procedures of the Chinese Semantic Orientation Corpus (CSOC) by presenting its research background, design plan, annotating system and processing steps. The CSOC is an unbalanced synchronic monolingual corpus with the purpose of researching linguistic subjective expressions. Shipped with a concordancing, retrieving and visualizing integrated toolkit, the one million Chinese character corpus is specially designed according to a multi-dimensional descriptive system of linguistic subjectivity. It is characterized by its high-quality, linguistic motivation and double usability for both linguistics and Natural Language Processing.

Key Words: semantic orientation; corpus; subjectivity; construction

1　引言

主观性（subjectivity）是语言的基本属性，语言意义的主观性是指话语中伴随命题内容产生的说话人的“自我”（self, ego）表达。日常话语中或多或少总是含有说话人“自我”的表现成分，说话人在说出一段话的同时也表明了自己对这段话的立场、态度和感情[1]。语言的主观性借助一定的语言手段、通过一定的语言形式得以实现，由此形成话语中的主观性表达（subjective expression），用以传递说话人的自我判断、感受、评价、意愿等主观性信息。对语言主观性以及主观性表达的关注，其实质是探索语言中“人”的因素，因为“语言不仅仅是客观地表达命题和思想，还要表达言语的主体即说话人的观点、感情和态度”[2]。

语言中的主观性表达是近年来语言学和自然语言处理领域的一个研究热点。语言学的相关研究着重从语言的角度探讨主观性表达的意义、使用、认知机制和描写手段，由此引发了对语言主观性的大量论述，使其逐渐成为认知语言学、功能语言学和语用学的元理论基础，并推动了“评价系统”的产生；自然语言处理的相关研究则主要从信息的角度关注主观性表达的辨识、抽取、分类和计算分析，从而产生了情感分析、观点挖掘、舆情监测等一批新兴研究方向。

研究语言中的主观性表达，不论是基于语言还是基于信息的视角，也不论是面向基础研究还是应用研究，都需要积累大量的语言素材，以帮助人们观察和把握语言事实，分析和研

* **基金项目：**教育部人文社会科学研究项目（11YJC740127）；湖南省教育厅科学研究优秀青年项目（14B068）

作者简介：杨江（1978—），男，副教授，主要研究领域为计算语言学；李薇（1978—），女，副教授，主要研究领域为对外汉语教学，应用语言学；彭石玉（1967—），男，教授，主要研究领域为功能语言学。

究语言的规律。具体而言，主要体现在其或为论证提供例句支持，或为描写提供统计数据，或为统计模型提供训练数据。这就要求建立基于既定标注体系、符合潜在研究需求、具有一定规模和加工深度的主观性表达语料库。

然而，就我们所知，目前国内外可获得的相关汉语语料库资源较少。古伦维等[3]的评价语料库对语料的篇章、句子、词语的情感倾向进行了标注，区分了显式和隐式观点持有者，但未能涉及词法分析信息；徐琳宏等[4]创建的100万字的情感语料库基于情感词汇本体[5]进行情感类别、主体、接受者、修辞类别等的标注，语料规模大，设计精细，标注信息详尽，但以句子为单位的加工层次略嫌粗糙；宋鸿彦等[6]完成了600余句的汉语意见型主观性文本标注语料库的标注，包含了词法和句法分析信息，但语料均为汽车评论，来源相对单一且规模较小；彭宣维等[7]遵循“评价系统”建立了100万词的汉英对应评价意义语料库，是首次按照一种语言理论体系构造的双语对应语料库，标注信息详尽，但其设计目的主要针对语言评价意义的研究；崔晓玲[8]构建了汉语网络新闻评论情感语料库，同样基于系统功能语言学的评价理论来设计，但其规模仅为13万字，语料来源均为单一的新闻评论，也不包含词法分析信息。除了上述的语料库以外，尚有一些零散或未经人工标注但值得一提的资源，如中文信息学会信息检索专业委员会提供的历届中文倾向性分析评测（COAE）语料，中国计算机学会主办的历届自然语言处理与中文计算会议（NLP & CC）提供的中文微博情感分析评测语料，谭松波[9]的中文情感挖掘语料等，但它们均用途单一，且难以形成规模。

由此可见，此前为研究汉语主观性表达而建设的语料库资源，由于标注体系不同，加工深度各异，应用目的多样，难以将其整合或统一；由于设计思路的差异，对领域研究认识的不同，其中的部分资源不能为当前研究背景和当下研究需求下的情感分析、语义倾向计算、观点挖掘等提供有力支持。在这样的背景下，我们从2011年开始，历经三年，完成了100万字的汉语语义倾向语料库（Chinese Semantic Orientation Corpus，CSOC）的标注工作，同时开发了集语料检索与统计、标注结果检查与可视化于一体的专用语料库工具箱系统（CSOC Toolkit）。汉语语义倾向语料库具有以下特点：

（1）从语言和计算两个角度综合考虑了语料的可用性，因而既能在语言学上为汉语主观性表达的基础研究所用，又能在自然语言处理上为主观性表达的计算和分析等应用研究所用；

（2）自觉地接受语言学理论的指导，每个加工环节、每项标注元素都既有语言学上的理据，又实实在在地面向相关研究和应用需要；

（3）标注体系遵从预先设计的“语言主观性多维度描述体系”；

（4）规模适中，同时尽量保证语料在领域、体裁、语体等方面的平衡性；

（5）标注过程有严格的质量保障机制，标注结果质量高。

2　设计思路和概念界定

汉语语义倾向语料库的设计思路遵循我们自行构建的“语言主观性多维度描述体系”。语言主观性多维度描述体系是一个以语言主观性理论为指导、面向文本主观性分析应用、衔接理论和应用的中间“接口”，它上连各种语言学理论、下接各类主观性分析，旨在为不同语言层级、不同颗粒度和不同应用目的的主观性分析提供统一的、跨语言的描述标准。该体系用类别、程度、形式、成分、关联和模式六个维度表示，每个维度反映语言主观性的一种属性，也代表一类研究视角，涵盖了当前学界正着力解决和未来可能进行的各项子任务。该体系的创建借鉴了Martin[10-11]的“评价系统”、Taboada等[12]和Read等[13]将“评价系统”应用于语义倾向计算所做的尝试性探索、Wiebe等[14]为建设MPQA观点标注语料库设计的个人心理状态（private state）标注框架、Kim等[15]面向观点挖掘为观点（opinion）制定的由主题（topic）、持有者（holder）、陈述（claim）、情感（sentiment）组成的四元组以及徐琳宏等[5]的情感词

汇本体，其框架结构如图1所示。篇幅所限，本文不对此展开详细论述。

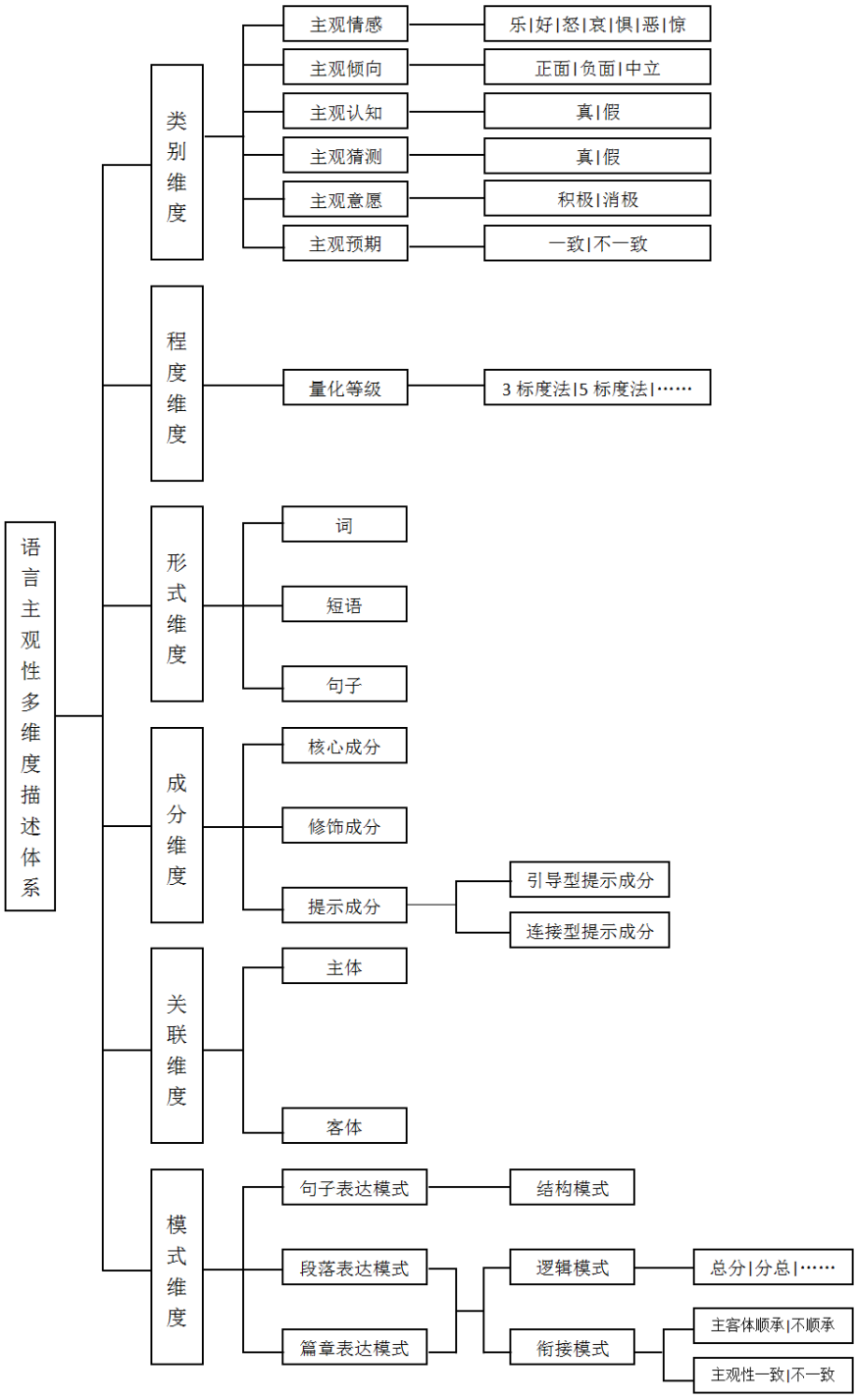

图1　语言主观性多维度描述体系框架结构图

语义倾向（Semantic Orientation）是语言主观性的一个子类，同其他子类一样，对它的刻画可以继承自语言主观性描述体系，只需在类别维度稍作修改，即可产生一个语义倾向描述子体系。汉语语义倾向语料库就是基本依据这个子体系设计的。需要指出的是，考虑到对语义倾向程度的描述大多以词典形式提供，加之句、段、篇的表达模式一般可以从其他维度的标注中间接推导得到，因而我们在标注体系中剔除了程度和模式两个维度。

下面对语料标注中涉及的一些基本概念进行界定和说明。

（1）语义倾向。语义倾向指倾向主体（subject）对倾向客体（object）所持有的赞成或反对、褒扬或贬抑、肯定或否定、积极或消极的态度、立场、观点或情感，分正面、负面和中立倾向三类。

（2）倾向主体。倾向主体是语义倾向的持有者、评价者或体验者，一般为有生命的人或由人组成的群体，在特殊语境下，如神话传奇、童话故事、科幻小说中，也可以是人格化的动物和物件。

（3）倾向客体。倾向客体是语义倾向的评价对象、接受者或针对方，通常为人、物、事件、动作行为等。

（4）正面倾向。指表达赞成、褒扬、肯定或积极类主观性的语义倾向。

（5）负面倾向。指表达反对、贬抑、否定或消极类主观性的语义倾向。

（6）中立倾向。指表达不偏不倚类主观性的语义倾向。

（7）核心成分。核心成分是表达语义倾向的中心和关键要素，形式上多为负载语义倾向的词和短语，少数情况下为句子（含小句），如“怀疑、善良、大公无私、让一切随风而去”。

（8）修饰成分。修饰成分指用以修饰核心成分，使其倾向程度增强或减弱的成分，以程度副词和否定副词居多，如“有点、非常、不”。

（9）提示成分。提示成分是本身不对核心成分产生影响，但具有引出或连接核心成分作用的成分。提示成分又分为引导型和连接型两类，其中，引导型提示成分用以引出核心成分，多数为表示心理状态的动词，如“想、认为、觉得、以为、希望”等；引导型提示成分用以连接两个或两个以上核心表达成分，即通常所说的关联词语，如“和、既…又…、虽然…但是…”等。

上述基本概念也即标注的主要元素，它们之间的关系可以用图 2 直观地表示。

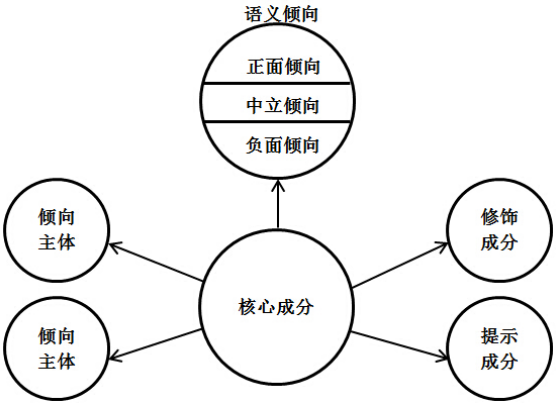

图 2　主要标注元素关系图

3　标注体系和标注方法

汉语语义倾向标注语料库的标注体系由文档结构标注体系和语义倾向标注体系构成，前者标注文档（即篇章）的层次结构，分为篇、段、句、词四级，后者则标注语义倾向的类别、形式、成分、关联四个维度的信息。语料标注遵守 Leech[16]提出的七条基本原则，采用国际通行的 TEI 标注模式，标注结果用 xml 格式文件储存。

文档结构标注体系表示成 text = (head, body)，其中，头信息表示成 head = (title, time, author, source, addr, info)，正文表示成 body = (para, sent, word)。此外，每级语言层次都附加了必要但并不完全相同的其他信息。例如，词、句、段三级都含有序号（id），而仅词语层级包含词性信息（pos）。文档结构标记集及其说明见表 1。

表 1　文档结构标记集及其说明

序号	标记	含义
1	<text> </text>	文档
2	<head> </head>	头信息
3	<time> </time>	发表时间
4	<author> </author>	作者
5	<sour> </sour>	出处
6	<addr> </addr>	网址
7	<info> </info>	其他需要说明信息，或非正文内容
8	<body> </body>	正文
9	<para> </para>	段
10	<sent> </sent>	句
11	<word />	词
12	id	段、句、词的序号，从 0 开始
13	cont	句、词的文字内容
14	pos	词的词性

不同的语言层级在语义倾向标注体系上略有差别。在篇、段级，我们标注其语义倾向类别和倾向客体，表示为 textSO/paraSO = (senti, obj)；在句一级，标注其语义倾向类别、句子核心话题、是否否定句、是否疑问句、是否修辞句，表示为 sentSO = (senti, topic, neg, que, fig)；而在词一级，我们围绕核心成分，标注它的语义倾向类别、成分、关联元素，表示为 coreSO = (senti, sub, obj, modi, clue)。语义倾向标记集及其说明见表 2。

表 2　语义倾向标记集及其说明

序号	标记	含义
1	senti	语义倾向
2	po	语义倾向类别，正面倾向
3	ne	语义倾向类别，负面倾向
4	zl	语义倾向类别，中立倾向
5	sub	倾向主体
6	obj	倾向客体
7	topic	句子核心话题
8	modi	修饰成分
9	clue	提示成分
10	neg	是否否定句
11	que	是否疑问句
12	fig	是否修辞句
13	span	核心成分跨距
14	yes	布尔值，真
15	+	分隔多个 id 所代表的词语
16	–	连接两个 id 所代表的词语
17	数字	词语 id

图 3 是一个句子的标注示例。

```
<para id="12" senti="ne" obj="这" >
  <sent id="0" neg="yes" que="" fig="" senti="ne" topic="2" cont="我以为这是最没有出息的人。">
    <word id="0" cont="我" pos="r" />
    <word id="1" cont="以为" pos="v" />
    <word id="2" cont="这" pos="r" />
    <word id="3" cont="是" pos="v" />
    <word id="4" cont="最" pos="d" />
    <word id="5" cont="没有" pos="v" />
    <word id="6" cont="出息" pos="n" senti="po" sub="0" obj="8+2" modi="5+4" clue="1" />
    <word id="7" cont="的" pos="u" />
    <word id="8" cont="人" pos="n" />
    <word id="9" cont="。" pos="w" />
  </sent>
  ……
</para>
```

图 3　一个句子标注示例

文档结构标注主要由机器自动完成，后期进行了必要的人工核查，主要针对分词和词性标注的错误；语义倾向标注主要由人工手动完成，后期辅以标注结果检查程序进行自动纠错，主要针对各级 id 错误、标记拼写错误、xml 合法性等问题。

如图 3 所示，在语义倾向标注上，对于 sub、obj、modi、clue 等属性的值，我们使用了数字，这些数字代表当前句子中词语的 id。由于每一个词都有唯一的 id，因此，为了节省存储空间，我们用其 id 代表其文字内容，这样做也能减轻标注人员的劳动强度。篇、段、句的标注内容基本相同，从图中可直观看出，不赘述。对于词一级的语义倾向各维度的属性，我们将其标注在核心成分上，这主要是考虑到核心成分在表达语义倾向时具有的关键作用；另外一重考虑则是针对含有多个核心成分的句子，这些句子中的 sub、obj、modi、clue 等属性会出现交错和重叠，而将其放置在核心成分上，相互之间的关系就会很清楚，层次感强，标注人员也方便理解和操作。

对于以下两种情形，我们引进 span 标记进行特殊处理：（1）句中的核心成分不是词，而是短语，如“没/得/说、吃/空饷”等；（2）核心成分被分词软件切分成了多个词，但从分词的角度看又并非错误，如“死守/不/放、功/在/当代”等。上述情形下，我们采用 span 标记将多个词组成的核心成分连接起来，将其视为一个整体，形如“span="id 起始-id 终止""，span 标记放置在终止 id 所代表的词语上。

4　研制过程

汉语语义倾向语料库是一个百万字符级规模的共时、非平衡、单语标注语料库。主要的建设过程包括语料收集、预处理、标注和校对。

4.1　语料收集

语料选取的首要原则是来源语料中含有较丰富的语义倾向，在满足这一前提后，尽量保证语料在语体、文体、领域等属性上的平衡。根据这个思路，我们收集了来自文艺期刊、童话故事、小说戏剧、语文课本、网络评论的文本 960 篇，各类来源的字数控制在约 15-30 万之间。表 3 列出了语料的组成信息。

表 3　汉语语义倾向语料库的组成信息

语料来源	内容简介	字符数	词数	句数	段数	篇数
文艺期刊	《读者》、《青年文摘》选篇	305520	211134	10184	3659	192
童话故事	《安徒生童话》、《格林童话》、《王尔德童话》、《一千零一夜》、《郑渊洁童话》、《中国童话故事百篇》选篇	200459	139756	8265	3218	60
小说戏剧	《雷雨》、《北京人在纽约》、《新结婚时代》、《五星大饭店》全本	192717	151136	13050	6618	25
语文课本	人教版小学、初中、高中语文课本、对外汉语教材《桥梁》、《中级汉语》、新加坡小学语文课本选篇	200555	144412	7789	2061	162
网络评论	书籍、电影、电视剧、酒店、手机、电脑等网络评论	143884	93892	4230	1728	521
总计		1043135	740330	43518	17284	960

4.2　语料预处理

生语料文本经过清洗、核对和文档规格化处理后，进入文档结构标注和词法分析序列。文档结构标注环节主要完成篇章内段落和句子的切分，词法分析环节则完成词语切分和词性标注任务。词法分析采用中国传媒大学文本切分标注系统（CUCBst 1.0），这是一个基于规则的词法分析系统，整体正确率超过 97.45%。生语料文本经过上述步骤后被转换成类似图 3 所示的 xml 格式待标文件，其中尚存的各种错误在语义倾向标注时一并纠正。

4.3　语料标注

语义倾向标注在文本编辑软件 UltraEdit 上进行，标注过程包括培训、试标、讨论、正式标注等环节。首先由研究人员对标注人员进行标注培训，讲解相关背景、应用前景、设计思路、标注体系、有关概念、注意事项等知识，然后 10 名标注人员按语料来源分成五组，各组成员分别独立进行前期五万字的试标注。试标注完成后，小组内部各自展开讨论，对小组成员的标注结果进行比对，利用文本比较软件 Beyond Compare 找出异同，重点讨论差异之处，并记录无法达成一致的标注差异。小组讨论后再召开全体人员大会，逐条讨论各组提出的标注差异，共同探讨以解决分歧，逐步完善标注体系和标准。接着开始正式标注，研究人员分批次将任务发放给各组，各组内人员同时标注相同语料。每批次标注完成后，各组仍先行在组内讨论，再进行全体讨论。如此反复，直至全部任务结束。标注过程中严格遵循“分批次发放任务—组员独立标注—小组讨论—大会讨论—返修—提交结果”的循环工作模式，基本保证了人工标注的一致性。

标注一致性(Inter-Annotator Agreement)是衡量语义标注语料库质量的一个重要指标，常用 Kappa 统计量衡量。我们统计了各组内部标注人员在各阶段对部分主要标注元素的完全相同实例数量（严格相等），用公式（1）在 SPSS 中计算了对应的 Kappa 系数值，以掌握标注语料的状况。详细数据见表 4。

$$K = \frac{Pa - Pe}{1 - Pe} \tag{1}$$

其中，Pa 表示两名标注者评定一致的百分比，Pe 表示理论上评定一致的百分比。

表 4　各组标注一致性统计

*K*值（%）	第 1 批（试标注）			第 2 批			第 3 批		
	核心成分	倾向主体	倾向客体	核心成分	倾向主体	倾向客体	核心成分	倾向主体	倾向客体
组 1	79.8	60.6	72.2	91.9	74.7	83.5	91.7	75.9	85.8
组 2	80.3	66.3	73.8	93.2	77.1	83.4	92.8	79.6	83.3
组 3	78.1	64.4	70.4	90.6	75.9	79.7	91.1	78.4	82.5
组 4	79.5	70.8	74.3	91.4	81.3	83.6	91.3	81.2	83.9
组 5	82.4	68.2	74.9	92.7	78.1	84.2	93.2	80.5	84.6

4.4　质量保障

人工标注的语料质量主要体现在标注的正确性上，这又可以从两个方面来衡量：一是对标注规范的理解是否准确，二是标注结果是否一致，尤其是由多人完成的大型标注工作。虽然我们在标注过程中采取了一定的措施，以尽量保证标注人员理解准确，标注一致，但仍然无法避免问题和错误的存在，因此，仍有必要对标注语料进行人工校对。校对的步骤与标注过程大致相似。保障校对质量的手段包括：（1）研究人员编制了详细的校对操作手册，集中阐释了标注过程中遇到的典型难点、疑点问题（如倾向主体和倾向客体的标注），并提供给校对人员参考；（2）研究人员与校对人员集体办公，以便随时讨论。

由于标注和校对都是人工进行的，在标记的输入、更改上难免出现输入错误，加之标注文件和校对文件都是具有结构层次关系的 xml 格式文件，极易破坏原有格式，而这些错误人工往往难以识别。因此，我们专门编制了一系列辅助检查和自动纠错工具软件，保证了标注和校对结果文件的完整、合法和正确。

通过上述步骤，我们完成了汉语语义倾向语料库的建设。表 5 列出了标注语料的部分统计信息。

表 5　汉语语义倾向语料库的部分标注结果统计信息

语料类别	核心成分	倾向主体	倾向客体	修饰成分	提示成分	倾向句	正面倾向句	负面倾向句	非倾向句	否定倾向句
文艺期刊	16764	5528	8668	2574	2317	5294	2985	2126	4890	100
童话故事	7630	2630	3611	928	484	412	167	238	7853	53
小说戏剧	8957	5195	4440	1666	311	4533	1361	3064	8517	69
语文课本	7018	2853	3424	1172	818	1267	746	513	6522	62
网络评论	9937	2337	6887	2382	638	3484	1589	1820	746	400
总计	50306	18543	27030	8722	4568	14990	6848	7761	28528	684

5　汉语语义倾向语料库专用工具箱系统

为了更好地利用汉语语义倾向语料库，我们开发了 CSOC Toolkit 专用工具箱系统。它由四大模块组成：检查抽取工具集、检索模块、统计模块和可视化模块。

（1）检查抽取工具集。工具集的开发初衷本是为了在标注时辅助人工完成检查和纠错任务，随着需求的不断增加，新添功能逐渐增多，于是将其整合到一起，作为工具箱的一个

独立模块。除了能够检查标注错误和对一部分错误进行自动纠错外，工具集还提供了标注语料信息概览、原始语料抽取等功能。

（2）检索模块。这个模块提供两类的检索功能：一类是固定的与语义倾向相关的内容检索，如倾向词、倾向句、倾向主体、倾向客体等的检索，另一类是任意字符串或标记的检索。检索完成后可以纯文本或富文本格式保存结果。图 4 是倾向词语检索的某个结果截图。

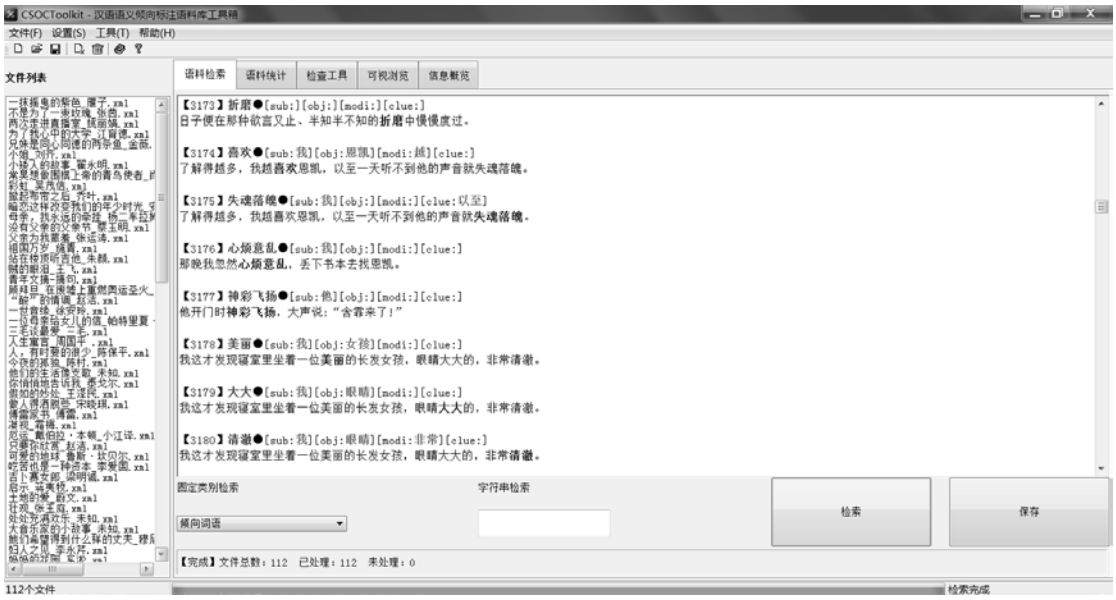

图 4　固定类别“倾向词语”项的检索结果

（3）统计模块。该模块提供对固定项的统计，如统计语义倾向成分、倾向句、非倾向句、正面倾向句、负面倾向句、否定倾向句等，统计结果以表格的形式呈现，并提供排序功能。统计结果可存为纯文本或 Excel 表格格式。

（4）可视化模块。为了方便人对语义倾向成分标注结果的直观观察，我们特别开发了可视化模块，在其中可以逐句浏览原始文本、分词文本、词性标注文本和语义倾向标注文本。语义倾向标注结果在呈现时，用不同颜色突出显示相关文本内容，并在文本顶部用带颜色和箭头的弧线表示他们之间的语义倾向关系，词性标记则在文本的底部显示。图 5 是《恶毒的王子》标注结果的可视化显示效果。

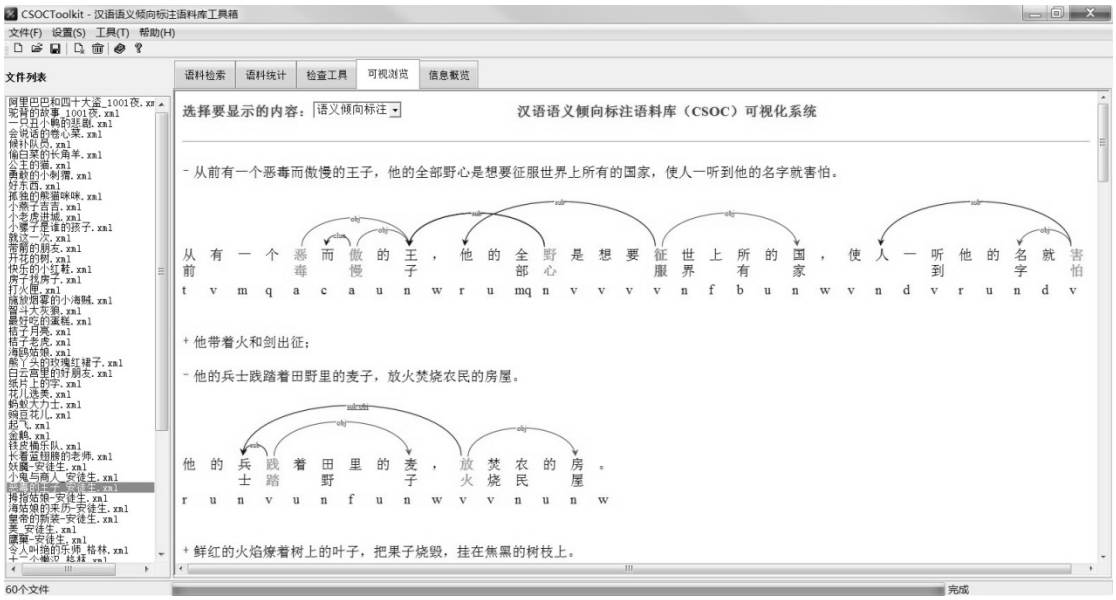

图 5　《恶毒的王子》标注结果可视化显示效果

6　结语

基于语言主观性多维度描述体系，我们构建了一个中等规模的汉语语义倾向语料库，并为之配备了相应的检索、统计和可视化工具，这项工作所产出的资源既适用于汉语主观性表达的基础研究，又适用于与主观性相关的应用研究。

语言中的主观现象日益受到学界和业界的重视，近 10 年间的相关工作成绩喜人，但总的来说，人们对于语言表达主观性的形式、方式、机制、规律、特点、差异等方方面面的问题所知尚浅，认识仍待深入。比如，语言中主观性表达的分布状况如何，各级语言单位在表达主观性上分别具有怎样的特点和规律，不同语言或同一语言的不同文体在表达主观性时有何差异，等等。对这些问题的回答和解决都有赖于对大量真实文本的有效统计和分析，本文的工作有望为这些研究提供一定的帮助，从而共同推动领域研究的发展。

参考文献

[1]沈家煊.语言的“主观性”和“主观化”[J].外语教学与研究, 2001,33(4):268-275.

[2]沈家煊.汉语的主观性和汉语语法教学[J].汉语学习, 2009,(4):3-12.

[3]Lun-Wei Ku,Tung-Ho Wu,L-i Ying Lee and Hsin-His Chen.Construction of an Evaluation Corpus for Opinion Extraction[C]//Proceedings of NTCIR-5 Workshop Meeting, Tokyo, Japan, 2005.

[4]徐琳宏,林鸿飞,赵晶.情感语料库的构建和分析[J].中文信息学报,2008,22(1):116-122.

[5]徐琳宏,林鸿飞,潘宇等.情感词汇本体的构造[J].情报学报,2008,27(2):180-185.

[6]宋鸿彦,刘军,姚天昉等.汉语意见型主观性文本标注语料库的构建[J].中文信息处理 2009,23(2):123-128.

[7]彭宣维,杨晓军,何中清.汉英对应评价意义语料库[J].外语电化教学,2012,247(9):3-10.

[8]崔晓玲.基于汉语网络新闻评论的情感语料库标注研究[J].北京邮电大学学报（社会科学版）,2013,15(6):21-29.

[9]谭松波.中文情感挖掘语料[DB/OL].(2010-06-29)［2013-07-20］. http://www.searchforum.org.cn/tansongbo/corpus-senti.htm

[10]Martin, J.R. Beyond Exchange: APPRAISAL Systems in English[C]. *Evaluation in Text*, Hunston, S. & Thompson, G. (eds), Oxford: Oxford University Press, 2000:142-175.

[11]Martin, J.R., and White, P.R.R. *The Language of Evaluation: Appraisal in English*, New York: Palgrave Macmillan, 2005.
[12]Taboada, M., and Grieve, J. Analyzing Appraisal Automatically[C]//Proceedings of American Association for Artificial Intelligence Spring Symposium on Exploring Attitude and Affect in Text, Stanford, USA, 2004:158-161.
[13]Read, J., Hope, D., and Carroll, J. Annotating expressions of appraisal in English[C]//Proceedings of Linguistic Annotation Workshop, ACL 2007, Prague, Czech, 2007:93-100.
[14]Wiebe, J., Wilson, T., and Cardie, C. Annotating expressions of opinions and emotions in language[J]. Language Resources and Evaluation, 2005, 39(2-3):165-210.
[15]Kim, S.-M. and Hovy, E.Determining the Sentiment of Opinions[C]// Proceedings of the COLING Conference 2004, Geneva, 2004:1367-1373.
[16]Leech,G.Corpus annotation schemes,*Literary and Linguistic Computing*, 1993, 8(4):275-81.