

面向俄文情感分析的新闻评论语料库建设与应用*

朱珊珊 原伟

(信息工程大学, 洛阳 471003)

摘要:从文本中挖掘情感态度是当前计算语言学研究的重点领域和热点方向。论文在分析梳理情感分类方法和俄语情感表达主要手段的基础上,将人的主要情感划分为4大类、19个子类,并基于网络自动构建约730万词的新闻及评论语料库用于实验。为考察俄语情感表达的基本词汇手段,设计一种基于TEI标准的语料情感标注系统,对语料库中热点新闻及其8031条用户评论进行分类标注,所标注语料成功运用到俄语新闻评论的情感自动分析和领域情感词典的构建中,实验结果证明所用方法的有效性和实用性,获得包含6321项情感表达手段的机用词典,后续应用潜力较大。

关键词:情感分析;俄语;语料库;新闻评论

中图分类号:H030 文献标识码:A 文章编号:1000-0100(2020)01-0024-6

DOI编码:10.16263/j.cnki.23-1071/h.2020.01.003

Building and Using News Comments Corpus for Russian Sentiment Analysis

Zhu Shan-shan Yuan Wei

(Information Engineering University, Luoyang 471003, China)

Mining emotional attitudes from text is the focus and hotspot of current computational linguistics research. After analyzing sentiment classification methods and main means of Russian sentiment expression, this paper divides main human emotions into 4 major categories and 19 sub-categories, and automatically builds a corpus of news and reviews based on the Internet for experiment, which contains over 7.3 million words. In order to examine basic vocabulary means of Russian emotional expression, a sentiment labeling system based on the TEI standard was designed. The news in the corpus and its 8031 user comments were classified and labeled. The labeled corpus successfully applied in sentiment analysis of Russian news comments. The experimental results proved the effectiveness and practicability of our methods. A sentiment dictionary containing 6,321 emotion expressions was obtained, which has great potential for subsequent applications.

Key words: sentiment analysis; Russian; corpus; news commentary

1 引言

情感分析(Анализ тональности)是计算语言学的重要研究方向,即借助计算机对文本语料中和包含的情感与态度进行分类、分析与挖掘,在语言学领域可运用于语义学、话语分析翻译和外语教学中,在信息科学领域可使用在倾向性分析、舆情监控、话题识别和用户评价分析等场景中,研究前景十分广阔,应用潜力巨大。

从传统语言学的视角看,在俄语情感意义与

表达手段研究方面,学界已取得一些成果,可分为以下两类:第一类针对动词、俄语成语、叹词、情感—表情词、标题等语言结构单位进行情感意义考察,梳理其结构特点和语义内涵,揭示其评价意义和特有民族文化色彩(于金玲 2010:96;李慧 2013:3;蔡晖 2014:15;杨婷 2014:14;曹广跃 2016:11);第二类从语言意识、语义范畴化等理论视角切入,对俄语情感意义、伴随和意义、词汇联想等问题进行讨论(林楠 2011:5,杨利芳

* 本文系国家哲学社会科学基金项目“基于本体的俄汉可比语料库构建与评估”(14CYY051)和“基于可比语料库和本体的俄汉网络新闻话题监测与情感识别研究”(18BYY235)的阶段性成果。

作者电子邮箱:peimingshanshan@163.com(朱珊珊)

2012:30,杨瑞 2015:115)。上述研究对为分析俄语情感表达手段,揭示其结构特点、语义内涵和文化意义做出重要贡献,然而从语言处理工程的角度来讲,一些理论方法、阐释与分类标准在计算机形式化描写层面较难完整表述或具体实现。

从计算语言学的视角来看,近年来国内面向英语、维吾尔语、藏语等语言构建一系列用于文本情感分析的语料库及应用系统(汪正中 张洪渊 2011:153,吐尔贡等 2017:67,袁斌等 2016:682),但未见面向俄语的情感分析研究。在国外,面向俄语的情感分析系统数量不多,较有代表性的有 Senti Strength^①、Аналитический курьер^②、Ваал^③、RCO Fact Extractor^④等。Senti Strength 系统主要面向简短、非结构化英文文本的情感分析任务,也包含部分处理俄语文本的功能,工作原理是将文本中的词汇分为两种评价色彩,即消极与积极,并赋予9个评价(从-4到+4),统计综合积极和消极词汇评价进行加权评估,测算最大情感意义值。Аналитический курьер 系统采用基于规则和词典的方法,对句子标注具有情感色彩的词汇,将其与邻接无情感意义词组成词链(цепочка),最后综合考虑句子的情感性分析文本。Ваал 系统基于频率词典,并将部分词汇划分至心理语言学范畴,定位于评估人们无意识的情感互动。RCO Fact Extractor 系统基于规则的方法搭建,考虑文本的句法结构以及不同类型词之间的相互作用。可以看出,上述系统在俄语情感分析领域做出十分积极的工作,但也存在如下不足:首先,上述系统都是封闭和付费的,数据集和应用接口无法调用共享;其次,未充分考虑俄语的自身特点,如形态多样性、语序自由性等;第三,情感评价标准过于简单,多采用二元化(褒贬或正面负面)准则,主要关注文本中评价色彩的总体极性和倾向性,对评价色彩的强度描述不足。

1 情感分类和俄语情感表达手段

参考中文情感分析相关成果的分类方法(徐琳宏等 2008:118),本文将情感描述为4个基本类及其19个子类:喜(愉、舒、敬、扬、信、爱、愿)、恶(烦、憎、贬、妒、疑、怒)、哀(悲、憾)、惧(恐、羞、惊、慌)。其他复杂情感可以被描述为这些基本情感的细化、加强、迁移和混合。当然目前也存在更为复杂精细的情感划分方法,本文认为过度细分情感大类别难免会过多地带入研究者的主观性,在语料标注时也会造成类别部分重合难以严谨描述,对研究的客观性带来的混乱。

对于俄语情感表达手段的问题,已有学者(王向丽 2004:5)基于语义与形态特征结合的方法将其分为构词手段、修辞手段、言语手段和词汇手段,这也较为符合学界的普遍认知。一般来说,构词手段指为动词、名词、形容词等词汇添加构词词缀,表达喜爱、厌恶、讽刺等复杂情感意义(如 мама-мамочка, солдат-солдатчина, малыш-малышня);修辞手段可包括不同语体修辞色彩的同义词使用(如 глаза-очи, красота-краса, бросать-кидать);言语手段可包括在言语交流使用表达情感评价意义的称谓、句式等。除以上手段外,还存在诸多非常规使用标点(如问号、感叹号结合使用、标点重复使用、添加引号等)、大写(如“ТРАМП, Это АМЕРИКАНСКИЙ президент”)、表情符号、拟声等手段表达情感的例子。本文重点考察俄语情感表达的词汇手段,并用于语料库的数据采集、标注与构建中。而对构词、修辞和言语手段,目前学界理论体系尚不完善,难以完整地形式化描述,在未来的工作中会深入讨论。

2 语料库构建与情感标注

2.1 语料采集

网络新闻本身承载着发布方的立场与态度,而新闻评论则富含用户情感信息,具有主观性强、互动积极、反馈及时和立场鲜明等特点。因此,本文选择俄罗斯权威网络媒体发布的网络新闻及新闻评论为语料的主要来源构建语料库,为深度分析这些媒体及用户的态度与情感奠定基础。

表1 新闻及评论语料采集结果

语料类型	数量篇(条)	总词数(个)	容量(KB)	平均(词)	最长(词)	最短(词)
正文	1647	465345	5948	247	2345	34
评论	235884	6840637	84704	29	1347	1
总计	237531	7305982	90652	—	—	—
比值	1:143.2	1:14.7	1:14.2	—	—	—

具体来说,本文选择俄罗斯网站与美国总统特朗普(Трамп)相关的新闻与评论为实验语料,采用面向特定新闻网站、使用关键词搜索、获取下载网页并抽取正文与相应评论的方法收集数据(原伟 2017:37),采集来自 РИА^⑤新闻网站的语料。根据2017年10月26日 РИА 网站的搜索结果,共查询到17742个与 Трамп 相关的文本(视频、图片新闻除外)。按照网站用户评论数量排序(Обсуждаемое)后,发现在1647篇新闻后评论数降低到50以下,因此以此为节点挑采集1647

篇新闻附带评论正文作为实验语料,保存为纯文本格式,总词数约 730.59 万,语料规模较为可靠,具体统计数据见表₁。

2.2 语料标注

首先,对语料进行预处理。作为屈折语的典型代表,俄语的词形变化复杂多样,语法体系严谨,语序灵活多变。面向俄语进行标注处理,首要任务是形态分析和词形还原。目前,较为知名的俄语文本自动处理工具(或系统)有 AOT^⑥,Стемка^⑦和 MyStem^⑧等,根据其功能不同各具特色。本文俄语处理采用 Yandex 公司研制开发的 MyStem 系统,该系统可有效识别并处理俄语词性和时间表达,还原名词、动词、形容词、代词、数词等词类的原始形式,标注命名实体(人名、地名和机构名等)等核心信息。本文调用 MyStem 的 Python 应用接口对收集的新闻语料进行形态分析词形还

原和命名实体识别,用于下一步的情感分析。

其次,根据本研究的目标与任务,将语料承载的信息分为 3 个部分:篇头信息、篇体信息和情感信息。篇头信息指整篇语料的属性和特征信息,针对本研究而言,指新闻语料的标题、时间、体裁、来源、作者、相关评论等,评论语料的用户、评论时间、相关新闻等;篇体信息一般指语料中的语法、语义和语用信息等,本文只涉及文本语法信息中的词类信息,借助俄语形态分析工具自动处理;而情感信息特别针对本文的情感语料库构建任务,主要包括情感对象、情感类别、子类等。此外,由于所建立的语料库需要体现一篇新闻语料及其评论语料之间的关联,基于上述考虑,我们基于较为通用的 TEI 标注集设计一套面向俄语情感分析的标注体系(表₂)。

表₂ 语料库标注体系

新闻语料标注			评论语料标注		
类别	编码	说明	类别	编码	说明
篇头信息	N_ID	新闻 ID	篇头信息	C_ID	评论 ID
	Title	标题		N_ID	相关新闻
	Time	时间		User	用户名
	Type	体裁	Comtime	评论时间	
	Source	来源	篇体信息	Comtext	评论原文
	Author	作者		Post_comtext	处理后
篇体信息	C_ID	相关评论	情感信息	means_s\p\g\so\dg	情感手段_名_形_动_搭_其他
	text	新闻正文		X_ra\sp\uv\xa\ve\lyu\je	好_愉\舒\敬\扬\信\爱\愿
	Post_text	处理后		H_be\ne\pl\re\so\gn	恶_烦\憎\贬\妒\疑\怒
				P_pe\ja	哀_悲\憾
				S_ko\st_ud\vo	惧_恐\羞\惊\慌

最后,进行语料标注。篇头信息与篇体信息的标注我们采用自动标注与人工校对相结合的方法,标注过程较为高效。情感标注是整个标注过程中工作量最大的部分,而目前由于俄语情感分析普遍缺乏可用数据资源、参考标准和处理工具,因此全人工标注虽然成本极高,但却势在必行。同时,为最大程度降低不同标注者对文本的情感判断的差异性,我们采用人工双重标注方法,标注者首先统一进行标注集学习和标准一致性讨论,随后采用对同一语料双人分别标注的方法,若标注结果一致,则通过标注结果;否则,采用第三人校对加以确认、判断与更正。在此基础上,我们对单篇评论数超过 600 条的新闻以及包含对总计 8031 条评论进行标注,面向俄语情感分析的实验新闻语料库初步

建成。新闻及评论语料标注示例如下:

```
<News_header> //新闻篇头信息
<N_ID> Трамп задался вопросом, не были Обама слишком мягок к России _ 201702151550 </N_ID> //
<Title> Трамп задался вопросом, не были Обама слишком мягок к России </Title>
<Time> 201702151550 </Time>
<Type> новость </Type>
<Source> https://ria.ru/world/20170215/1488044272.html </Source>
<Author> Brendan Smialowski </Author>
<C_IDs> Сергей Свой_201702151653, Александр_201702151638, ... </C_IDs>
```

```

</News_header >
< News_body > //新闻篇体信息
< text >... Крым был взят Россией во вре-
мя пребывания в должности Обамы... </text >
< Post_text >... Крым { крым, Gramme.
Geo:158 } был { быть, Gramme. Verb:137 }
взят { взять, Gramme. Verb:137 } Россией { рос-
сия, Gramme. Geo:158 }... </Post_text > //
</News_body >
.....
< Comments_header > //评论篇头信息
< C_ID > Александр_201702151638 </C_ID > //
< N_ID > Трамп задался вопросом, не был
ли Обама слишком мягок к России _ 20170
2151550 </Title >
< Comtime > 201702151638 </Comtime >
< User > Александр </User >
</Comments_header >
< Comments_body > //评论篇体信息
< Comtext > У Трампа авитаминоз что ли?
Это как "слишком мягок", а он что сделает?
Нападет на Россию? Нет это не политик, а ка-
кой-то уже затравленный зверек, который ки-
дается из крайности в крайность. Его загоняют
методично в угол. Печально. </Comtext >
< Post_comtext > У { у, Gramme. Preposi-
tion:135 } Трампа { трамп, Gramme. Surname:
156 } авитаминоз { авитаминоз, Gramme. Sub-
stantive:136 }... </Post_comtext >
</Comments_body >
< Comments_emotion > //情感信息标注
< E_objects > Трамп, Россия < E_objects >
< E_means > "авитаминоз" { means_s, H_
be }, "затравленный зверек" { meas_so, H_pl },
"кидаться из крайности в крайность" { meas_
so, H_pl }, "печально" { means_dg, P_ja } </E_
means >
</Comments_emotion >

```

3 语料库应用

3.1 新闻评论的情感分析实验

结合上文,在对 8031 条新闻评论进行篇头、篇体以及情感标注之后,基于该实验语料库,我们对其中所标注的情感类(好、恶、哀、惧)、情感表达手段(名词、动词、形容词、固定搭配)进行数量统计(表₃)。根据统计结果,如果将情感类“好”

作为积极情感,而将“恶、哀、惧”归为消极情感,那么积极情感与消极情感表达手段数量的比率约为 1:2.5 (5109:12933),也就是消极情感手段占主体,可初步判定俄罗斯民众对美国总统特朗普的总体情感态度为消极。

表₃ 情感表达手段的统计分析表

情感类	表达手段实例/个				
	名词	动词	形容词	固定搭配	手段总计
好	1063	1115	1637	1294	5109
恶	1292	1434	2315	1531	6572
哀	343	475	1192	495	2505
惧	921	1123	786	1026	3856
类总计	3619	4147	5930	4346	18042

显然,上述通过简单统计方法的情感分析判定较为粗糙。为更加精确的对评论语料库的情感态度做出判断,需要借助文本倾向性分析手段实现。文本倾向性分析的方法很多,这不是本文讨论的重点,我们采用基于支持向量机(SVM)的特征分类方法来验证本文语料情感标注的有效性。基于 SVM 的情感倾向性分析关键是情感特征项的选择,本文的情感特征项即评论语料中所标注的带有情感色彩的名词、形容词、动词和固定搭配,按照“积极类”和“消极类”区分,作为向量空间的维,将在文本中的出现频率作为特征的权重。随后设计情感分类器,将积极情感表达手段的特征值定位为正值(+1),将消极情感的特征值定位为负值(-1),如评论文本中无情感特征项或积极情感与消极情感项数量相当则判定为中性情感(0)。通过这样的方法,我们对 8031 条新闻评论进行了计算机情感倾向性自动判定。为对其准确性做出判断,组织人工阅读判定上述新闻评论的情感基调,两者结果比较见表₄。

表₄ 评论情感倾向性的人工与自动判定结果比较

情感倾向	人工判定		自动判定		误差条数
	评论条数	占比	评论条数	占比	
积极	2332	29.04%	2989	37.21%	657
中性	832	10.36%	593	7.39%	239
消极	4867	60.60%	4449	55.40%	418
总计	8031	—	8031	—	1314

从结果来看,计算机自动判定的总体结果同之前的统计分析基本一致,即消极评论占主体,但从数据细节来看,计算机存在一定比例的误判。

这里分析机器判定错误的原因,我们认为主要有以下 3 点:第一,还未考虑其他情感表达手段的影响因素,如副词、否定表达、标点符号和表情符号等;第二,还未对情感表达手段的感情强度进行程度赋值,如比较级、程度副词对情感表达的影响等;第三,未考虑语言中的讽刺、比喻、隐喻等修辞手法的使用,这都会使得机器难以准确判断文本的情感倾向。可以说,上述这些问题都是未来我们工作的重点。

3.2 新闻评论领域的情感词典构建实验

通常来说,情感词典根据适用领域的不同分为通用型和领域型。通用情感词典能满足大部分情感分析任务的需求。然而为解决特定领域的情感分析任务并提高分析精度,需要使用领域情感词典,即根据某领域大量语料构建的情感词库,具有领域特定、时效性高等特点。在上文中,我们已经对 8031 条新闻评论中的情感信息进行标注,将表达情感语义的名词(3619 个)、动词(4147 个)、形容词(5930 个)和固定搭配(4346 个)进行标注。这里我们借助已标注信息,经过形态分析还原和去重处理后,初步自动抽取形成一部面向新闻评论领域的情感词典,包括各类表达手段总计 6321 条。这里限于篇幅,将情感词典片段示例如表₃。

表₃ 自动抽取的新闻评论领域情感词典片段

情感类	子类	表达手段			
		名词	形容词	动词	固定搭配
好	愉	радость\удовольствие	радостный\весёлый	радоваться\веселиться	в восторге\быть в восторге
	舒	успокоение\удобство	спокойный\удобный	успокоиться\утешаться	как ни в чем не бывало\без напряга
	敬	почтение\уважение\честь	почтенный\почтительный	уважать\почитать\почтить	с почтением\с уважением\с честью
	扬	красавица\отличник\талант	отличный\разумный	хвалить\одобрить	лучшие из лучших\высший класс
	信	доверие\вера\уверенность	верный\убеждённый	доверяться\полагаться	был уверенным в\иметь уверенность в
	爱	привязанность\охота	любимый\влюблённый	любить\нравиться	по душе\по сердцу\располагать к себе
恶	愿	мечта\надежда\желание	здоровый\благополучный	надеяться\поздравлять	ни пуха ни пера\счастливого пути
	烦	беспокойство\тоска\скука	беспокойный\печальный	надоедать\беспокоить	впасть в смуту\прийти в расстройство
	憎	антипатия\ненависть\злоба	ненавистный\отвратительный\возмутительный	ненавидеть\злобствовать	ненавидеть до мозга костей\питать злобу
	贬	идиот\лицемер\кретин	неискренний\фальшивый	ругать\охаивать\оскорблять	иметь одну только внешность
	妒	ревность\зависть	ревнивый\завидный	ревновать\завидовать	танть зависть\ревновать к заслугам других
	疑	недоверие\мнительность	подозрительный	сомневаться\подозревать	вериться с трудом\строить догадки
哀	怒	гнев\ярость\раздражение\возмущение	гневный\раздражающий\удручённый	сердиться\раздразниться\злиться	выходить из себя\в гневе
	悲	мука\страдание\печаль	печальный\плаченный	скорбеть\печалиться	болеть душой\в глубокой печали
惧	憾	сожаление\досада	безнадёжный\досадливый	сожалеть\досадовать	терять надежду\впасть в отчаяние
	恐	страх\кошмар\опасность	страшный\испуганный	бояться\пугаться\трусить	терять голову
	羞	стыд\позор\стыдливость	стыдный\пристыженный	смышаться\конфузиться	укоры совести\утрачения совести
	惊	удивление\тревога	удивлённый\тревожный	тревожиться\содрогнуться	сердце тревожится беспрестанно
	慌	смятение\паника\волнение	растерянный\торопливый	суетиться\волноваться	в замешательстве\в растерянности

4 结束语

本文分析俄语情感表达的基本手段,将情感描述为 4 大类、19 子类的分类体系,设计面向俄语情感分析的语料库标注体系并在新闻评论语料中进行验证,并基于标注语料进行俄语情感分析实验并初步构建包含 6321 条情感表达手段的情

感词典。可以说,俄语情感词典和标注语料库将为语言学诸多研究工作提供数据支持和研究思路,同时也可以为俄语文本倾向性分析、话题检测、舆情监控等自然语言处理研究与应用提供训练语料和核心数据资源,后续研究潜力巨大。对俄语情感表达手段的研究还需进一步深入,情感

词典与语料库规模还有待进一步扩展与提升,语料库结构和标注体系有待加强,情感倾向性判定算法还应继续优化,期待后续研究在这些方面能有新的发现与进步。

注释

- ①参见:<http://sentistrength.wlv.ac.uk>
 ②参见:http://www.i-teco.ru/solutions/business_intelligence_products/analytical_courier
 ③参见:<http://www.vaal.ru>
 ④参见:http://www.rco.ru/product.asp?ob_no=5047
 ⑤参见:<https://ria.ru/search/?query=трамп>
 ⑥参见:<http://www.aot.ru>
 ⑦参见:<http://www.keva.ru/stemka>
 ⑧参见:<https://tech.yandex.ru/mystem>
 ⑨TEI(Text Encoding Initiative)是机读语篇的国际信息编码方案,普遍用于大型语料库标注工作中,俄语国家语料库(НКРЯ)的标注体系也同TEI相兼容。

参考文献

- 蔡晖. 俄语情感动词的语义聚合体[J]. 中国俄语教学, 2014(4). || Cai, H. Semantic Aggregation of Russian Affective Verbs [J]. *Teaching Russian in China*, 2014(4).
 曹广跃. 俄语情感类心理动词的级次性研究[D]. 哈尔滨师范大学博士学位论文, 2016. || Cao, G.-Y. A Study of the Grades of Emotional Psychological Verbs in Russian[D]. Harbin Normal University, 2016.
 李慧. 带情感意义俄语成语的综合研究[D]. 辽宁大学博士学位论文, 2013. || Li, H. Comprehensive Research on the Russian Phrase with Emotion[D]. Liaoning University, 2013.
 林楠. 俄语词汇语义体系中的情感范畴化问题研究[D]. 首都师范大学, 2011. || Lin, N. *A Study of Affective Categorization in Russian Lexical Semantic System* [D]. Capital Normal University, 2011.
 汪正中 张洪渊. 基于英文博客文本的情感分析研究[J]. 计算机技术与发展, 2011(8). || Wang, Z.-Z., Zhang, H.-Y. Research of Sentiment Analysis on English Blog Text[J]. *Computer Technology And Development*, 2011(8).
 王向丽. 论俄语情感意义的表达手段[D]. 山东大学博士学位论文, 2005. || Wang, X.-L. Means of Express-

ing Russian Emotional Meaning[D]. Shandong University, 2005.

- 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(1). || Xu, L.-H., Lin, H.-F., Zhao, J. Construction and Analysis of Emotional Corpus [J]. *Journal of Chinese Information Processing*, 2008(1).
 杨利芳. 现代俄语中情感评价语义的范畴化[J]. 解放军外国语学院学报, 2012(3). || Yang, L.-F. Categorization of Sentiment Evaluation Meanings in Modern Russian[J]. *Journal of PLA University of Foreign Languages*, 2012(3).
 杨瑞. 俄语情感词的语言意识研究[J]. 吉林省教育学院学报, 2015(7). || Yang, R. A Study of Language Awareness of Russian Sentiment Words[J]. *Journal of Educational Institute of Jilin Province*, 2015(7).
 杨婷. 当代俄语报刊标题的情感—表现力手段研究[D]. 上海外国语大学, 2014. || Yang, T. *A Study on the Affective-expressive Means of Contemporary Russian Newspaper Titles*[D]. Shanghai International Studies University, 2014.
 伊尔夏提·吐尔贡 吾守尔·斯拉木 热西旦木·吐尔洪太 于清. 维吾尔文情感语料库的构建与分析[J]. 计算机与现代化, 2017(4). || Tuerhong, Y., Silamu, W., Tuserhongtai, R., Yu, Q. Construction and Analysis of Uighur Emotional Corpus[J]. *Computer and Modernization*, 2017(4).
 于金玲. 浅议俄语中的情感—表情词汇[J]. 东北农业大学学报(社会科学版), 2010(6). || Yu, J.-L. Words of Emotion-expression in Russian[J]. *Journal of Northeast Agricultural University (Social Science Edition)*, 2010(6).
 袁斌 江涛 于洪志. 基于语义空间的藏文微博情感分析方法[J]. 计算机应用研究, 2016(3). || Yuan, B., Jiang, T., Yu, H.-Z. Emotional Classification Method of Tibetan Micro-blog Based on Semantic Space[J]. *Application Research of Computers*, 2016(3).
 原伟. 俄汉新闻可比语料库的构建、评估及应用展望[J]. 解放军外国语学院学报, 2017(6). || Yuan, W. Building Russian-Chinese Comparable News Corpora, Their Comparability Evaluations and Further Applications[J]. *Journal of PLA University of Foreign Languages*, 2017(6).