

朝鲜语特殊短语结构的语言特征分析及自动提取* Structure Characteristic Analysis and Automatic Ex- traction for Korean Special Phrase

毕玉德, 扈毅, 安帅飞

BI Yude, HU Yi, AN Shuaifei

(战略支援部队信息工程大学洛阳校区, 河南洛阳 471003)

(Luoyang Campus, Information Engineering University of PLA Strategic Support Forces,
Luoyang, Henan 471003, China)

摘要: 朝鲜语中存在大量特殊短语结构, 因此在朝汉翻译中, 如何准确翻译这些特殊短语显得尤为重要, 此举有利于提高机器翻译的精度与效率。本文基于韩国“世宗计划”标注语料库, 通过对特殊短语结构进行语言特征分析, 构建规则库, 以迭代方式提取特殊短语结构及其分布, 并以中心词为“-따르다”的特殊短语为例, 进行自动提取实验, 取得满意的效果。

关键词: 朝鲜语 标注语料库 特殊短语 自动提取

中图分类号: TP391.1 **文献标识码:** A **文章编号:** 1002-7378(2018)01-0057-06

Abstract: There are a large number of special phrase structures in Korean, so it is especially important to accurately translate these special phrases in the process of Korean-Chinese translation, which is helpful to promote the accuracy and efficiency of machine translation. This article was based on the tagged corpus made by Korea's Sejong plan. Through analyzing the language features of special phrases, a rule base was constructed, and the special phrase structure and its distribution were extracted with iterative method. The headword of “-따르다” in special phrases was used as an example to conduct an automatic extraction experiment and achieved satisfactory results.

Key words: Korean, tagged corpus, special phrase, automatic extraction

0 引言

机器翻译, 又称自动翻译, 是利用计算机将一种语言翻译成另一种语言的过程^[1]。其概念产生于20世纪30年代, 经过几十年的研究、发展, 特别是随着计算机技术的高速发展, 机器翻译已逐步进入了人们的日常生活中。那么, 如何更好、更快、更

高效地进行机器翻译便成为此项技术的关键问题。但目前, 机器翻译的译文有时还比较生硬, 语句不够通顺, 甚至由于受存储单词和专业术语的限制, 会出现错译, 翻译后通常需要进行人工校译和编辑加工。

朝鲜语中存在大量特殊短语结构, 如果能在朝汉翻译中准确地翻译这些特殊短语, 将对提高朝汉机器翻译的准确率和效率起到十分关键的作用。首先, 句子的翻译往往会因为语序混乱而导致译文不通顺, 而通过提取特殊短语, 建立相应的短语库, 翻译时便可以方便快捷地得到语序正确的短语译文。其次, 特殊短语中, 构成短语的固有动词很大一部分是一词多义, 短语有固定搭配, 如果能够建立相应的

收稿日期: 2017-11-10

作者简介: 毕玉德(1967-), 男, 教授, 主要从事计算语言学、朝鲜语句法语义学研究, E-mail: biyude@gmail.com。

* 国家社会科学基金项目“面向语言信息处理的朝鲜语连接词尾语法知识库研究”(16BYY157)资助。

短语库,将能够更加快速、更加准确地进行匹配,提高翻译效率。故本文将基于标注语料库,对特殊短语结构进行语言特征分析,并将实现特殊短语的自动提取。

1 特殊短语结构的语言特征分析

短语是由两个或两个以上的词组成的,如:“학교에 가다/去学校,밥을 먹다/吃饭”等,但上述短语组合结构较为自由,本文研究的特殊短语,其结构较为固定,基本结构为:助词+中心词+词尾,如:“-에 대해, -을 비롯한”,也存在个别特殊情况,如:“-기 위해, -과 같이”。该类结构中谓词活用形式(谓词词干+词尾)有限,与之搭配的名词性成分的语法标记(助词)相对固定,在朝鲜语中出现频率较高。

本文将特殊短语按照短语结尾的类型分为以连接词尾结尾的特殊短语(例如:“말뭉치를 기반해/基于语料库”)和以冠形词转成词尾结尾的特殊短语(例如:“미국을 비롯한/以美国为首的”),但也存在特殊情况,即以一般副词结尾的特殊短语(例如:“다음과 같이”)。本文将分别对三者进行语言特征分析,为确定左右边界提供语言学基础。

1.1 以连接词尾结尾的特殊短语结构

该结构如下所示:

이번 훈련을 통해
此次 训练 通过

通过 此次 训练

上例中包含“이번(一般名词,代码:NNG)、훈련(一般名词,代码:NNG)、을(宾格助词,代码:JKO)、통하(一般动词,代码:VV)、아(接续词尾,代码:EC)”等成分。朝鲜语中,谓语(动词和形容词)占一个句子的支配地位。在一个句子中谓语作为语言的语义中心要素,反映着人的经验,它决定着什么样的名词以什么样的关系与之搭配^[2]。特殊短语作为一个小的句子单元,其中心词为该短语中的动词(例如:“이번 훈련을 통해”中的“통하다”是中心词)或形容词(例如:“오늘날과 같은”中的“같다”是中心词),是短语翻译的核心,与该中心动词搭配的名词也十分重要,只有两者相互合理搭配,短语才有实际意义,才能够完成对短语的翻译。

根据与动词搭配的名词性成分的语法标记(助词)分类如下,其中还存在包含名词性转成词尾“-기”的特殊情况。

(1) 包含副词格助词的特殊短语结构

该特殊短语的结构:NNG+JKB+[XR/VV/(NNG+XSV)]+EC。

NNG	JKB	XR/VV/(NNG+XSV)	EC
名词	副词格助词	词根/动词/(名词+动词派生后缀)	连接词尾

- (a) NNG+와/과(JKB)+비슷(XR)+하(XSV)+아(EC):
어제와 비슷하여/与昨天类似
- (b) NNG+에(JKB)+따르(VV)+아(EC):
보도에 따라/据报道
- (c) NNG+에(JKB)+입각(NNG)+하(XSV)+아(EC):
이 시각에 입각해/立足于此视角

(2) 包含宾格助词的特殊短语结构

该特殊短语的结构:NNG+JKO+[XR/VV/(NNG+XSV)]+EC。

NNG	JKO	XR/VV/(NNG+XSV)	EC
名词	宾格助词	词根/动词/(名词+动词派生后缀)	连接词尾

- (a) NNG+를/을(JKO)+기(XR)+하(XSV)+아(EC):
3월 1일을 기해/以3月1日为起点
- (b) NNG+를/을(JKO)+위하(VV)+아(EC):
공산주의를 위하여/为了共产主义事业
- (c) NNG+를/을(JKO)+제외(NNG)+하(XSV)+고(EC): 학생을 제외하고/除了学生

(3) 包含接续助词的特殊短语结构

该特殊短语的结构:NNG+JC+[XR/VV/(NNG+XSV)]+EC。

NNG	JC	XR/VV/(NNG+XSV)	EC
名词	接续助词	词根/动词/(名词+动词派生后缀)	连接词尾

- (a) NNG+와/과(JC)+비슷(XR)+하(XSV)+아(EC):
남자와 비슷하여/与男人相类似
- (b) NNG+와/과(JC)+더불어(VV)+어(EC):
그와 더불어/和他一起
- (c) NNG+와/과(JC)+관련(NNG)+하(XSV)+아(EC):
영토문제와 관련해/有关领土问题

(4) 包含名词性转成词尾的特殊短语结构

该特殊短语的结构:NNG+JKO+[VV/(NNG+XSV)]+ETN+VV+EC。

NNG	JKO	XR/VV/(NNG+XSV)	ETN	VV	EC
名词	宾格助词	词根/动词/(名词+动词派生后缀)	名词性转成词尾	动词	连接词尾

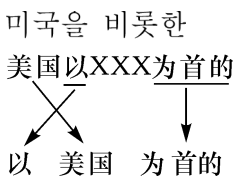
(a) NNG+를/을(JKO)+맞이(XR)+하(XSV)+기(ETN)+
 위하(VV)+아(EC):손님을 맞이하기 위해/为迎接
 客人

(b) NNG+를/을(JKO)+쏘(VV)+기(ETN)+위하(VV)+
 아(EC):나로호를 쏘기 위해/为发射罗老号

(c) NNG+를/을(JKO)+실현(NNG)+하(XSV)+
 기(ETN)+위하(VV)+아(EC):
 통일을 실현하기 위해/为实现统一

1.2 以冠形词转成词尾结尾的特殊短语结构

该结构如下所示:



该短语结构与以连接词尾结尾的特殊短语类

似,具体的分类如下:

(1)包含副词格助词的特殊短语结构

该特殊短语的结构:NNG + JKB + [XR/VV/
 (NNG + XSV)] + ETM.

NNG	JKB	XR/VV/(NNG+XSV)	ETM
名词	副词格助词	词根/动词/(名词+动词派生后缀)	冠形词转成词尾

(a) NNG+와/과(JKB)+무관(XR)+하(XSV)+
 ㄴ(ETM):대국과 무관한/无关大局的

(b) NNG+에(JKB)+대하(VV)+ㄴ(ETM):
 이 영화에 대한/对于这部电影的

(c) NNG+에(JKB)+관련(NNG)+하(XSV)+ㄴ(ETM):
 학교폭력 사건에 관련한/有关校园暴力事件的

(2)包含宾格助词的特殊短语结构

该特殊短语的结构:NNG + JKO + [XR/VV/
 (NNG + XSV)] + ETM.

NNG	JKO	XR/VV/(NNG+XSV)	ETM
名词	宾格助词	词根/动词/(名词+动词派生后缀)	冠形词转成词尾

(a) NNG+를/을(JKO)+맞이(XR)+하(XSV)+ㄴ(ETM):
 회갑을 맞이한/年近花甲的

(b) NNG+를/을(JKO)+위하(VV)+ㄴ(ETM):
 문화 발전을 위하/为了文化发展的

(c) NNG+를/을(JKO)+포함(NNG)+하(XSV)+
 ㄴ(ETM):연금을 포함한/包括退休金在内的

(3)包含接续助词的特殊短语结构

该特殊短语的结构:NNG + JC + [XR/VA/
 (NNG + XSV)] + ETM.

NNG	JC	XR/VA/(NNG+XSV)	ETM
名词	接续助词	词根/形容词/(名词+动词派生后缀)	冠形词转成词尾

(a) NNG+와/과(JC)+무관(XR)+하(XSV)+ㄴ(ETM):
 자기와 무관하다는/与自己无关的

(b) NNG+와/과(JC)+같(VA)+은(ETM):
 이 소리, 냄새와 같은/与这种声音和气味相似的

(c) NNG+와/과(JC)+비교(NNG)+하(XSV)+ㄴ(ETM):
 스탈린주의와 비교한/相比于斯大林主义的

(4)包含名词性转成词尾的特殊短语结构

该特殊短语的结构:NNG + JKO +
 [VV/(NNG + XSV)] + ETN + VV + ETM.

NNG	JKO	XR/VV/(NNG+XSV)	ETN	VV	ETM
名词	宾格助词	词根/动词/(名词+动词派生后缀)	名词性转成词尾	动词	冠形词转成词尾

(a) NNG+를/을(JKO)+기(XR)+하(XSV)+기(ETN)+
 위하(VV)+ㄴ(ETM):통일을 기하기 위하/以求
 统一的

(b) NNG+를/을(JKO)+되(VV)+기(ETN)+위하(VV)+
 ㄴ(ETM):배우를 되기 위하/为了成为演员的

(c) NNG+를/을(JKO)+확립(NNG)+하(XSV)+기(ETN)+
 위하(VV)+ㄴ(ETM):제도를 확립하기 위하/为了确立制度的

1.3 以一般副词结尾的特殊短语结构

特殊短语结构也存在特殊情况,即结尾为一般副词的特殊短语,如:“다음과 같이/如下”。这种特殊短语结构中,与中心词相连的助词分类方法不同,仅当短语包含副词格助词与接续助词时有实际意义,故对该特殊短语结构进行如下分类:

(1)包含副词格助词的特殊短语结构

该特殊短语的结构:NNG + JKB + MAG.

NNG	JKB	MAG
名词	副词格助词	一般副词

例如:

NNG+와/과(JKB)+같이(MAG):다음과 같이/如下

(2)包含接续助词的特殊短语结构

该特殊短语的结构为:NNG + JC + MAG.

NNG	JC	MAG
名词	接续助词	一般副词

例如:

NNG+와/과(JC)+달리(MAG):대중문화와 달리/与
 大众文化不同

2 特殊短语的左右边界问题

自动提取特殊短语,即按照相关规则,提取范围内的短语。而实现自动提取的核心则是确定其左右边界。首先确定左边界,特殊短语的最左端为单个名词或名词短语,故其左边界应为在助词前与动词搭配的名词性成分(如:“자료에 따라”中副词格助词“-에”前与“따르다”搭配的名词“자료”,“단장님을 비롯한”中宾格助词“-을”前与“비롯하다”搭配的名词“단장님”)。安帅飞等^[3]利用世宗语料库中的二十多万句子,提取归纳了单个名词之外的8大类名词短语,名词短语的左边界等同于特殊短语的左边界。本文据此,构建了特殊短语的左边界规则集。

然后确定右边界,通过语料中的大量特殊短语实例可以发现特殊短语的最右端为连接词尾、冠形词转成词尾以及一般副词,故右边界为连接词尾(例如:“자료에 따라”中动词“따르다”后的连接词尾“-아”)、冠形词转成词尾(例如:“단장님을 비롯한”中动词“비롯하다”后的冠形词转成词尾“ㄴ”)以及一般副词(例如:“다음과 같이”中的一般副词“같이”)。故将其右边界分为上述3类。

以包含副词格助词的特殊短语为例,该特殊短语左边界为名词,包括单个名词与名词短语,将左边界按照单个名词与名词短语进行分类研究,具体情况如下。

2.1 右边界为连接词尾的特殊短语的左边界问题

(1)左边界为单个名词的特殊短语

이(NP)+에(JKB)+따르(VV)+아(EC)/据此

(2)左边界为名词短语的特殊短语,可分为以下8种情况^[3]:

①名词|代词+의+名词|名词叠加

오늘(NNG)+의+상황(NNG)+에(JKB)+의하(VV)+아(EC)/依据今天的情况

②两个或两个以上名词(代词)混合叠加

교육(NNG)+문제(NNG)+에(JKB)+대하(VV)+아(EC)/对于教育问题

③名词|代词+接续助词|特殊的副词+名词|代词

전통(NNG)+과+습관(NNG)+에(JKB)+의하(VV)+아(EC)/按照传统与习惯

④冠形词+名词|代词

이런(MM)+물음(NNG)+에(JKB)+관하(VV)+아(EC)/有关此类疑问

⑤数字|数词+名词

21(SN)+세기(NNG)+에(JKB)+관하(VV)+아(EC)/有关21世纪

⑥名词|名词叠加+적+名词

사회(NNG)+적+필요(NNG)+에(JKB)+따르(VV)+아(EC)/根据社会的需要

⑦名词+名词派生词尾+肯定指示词+冠形词性转成词尾+名词

보편(NNG)+적(XSN)+이(VCP)+ㄴ(ETM)+가지(NNG)+에(JKB)+의하(VV)+아(EC)/依据普遍的价值

⑧名词|代词+数词+(依存名词)

나(NP)+하나(NR)+에(JKB)+대하(VV)+아(EC)/对于我一个人而言

2.2 右边界为冠形词转成词尾的特殊短语的左边界问题

(1)左边界为单个名词的特殊短语

문화(NNG)+에(JKB)+대하(VV)+ㄴ(ETM)/对于文化的

(2)左边界为名词短语的特殊短语,可分为以下8种情况^[3]:

①名词|代词+의+名词|名词叠加

고조선(NNG)+의+정체(NNG)+에(JKB)+대하(VV)+ㄴ(ETM)/对于古朝鲜政体的

②两个或两个以上名词(代词)混合叠加

군사(NNG)+분야(NNG)+에(JKB)+대하(VV)+ㄴ(ETM)/对于军事领域的

③名词|代词+接续助词|特殊的副词+名词|代词

조선(NNP)+과+일본(NNP)+관계(NNG)+에(JKB)+관하(VV)+ㄴ(ETM)/有关朝鲜与日本关系的

④冠形词+名词|代词

이(MM)+이론(NNG)+에(JKB)+의하(VV)+ㄴ(ETM)/根据该理论的

⑤数字|数词+名词

18(SN)+세기(NNG)+에(JKB)+관하(VV)+ㄴ(ETM)/有关18世纪的

⑥名词|名词叠加+적+名词

철학(NNG)+적+문제(NNG)+에(JKB)+대하(VV)+ㄴ(ETM)/对于哲学问题的

⑦名词+名词派生词尾+肯定指示词+冠形词性转成词尾+名词

보편(NNG)+적(XSN)+이(VCP)+ㄴ(ETM)+가지(NNG)+에(JKB)+관하(VV)+ㄴ(ETM)/有关普世价值的

⑧名词|代词+数词+(依存名词)

방(NNG)+하나(NR)+에(JKB)+관하(VV)+ㄴ(ETM)/有关一间房子的

2.3 右边界为一般副词的特殊短语的左边界问题

(1)左边界为单个名词的特殊短语

다음(NNG)+과(JKB)+같이(MAG)/如下

(2)左边界为名词短语的特殊短语,可分为以下8种情况^[3]:

①名词|代词+의+名词|名词叠加

제사(NNG)+의+절차(NNG)+와(JKB)+같이(ETM)/与祭祀程序相类似

②两个或两个以上名词(代词)混合叠加

막(NNG)+목사(NNG)+와(JKB)+함께(MAG)/和朴牧师一起

③名词|代词+接续助词|特殊的副词+名词|代词

전통(NNG)+과+습관(NNG)+와(JKB)+같이(ETM)/与传统习惯相类似

④冠形词+名词|代词

한(MM)+TV 프로그램(NNG)+에(JKB)+불과(MAG)/不过是一个电视节目

⑤数字|数词+名词

일(NR)+년(NNB)+에(JKB)+불과(MAG)/不过一年

⑥名词|名词叠加+적+名词

형식(NNG)+적+필요(NNG)+에(JKB)+불과(MAG)/不过是形式上的需要

⑦名词+名词派生词尾+肯定指示词+冠形词性转成词尾+名词

보편(NNG)+적(XSN)+이(VCP)+ㄴ(ETM)+가지(NNG)+과(JKB)+같이(MAG)/如普世价值一样

⑧名词|代词+数词+(依存名词)

방(NNG)+하나(NR)+에(JKB)+불과(MAG)/不过一间房子

3 特殊短语结构提取及短语库的建立

Perl语言的最初设计者为拉里·沃尔(LARRY WALL),Perl语言借取了C、sed、awk、shell scripting以及很多其他程序语言的特性。其中最重要的特性是它内部集成了正则表达式的功能^[4]。而实现特殊短语的自动提取的关键,就是Perl语言的正则表达式功能,根据相应的语言特征和确定的左右边界,自动提取规则内的短语,这是建立短语库的重要一环。

另外,短语库的建立还需要对自动提取的短语进行人工筛选,选出有实际意义的短语加入短语库。

3.1 特殊短语结构自动提取

将该特殊短语按照助词的种类进行分类,根据其语言特征,使用Perl语言实现特殊短语自动提取,即以助词为提取特殊短语结构的核心。以包含副词格助词(JKB)的特殊短语为例,对特殊短语的自动提取进行说明。

基于韩国“世宗计划”标注语料库中的20余万句语料,共提取出56216个短语结构,其部分结果如表1所示。

表1 部分特殊短语的提取结果

Table 1 Some extractions of special phrases

序号 No.	特殊短语结构 Structures of special phrase	频数 Frequency
1	에/JKB 대하/VV ㄴ/ETM	5 808
2	에/JKB 대하/VV 아/EC	2 338
3	에/JKB 따르/VV 아/EC	2 190
4	에/JKB 의하/VV 아/EC	1 348
5	에/JKB 대하/VV 아서/EC	1 159
6	에/JKB 관하/VV ㄴ/ETM	1 084
7	와/JKB 같/VA 은/ETM	854
8	과/JKB 같/VA 은/ETM	817
9	에/JKB 따르/VV 면/EC	784
10	에/JKB 의하/VV ㄴ/ETM	556

自动提取的结果存在非特殊短语的结构(如“-에 지나지”等)(表1),故需要进行二次分析。观察结果分析得出,符合以下规则的短语不属于特殊短语。

连接词尾后与补助动词(VX)相连的短语,如下例所示:

(a)에(JKB)+자리(NNG)+잡(VV)+고(EC)+있(VX)+다(EF)

(b)에(JKB)+지나(VV)+지(EC)+않(VX)+아요(EF)

(c)으로(JKB)+알리(VV)+러(EC)+지(VX)+있(EP)+다(EF)

经过二次分析,根据规则再次抽取短语,避免了上述不符合特殊短语结构情况的发生,取得了较好的效果。采用自动提取与人工筛选相结合的方式,按照不同的中心词进行分类,共抽取了50余种特殊短语结构:

대하다, 의하다, 따르다, 관하다, 같다, 비하다, 인하다, 더볼다, 맞추다, 관련하다, 불과하다, 속하다, 걸치다, 입각하다, 비슷하다, 근거하다, 기만하다, 비롯하다, 삼다, 말미암다, 기하다, 포함하다, 대비하다, 다르다, 의존하다, 반하다, 비교하다, 의거하다, 일치하다, 관계되다, 무관하다, 통하다, 통과하다, 즈음하다, 직면하다, 경과하다, 빙자하다, 기초하다, 건주다, 제외하다, 구분하다, 연관되다, 시작하다, 비추다, 대비하다, 거치다, 여기다, 겨누다, 조준하다。

3.2 特殊短语库的建立

根据自动提取结果,依据所总结的规则,进行人工筛选,最终选取符合规则的短语加入短语库。以包含副词格助词(JKB)的特殊短语结构为例,对短语库的建立进行说明。

将经过自动提取与人工筛选的结果按照不同的中心词进行分类,部分分类结果如表2所示。

表2 特殊短语库

Table 2 Corpus of special phrase

中心词 Headword	特殊短语结构 Structures of special phrase	译文 Translation
-따르다	에(JKB)+따르(VV)+아(EC)	根据XX
	에(JKB)+따르(VV)+면(EC)	根据XX
	에(JKB)+따르(VV)+아서(EC)	根据XX
	에(JKB)+따르(VV)+ㄹ(ETM)	根据XX的
	에(JKB)+따르(VV)+는(ETM)	根据XX的
-대하다	에(JKB)+대하(VV)+아(EC)	对于XX
	에(JKB)+대하(VV)+아서(EC)	对于XX
	에(JKB)+대하(VV)+ㄹ(ETM)	对于XX的
-의하다	에(JKB)+의하(VV)+아(EC)	依据XX
	에(JKB)+의하(VV)+면(EC)	依据XX
	에(JKB)+의하(VV)+아서(EC)	依据XX
	에(JKB)+의하(VV)+며(EC)	依据XX
	에(JKB)+의하(VV)+ㄹ(ETM)	依据XX的

3.3 实例测试

对信息提取系统的评价,最为常用的两个指标是召回率和准确率^[5],因此这两个指标也是衡量特殊短语自动提取成功与否的重要标准。

根据所确定的特殊短语的左右边界,编写相应代码,使用计算机对特殊短语进行自动提取,人工判断结果是否全部被召回,是否全部正确,并计算其召回率和准确率。

选取5000句训练语料,以包含“-에 따르다”且右边界为连接词尾的特殊短语为例进行测试,部分结果如表3所示。

经过人工筛选得出,该语料中共有48个包含“-에 따르다”且右边界为连接词尾的特殊短语,自动提取出48个结果,召回率为100%,提取结果中,出现3个错误结果,准确率为93.8%。其中的错误结果均是由于名词前做修饰成分的定语过长造成的,但此种情况在朝鲜语中并不多见,且长定语结构较

为复杂,难以研究其规则,故可以暂不考虑此种情况的发生。所以总体上,测试取得了满意的效果。

表3 基于5000句语料的部分提取结果

Table 3 Some extractions from 5 000 Korean sentences

序号 No.	特殊短语 Special phrases	频数 Frequency
1	이/NP 에/JKB 따르/VV 아/EC	6
2	그/NP 에/JKB 따르/VV 면/EC	2
3	경우/NNG 에서/JKB 따르/VV 아/EC	2
4	문항/NNG 에서/JKB 따르/VV 아/EC	2
5	실력/NNG 에/JKB 따르/VV 아/EC	1
6	개혁/NNG 안/NNG 에/JKB 따르/VV 면/EC	1
7	이/MM 지적/NNG 에/JKB 따르/VV 아/EC	1
8	이/NP 에/JKB 따르/VV 면/EC	1

4 结束语

本文通过分析特殊短语的语言特征,寻找特殊短语自动提取的左右边界,进而建立规则,编写代码,实现自动提取,建立特殊短语库,并经过测试,取得了满意的效果。

但本文的工作还有两点不足:(1)虽然本文依据的标注语料库达50万句之多,但语料所涉及的范围仍旧有限,所以有必要适当扩大语料的范围。(2)虽然召回率与准确率都较高,但错误结果仍然存在,所以需要进一步分析论证,将错误率降低。

参考文献:

- [1] 赵铁军. 机器翻译原理[M]. 哈尔滨:哈尔滨工业大学出版社,2000.
ZHAO T J. Theory of machine translation[M]. Harbin: Harbin Institute of Technology Press, 2000.
- [2] 毕玉德. 现代朝鲜语动词语义组合关系研究[M]. 北京:民族出版社,2005:28.
BI Y D. Research on the semantic combinations of modern Korean verbs[M]. Beijing: Nationalities Publishing House, 2005:28.
- [3] 安帅飞,毕玉德. 韩国语名词短语结构特征分析及自动提取[J]. 中文信息学报,2013,27(5):205-210.
AN S F, BI Y D. Structure characteristic analysis and automatic extraction of Korean noun phrase[J]. Journal of Chinese Information Processing, 2013, 27(5): 205-210.
- [4] 施瓦茨,菲尼克斯,福瓦. Perl语言入门[M]. 盛春,蒋永清,王晖,译. 南京:东南大学出版社,2012.
SCHWARTZ R L, PHOENIX T, FOY B. Learning Perl[M]. SHENG C, JIANG Y Q, WANG H (eds.). Nanjing: Southeast University Press, 2012.
- [5] 俞士汶. 计算语言学概论[M]. 北京:商务印书馆,2003.
YU S W. Introduction of computational linguistics [M]. Beijing: Commercial Press, 2003.

(责任编辑:米慧芝)