

基于规则的哈萨克语动词短语识别研究

古丽拉·阿东别克^{1,2,3}, 古丽扎达·海沙^{1,3}

(1.新疆大学 信息科学与工程学院, 乌鲁木齐市 830046; 2. 新疆多语种信息技术实验室

3.国家语言资源监测与研究中心少数民族语言中心哈萨克和柯尔克孜语文基地

摘要: 本文进行了哈萨克语基本动词短语 (KzBaseVP) 的定义、性质、分类、结构, 确立了哈萨克语基本动词短语的句法功能分类框架, 初步描述了现代哈萨克语短语结构所需要的句法体系和比较完整短语功能分类体系。对哈萨克语基本动词短语的结构进行了统计与分析; 接着确定哈萨克语基本动词短语的定界确定规则, 最终识别了哈萨克语基本动词短语, 实验结果表明: 哈萨克语动词短语的识别精确率达到 87.92%。

关键词: 哈萨克语; 动词短语; 识别; 规则集

Research on identification of Verb phrases in Kazakh with Rule

Gulila Altenbek^{1,2,3}, Gulizada Haisa^{1,3}

(1.Department of Information Science and Engineering College, Xinjiang University, Urumqi, Xinjiang 830046, China

2.Xinjiang Laboratory of Multi-language Information Technology

3. The Base of Kazakh and Kirghiz Language of National Language Resource Monitoring and Research Center Minority Languages)

Abstract: This study clearly put forward the Kazakh basic verb phrase (KzBaseVP) definition, nature, classification, structure, and established the Kazakh basic verb phrase syntactic function classification framework, the initial description of the modern Kazakh phrase structure of the syntactic system and relatively complete phrases functional classification system. Second, statistics and analysis the Kazakh basic structure of the verb phrase; followed to determine the Kazakh basic verb phrase collocation, and extract the Kazakh basic verb phrase base on rules. Experiments reported the accuracy rate reached 87.92%.

Key words: Kazakh; Verb Phrase; Identification; Rules collection

1 引言

哈萨克语在语序结构上属于“SOV 型”语言, 哈萨克语自然语言信息处理技术在文字处理、词法分析、资源库建设等阶段均取得了一定的成果, 目前已进入句法处理阶段, 即如何自动分析短语结构、短语定界、短语内部句法关系分析等。短语结构识别是其句法结构研究的核心, 为哈萨克语句法分析阶段提供良好的基础; 为机器翻译、搜索引擎、文本分类、信息提取等应用研究提供很好的中介作用。

浅层句法分析和短语识别研究领域里在国外研究者有 Abeny^[2], Church^[3] 等人, 国内研究者有赵军, 黄昌宁^[4,5], 周强, 孙茂松^[6], 周雅倩^[7], 李素建, 刘群^[8], 李荣, 曹建芳^[1], 白妙青, 郑家恒^[9], 霍亚格, 黄广君^[10] 等人, 少数民族研究者有达胡白乙

拉^[11], 古丽拉^[12,13] 等人, 都获得了一定的研究成果。

本文重点研究哈萨克语动词短语自身特点, 使用基于规则的方法, 研究与实现哈萨克语动词短语自动识别。

2 哈萨克语基本动词短语

哈萨克语中短语按功能可分为两大类: 名词短语 (NP) 和动词短语 (VP)。本文围绕哈萨克语动词语法特点和结构研究, 参考英语和汉语的基本动词短语特点, 结合哈萨克语自身语法特点给出哈萨克语基本动词短语的定义如下:

定义: 哈萨克语基本动词短语 (Kazakh Base Verb Phrase, 缩写形式: KzBaseVP) 是以动词为中心, 不交叉、不包含内嵌基本动词短语和其他词类基本短语的语言单位。基本动词短语用法同动词, 在句中充

当谓语，主语，定语等成分。KzBaseVP 的中心词是动词，通常句子的核心词在句末。大多数哈萨克语基本动词短语在哈萨克语句末，很少出现在句中和句首。

哈萨克语的动词短语是表示行为动作，内心活动，事物发展变化的短语。通常动词或者其他词类配合使用。

3 哈萨克语浅层句法

短语识别在自然语言处理中居于词汇和句子中间的一种语法单位。哈萨克语短语是由两个或者两个以上单词在词汇意义和语法结构上有联系的词构成的语言单位是短语。

哈语是外部形态变化丰富的语言，而这种形态在短语结构上表现最突出，哈语短语可分为并列结构 (سالالاس قۇرلىمى)，一致结构 (قييسۇ قۇرلىمى)，支配结构 (مەمكەتەرۇ قۇرلىمى)，领属结构 (ماتاسۇ قۇرلىمى)，附加结构 (قابىسۇ قۇرلىمى)，邻近结构 (جاناسۇ قۇرلىمى)。如：
بىيل ءبىزدەك اۆلىمىز استىقتان مول ءونىم الدى.
(今年我们家乡粮食庄稼获得丰收)。

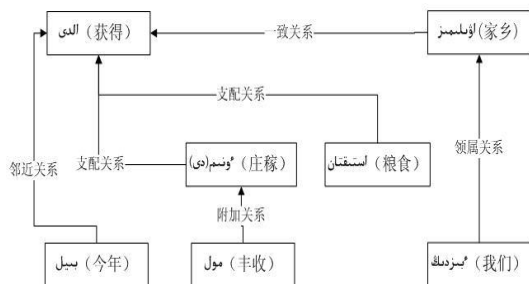


图1 哈萨克语形态结构示例

哈萨克语中有名词短语 (NP) 和动词短语 (VP) 两大类之分。进一步分为名词短语、动词短语、形容词短语、代词短语、数词短语、数量词短语、方位词短语、时间词短语、副词短语、后置词短语等 10 种 [14, 15, 16]。

本研究中采取的词性标注集为：v-动词，n-名词，adj-形容词，m-量词，num-数词，adv-副词，pron-代词，int-感叹词，ono-摹拟词，conj-助词等基本词性标注外，还有va表示动名词，vb 动词的各种时态，vc表示副动词，vd表示形动词。

4 哈语动词短语的定界规则

4.1 获取上下文信息的必要性

哈萨克语动词短语划分和自动标注中，应该考虑语法功能信息，即在动词短语内部搭配结构上，加上下文信息提供的信息、统计、分析，总结出动词短语出现的上下文信息和动词短语不肯出现的上下文信息，从而形成一系列上下文有关的规则。这样可以排除不可能构成动词短语的信息，并且提高系统的准确性。

4.2 动词短语规则库的建立

哈萨克语基本动词短语自动识别规则，包括边界识别规则和结构判定规则。

动词的性质和功能直接决定动词短语的范围，哈萨克语基本动词短语的界定涉及动词的各种类型与动词的态范畴，式范畴，时畴等问题，如图 2 所示，图 2 中前缀一般在外来词中应用，并且数量很少，因此可以忽略。

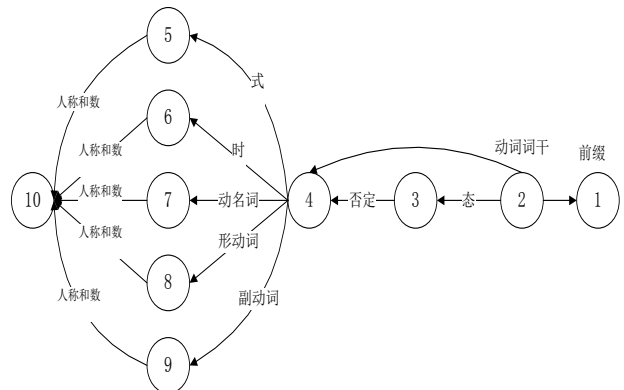


图2 哈萨克语动词范畴的连接规则框架

本文根据哈语基本动词短语自身的特点出发，采用基于规则的方法，归纳的动词短语规则进行形式化描述，最终建立了动词构形附加成分集，如表 1 所示。

从上述分析得出哈语动词附加成分中表示态范畴的有：dsen, dseo, dsez, dsee, dsed。表示时范畴的有：dsms, ds mk, dskz, dskd, dskb, dsh。表示态式畴的有：ds mo, ds ma, dsks, dskj, dsa。表示人称和数的附加成分有：dcr1, dcr2, dcr3。表示否定的附加成分有：dsl。根据哈语动词构形

附加成分标记集，建立了 17 个规则如下：

规则 1: vp->v1 v2

<v1>=<v[dsen|dseo|dsez|dsee|dsed|dsms|dsmk|dskz|dskd|dskb|dsh|dsmo|dsma|dsk|dskj|dsa|dcr1|dcr2|dcr3|dsl]>

<v2>=<v>

规则 2: vp->v1 v2

<v1>=<v>

<v2>=<v[dsen|dseo|dsez|dsee|dsed|dsms|dsmk|dskz|dskd|dskb|dsh|dsmo|dsma|dsk|dskj|dsa|dcr1|dcr2|dcr3|dsl]>

规则 3: vp->n v

<n>=<n>

<v>=<v[dsen|dseo|dsez|dsee|dsed|dsms|dskz|dskd|dskb|dsh|dsmo|dsma|dsk|dskj|dsa|dcr1|dcr2|dcr3|dsl]>

规则 4: vp->v1+v2+v3

<v1>=<v>

<v2>=<v[dsen|dseo|dsee|dsed|dsms|dsmk|dskz|dskd|dskb|dsh|dsmo|dsma|dsk|dskj|dsa|dcr1|dcr2|dcr3|dsl]>

<v3>=<v[dsen|dseo|dsez|dsee|dsms|dsmk|dskz|dskd|dskb|dsh|dsmo|dsma|dsk|dskj|dsa|dcr1|dcr2|dcr3|dsl]>

规则 5: vp->v1+v2+v3

<v1>=<v[dsen|dseo|dsez|dsee|dsed|dsms|dskz|dskd|dskb|dsh|dsmo|dsma|dsk|dskj|dsa|dcr1|dcr2|dcr3|dsl]>

<v2>=<v>

<v3>=<v[dsen|dseo|dsez|dsee|dsed|dsms|dsmk|dskz|dskd|dskb|dsh|dsmo|dsma|dsk|dskj|dsa|dcr1|dcr2|dcr3|dsl]>

规则 6: vp->v1+v2+v3

<v1>=<v[dsen|dseo|dsez|dsee|dsed|dsms|dsmk|dskz|dskd|dskb|dsh|dsmo|dsma|dsk|dskj|dsa|dcr1|dcr2|dcr3|dsl]>

<v2>=<v[dsen|dseo|dsez|dsee|dsed|dsms|dsmk|dskz|dskd|dskb|dsh|dsmo|dsma|dsk|dskj|dsa|dcr1|dcr2|dcr3|dsl]>

<v3>=<v>

规则 7: vp->v1+pron+v2

<v1>=<v>

<pron>=<pron>

<v2>=<v[dsen|dseo|dsez|dsee|dsed|dsms|dsmk|dskz|dskd|dskb|dsh|dsmo|dsma|dsk|dskj|dsa|dcr1|dcr2|dcr3|dsl]>

规则 8: vp->v1+pron+v2

<v1>=<v[dsen|dseo|dsez|dsee|dsed|dsms|dsmk|dskz|dskd|dskb|dsh|dsmo|dsma|dsk|dskj|dsa|dcr1|dcr2|dcr3|dsl]>

<pron>=<pron>

<v2>=<v>

规则 9: vp->adj v

<adj>=<adj>

<v>=<v[dsen|dseo|dsez|dsee|dsms|dsmk|dskz|dskd|dskb|dsh|dsmo|dsma|dsk|dskj|dsa|dcr1|dcr2|dcr3|dsl]>

kb|dsh|dsmo|dsma|dsk|dskj|dsa|dcr1|dcr2|dcr3|dsl]>

规则 10: vp->adv v

<adv>=<adv>

<v>=<v[dsen|dseo|dsez|dsee|dsed|dsms|dsmk|dskz|dskd|dskb|dsh|dsmo|dsma|dsk|dskj|dsa|dcr1|dcr2|dcr3|dsl]>

规则 15: vp->pron v

<pron>=<pron>

<v>=<v[dsen|dseo|dsez|dsee|dsed|dsms|dsmk|dskz|dskd|dskb|dsh|dsmo|dsma|dsa|dcr1|dcr2|dcr3|dsl]>

规则 11: vp->int v

<int>=<int>

<v>=<v[dsen|dseo|dsez|dsee|dsed|dsms|dsmk|dskz|dskd|dskb|dsh|dsmo|dsma|dsk|dskj|dsa|dcr1|dcr2|dcr3|dsl]>

规则 12: vp->n conj v

<n>=<n>

<conj>=<conj>

<v>=<v[dsen|dseo|dsez|dsee|dsms|dsmk|dskz|dskd|dskb|dsh|dsmo|dsma|dsk|dskj|dsa|dcr1|dcr2|dcr3|dsl]>

规则 13: vp->v1 conj v2

<v1>=<v[dsen|dseo|dsez|dsee|dsed|dsms|dsmk|dskz|dskd|dskb|dsh|dsmo|dsma|dsk|dskj|dsa|dcr1|dcr2|dcr3|dsl]>

<conj>=<conj>

<v2>=<v[dsen|dseo|dsez|dsee|dsed|dsms|dsmk|dskz|dskd|dskb|dsh|dsmo|dsma|dsk|dskj|dsa|dcr1|dcr2|dcr3|dsl]>

规则 14: vp->v1 conj v2

<v1>=<v[dsen|dseo|dsee|dsed|dsms|dsmk|dskz|dskd|dskb|dsh|dsmo|dsma|dsk|dskj|dsa|dcr1|dcr2|dcr3|dsl]>

<conj>=<conj>

<v2>=<v>

规则 15: vp->mum m v

<mum>=<num>

<m>=<m>

<v>=<v[dsen|dseo|dsez|dsee|dsed|dsms|dsmk|dskz|dskd|dskb|dsh|dsmo|dsma|dsk|dskj|dsa|dcr1|dcr2|dcr3|dsl]>

规则 16: vp->v1 prep v2

<v1>=<v>

<prep>=<prep>

<v2>=<v[dsen|dseo|dsez|dsee|dsed|dsms|dsmk|dskz|dskd|dskb|dsh|dsmo|dsma|dsk|dskj|dsa|dcr1|dcr2|dcr3|dsl]>

规则 17: vp->v1 prep v2

<v1>=<v[dsen|dseo|dsez|dsee|dsed|dsms|dsmk|dskz|dskd|dskb|dsh|dsmo|dsma|dsk|dskj|dsa|dcr1|dcr2|dcr3|dsl]>

<prep>=<prep>

<v2>=<v>

表 1 哈语动词构形附加成分集

类型	附加	类型	附加	类型	附加	类型	附加	类型	附加	类型	附加	类型	附加
dsms	مىشى	dsma	ەتىن	dskj	مايىنشا	dsh	شىلىق	dsez	ت	dsa	دى	dcr2	ئىز
dsms	شى	dsma	اتىن	dskj	مايىنشا	dsh	داي	dsez	ەر	dsa	مىدى	dcr2	ك
dsms	ماق	dsma	متىن	dskj	بەيىنشە	dsh	ۋ	dsez	سەت	dsa	مىنى	dcr2	ئىدەر
dsms	مەك	dsma	تىن	dskj	مەيىنشە	dsh	مۇ	dseo	لاس	drc3	سە	dcr2	مىز
dsms	باق	dsma	مىتىن	dskd	كەلى	dsh	دەق	dseo	س	drc3	سا	dcr2	ئىزدەر
dsms	بەك	Dsl	پە	dskd	قالى	dsh	مىش	dseo	س	drc1	بان	dcr2	ئىزدەر
dsms	باق	Dsl	ما	dskd	گەلى	dsh	شىلىك	dseo	لەس	drc1	پىز	dcr2	سەك
dsms	بەك	Dsl	مە	dskd	عالى	dsez	گىز	dseo	دەس	drc1	مىز	dcr2	مىسەك
dsms	عان	Dsl	با	dskb	پ	dsez	عىز	dseo	مىس	drc1	بىز	dcr2	سەككەر
dsms	كەن	dsl	پا	dskb	پ	dsez	كىز	dseo	تاس	drc1	مىز	dcr2	سەككەر
dsms	قان	dsl	بە	dskb	مىپ	dsez	قىز	dseo	تەس	drc1	مىز	dcr2	سىز
dsms	گەن	dskz	ا	dsh	ش	dsez	مىر	dseo	داس	drc1	مىن	dcr2	مىز
dsms	ەر	dskz	ە	dsh	دەي	dsez	مىر	dsee	مىل	drc1	مىز	dcr2	سىزدار
dsms	باس	dskz	ي	dsh	مىلى	dsez	ار	dsee	مىل	drc1	مىن	dcr2	سىزدار
dsms	ار	dskz	گەنشە	dsh	شى	dsez	دار	dsee	مىن	drc1	مىن	dcr2	ئىدار
dsms	ر	dskz	كەنشە	dsh	لى	dsez	دەر	dsee	مىندى	drc1	مىن		
dsms	ماس	dskz	عاشا	dsh	مىشى	dsez	مىر	dsed	نەپ	drc1	ك		
dsms	مەس	dskz	گەنشە	dsh	لىق	dsez	مىر	dsa	نى	drc1	م		
dsms	پەس	dskz	قانشا	dsh	لەك	dsez	تەت	dsa	مىنى	drc1	ق		
dsma	بىتىن	dskj	پەيىنشە	dsh	دەك	dsez	مىت	dsa	نى	drc1	پىن		

注：表中的 dsen 表示自身态附加成分，dseo 表示共同态附加成分，dsez 表示使动态附加成分，dsee 表示被动态附加成分，dskb 表示一般副动词附加成分，dskz 表示状态副动词附加成分，dskd 表示目的副动词附加成分，dskz 表示时限副动词附加成分，dskj 表示假使副动词附加成分，dsms 表示经历形动词附加成分，dsms 表示职业形动词附加成分，dskm 表示目的形动词附加成分，dsma 表示性质形动词附加成分，dsh 表示动名词附加成分，dsa 表示及物动词附加成分，drc1 表示第一人附加成分，drc2 表示第二人称附加成分，drc3 表示第三人附加成分。

4.3 动词短语定界规则的确定

哈萨克语基本动词短语的定界规则的正确性对识别结果有很大的影响，如果一个系统的定界规则比较完备，可以大大提高整个系统的准确率。定界规则应该有一定的实用性和固定性，在一个环境构成哈萨克语基本名词短语的定界规则，在另外一个环境中也适应。

本文将哈萨克语基本动词短语观察窗口确定为【-2，1】，即当前动词的前两个词与后一个词。为了提高动词短语识别的准确率，根据句子中的单词、词性和一些单词附加成分，总结出了动词短语的定界规则，用这些规则来构建本文规则库，以便于动词短语自动标注。本文对《新疆日报》语料中抽取 7 天的语料(规模为 5000 句的语料)进行了统计和分析，归纳出了以下能够构成哈萨克语基本动词短语的 16 条规

则，如表 2 所示：

表 2 哈萨克语基本动词短语搭配规则

序号	搭配规则	序号	搭配规则
1	n+v	9	n+conj+v
2	v+v	10	ono+v
3	adj+v	11	int+v
4	adv+v	12	num+m+v
5	v+adv	15	v+prep+v
6	v+v+v	14	v+prep
7	pron+v	15	v+conj+v
8	n+prep+v	16	pron+prep+v

4.4 动词短语歧义形式

对于 KzBaseVp 歧义，不能简单用规则搭配方法来消除，本文总结出以下 5 个 KzBaseVp 歧义格式。

(1) VD 格式 (v+adv)

例句 1: adv/تومەندىۋى v/قابىلدەۋى 是动词短语。

例句 1: adv/ەرەكشە v/قابىلدەۋى 是副词短语。

- (2) ND 格式 (n+adv, pron+adv)
 例句 1: adv/جاڭلاۋ n/كيسين 是动词短语。
 例句 1: adv/كەرمەت n/ناتىجەسى 是副词短语。
- (3) NPV 格式 (n+prep+v, pron+prep+v)
 例句 1: v/ۋىرەنۋ prep/تۇرالى n/نىتىماق 是动词短语。
 例句 1: v/دى prep/ئىلەن n/ئاسان 是名词短语。
- (4) VPV 格式 (v+prep+v)
 例句 1: v/كەتتى prep/دە v/كەلدى 是动词短语。
 例句 1: v/تۇسنىۋ prep/تۇرالى v/زەرتتەۋ 是副词短语。
- (5) VP 格式 (v+prep)
 例句 1: prep/بۇرىنداۋ v/سويلەۋدەن 是动词短语。
 例句 1: prep/جونىننە v/رەتتەۋ 是名词短语。

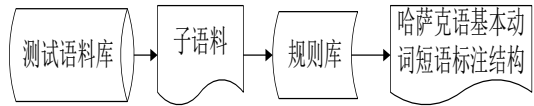


图 3 基于规则的动词短语识别方法系统

5 哈萨克语动词短语的自动识别

5.1 识别算法

输入：测试语料(test.xml)；

输出：第一个是标注动词短语识别的文件 result.xml；第二个是基本动词短语文件 resultVP.txt；

动词短语识别具体算法如下：

- (1) $i=1$;
- (2) while(! (test.xml))

① from right to left mach rule(rule data);

② if (find mach rule)

if (此规则不符合上下文排除规则)

置此短语左右边界并标注相应的结构类型；

③ $i=i+1$ (right move next word)

(3) Out put POS: result.xml and KzBaseVp:resultVP.txt file.

本算法中的测试语料是哈萨克语《新疆日报》语料中抽取一个月的语料，通过实验室目前使用的哈萨克语词性标注系统准备测试语料，然后通过人工方式消除未登录词和兼类词等问题的较理想的语料。

5.2 实验系统

根据以上所描述的识别算法，利用哈萨克语动词短语构成规则集，设计了基于规则的动词短语自动标注系统，系统是用 c#语言实现的，工作流程图如图 3 所示：

5.3 实验系统界面

第一步：根据哈萨克语基本动词短语的搭配规则，从语料中抽取哈萨克语基本动词短语，其系统界面图如 4 所示：



图 4 抽取哈萨克语基本动词短语界面

第二步：根据哈萨克语基本动词短语的搭配规则，对从语料中抽取的哈萨克语基本动词短语通过人工方式进行修正，选出最正确的哈萨克语基本动词搭配格式，其系统界面图如 5 所示：



图 5 修正已抽取的哈萨克语基本动词短语界面

通过上面的第一步或者第二步哈萨克语基本动词短语抽取方法来提取的基本动词短语，并保存到基本动词短语库里，其短语库如 6 图所示：

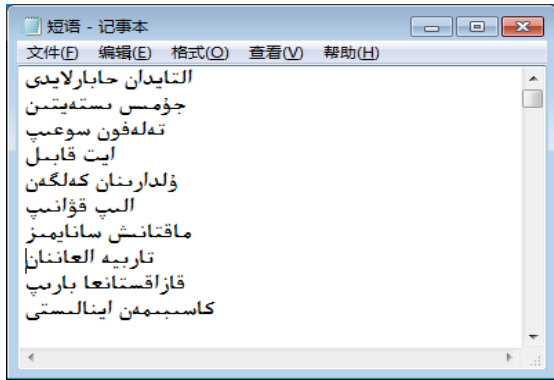


图6 已提取的动词短语

第三步：打开测试语料文本，再打开哈萨克语基本动词短语库，最后通过规则库中的信息对测试语料进行基本动词短语标注。其系统界面如图7所示：

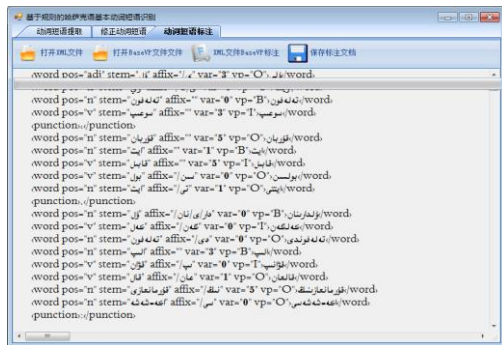


图7 基于规则的方法识别哈萨克语基本动词短语图

5.4 实验数据

本文采用以下几个指标：短语识别准确率，召回率，错误率和漏召率。

定义如下：

$$\text{召回率 (recall)} = a / (a+b) * 100\%$$

$$\text{正确率 (precision)} = a / (a+c) * 100\%$$

$$\text{漏召率 (leakage)} = b / (a+b) * 100\%$$

$$\text{错误率 (error)} = c / (a+c) * 100\%$$

$$\text{其中 } \text{recall} + \text{leakage} = 1;$$

$$\text{precision} + \text{error} = 1;$$

a 表示系统正确识别的短语数；

b 表示系统漏召的短语数；

c 表示系统错误识别的短语数目；

本系统从《新疆日报》已经标注的 3000 句短语作为训练语料，进行封闭测试，另外对 1000 句语料进行开放测试，测试结果如表 3 所示：

表3 基于规则的动词短语识别测试结果

测试类型	准确率 (%)	召回率 (%)	错误率 (%)	漏召率 (%)	F 值 (%)
封闭测试	91.27	89.93	8.73	10.07	90.60
开放测试	87.92	86.36	12.08	13.64	87.13

使用规则库的动词短语的各种结构类型，在开放测试的语料中准确率可以得到较好的短语标注结果，再通过人工方式修改定界错误，得到更高的准确率。

6 总结

本文以统计与分析哈萨克语基本动词短语形式，总结出哈萨克语基本动词短语的定界与对策短语搭配类型说明，提取基于规则的方法，实现哈萨克语基本动词短语识别。

哈语自由短语比较多而灵活，哈语语言特征在自由短语中表现突出，仅使用词性搭配识别动词短语是不能解决问题，哈萨克语动词短语结构相当复杂，有可能存在没有考虑到的规则产生错误识别问题。

研究哈萨克语基本动词短语的难点之一是歧义排歧问题。产生 KzBaseVp 歧义的原因主要有单词词性歧义问题，短语界限不易确定，词性搭配序列相同等，这些问题有待进一步研究。

本系统对实验室现有的“新疆日报”语料中抽取 30 天语料（规模为 20MB）中进行哈萨克语基本动词短语（KzBaseVP）识别。从实验结果表明规则方法对哈萨克语基本动词短语识别是可行的，在封闭和开发测试条件下可以得到令人较满意的效果。

参考文献

- [1] 李荣，曹建芳著，汉语名词短语和动词短语的自动识别方法研究[M]. 北京：兵器工业出版社，2008:5-8
- [2] Steven Abney. Parsing by chunks[M]. Dordrecht: Kluwer Academic Publishers,1991:257-278
- [3] Church K.A stochastic parts program and noun phrase parser for unrestricted text[J]. In Proceedings of the Second Conference on Applied Natural Language Processing. Texas, USA. 1988,19(8):136-143.

- [4] 赵军,黄昌宁. 结合句法组成版识别汉语基本名词短语的概率模型[J]. 计算机研究与发展. 1999, 36(11):1384-1390.
- [5] 赵军,黄昌宁. 汉语基本名词短语结构分析模型[J]. 计算机学报. 1999, 22(2):141-146.
- [6] 周强,孙茂松,黄昌宁. 汉语句子的组块分析体系 [J]. 计算机学报 .1999, 22(11) :1158—1165.
- [7] 周雅倩,郭以昆,黄萱菁等. 基于最大熵方法的中英文基本名词短语识别[J]. 计算机研究与发展. 2003, 40(3) :440-446.
- [8] 李素建,刘群,杨志峰. 基于最大熵模型的组块分析 [J]. 计算机学报 .2003. 26(12) : 1722-1727.
- [9] 白妙倩,郑家恒. 基于最大熵方法进行动词搭配的自动标注 [J]. 计算机工程与应用, 2009, 45(3) :135-137
- [10] 霍亚格,黄广君. 基于最大熵的汉语短语结构识别方法[J]. 计算机工程. 2011, 37(16) : 206-210.
- [11] 达胡白乙拉. 蒙古语基本动词短语自动识别研究[M]. 内蒙古大学博士论, 2005:12-13
- [12] 孙瑞娜,古丽拉·阿东别克. 基于规则的哈萨克语基本名词短语识别研究 [J]. 计算机应用研究, 2010, 27(12): 4511-4513.
- [13] 孙瑞娜,古丽拉·阿东别克. 哈萨克语基本名词短语自动识别研究与实现 [J]. 中文信息学报, 2010, 24(6) :114-119.
- [14] 张定京. 现代哈萨克语实用语法[M]. 北京:中央民族大学出版社, 2004. 15-198.
- [15] 耿世民,李增祥. 哈萨克语简志[M]. 北京:民族出版, 1985. 54-290.
- [16] 耿世民. 现代哈萨克语语法[M]. 北京:中央民族学院出版社, 1989