

文章编号: 1003-0077(2007)05-0118-08

语言学与统计方法结合建立汉语动词 SCF 类型集

冀铁亮,孙薇薇,穗志方

(北京大学 计算语言学研究所,北京 100871)

摘要: 动词子语类框架(Subcategorization Frame 以下简称 SCF)在句法分析、语义角色标注等方面的研究中具有不可或缺的重要作用。在子语类框架信息的获取过程中,首先要建立标准完备的子语类框架类型集。目前英语研究已经建立了获得普遍共识的子语类框架类型集。而汉语方面还没有标准的动词子语类框架类型集。本文提出一种语言学知识与统计方法相结合的汉语动词子语类框架类型集的半自动获取方案。初步建立起既符合统计结果又基本符合语言学理论的汉语动词子语类框架类型集。实验证明,加入语言学理论子语类框架类型集降低了对语料的依赖程度,比完全由分析语料产生的类型集更完备。

关键词: 计算机应用;中文信息处理;动词子语类框架;类型集;语言学与统计方法结合

中图分类号: TP391

文献标识码: A

The Acquisition of Chinese Verb's Subcategorization Frame Types Based on Linguistic Theory and Statistical Algorithm

JI Tie-liang, SUN Wei-wei, SUI Zhi-fang

(Institute of Computational Linguistics, Peking University, Beijing 100871, China)

Abstract: Subcategorization of verbs is an essential issue and plays an important role in syntactic parsing, semantic roles labeling and etc. A sufficient subcategorization frame type set is critical for subcategorization acquisition. By now, a set of subcategorization frame types has come to an agreement in English, while no standard subcategorization frame type set for Chinese verbs has been achieved. In this paper we apply a semi-supervise method for subcategorization frame type acquisition with linguistic theory and statistical algorithm. Firstly we create a set of seeds of subcategorization patterns according to linguistics theory. And then a semi-supervise machine learning method is applied to analyze the corpus for extending the seeds. Contrasted with a corpus based subcategorization frame type acquisition method, our method gains better precision and coverage.

Key words: computer application; Chinese information processing; verb's subcategorization frame; lexicon; the hybrid of linguistic theory and statistical algorithm

1 引言

子语类框架(Subcategorization Frame)也翻译作“子类化”、“子范畴化”、“子范畴”,是根据动词可带句法成分对动词进行分类的一种方法。它描述的是动词与其搭配的成分之间的搭配模式及动词和其搭配成分之间的句法语义关系。例如:

“给”的子语类框架为:给 NP NP(给你一本书)。

动词子语类框架的信息在语言学上有重要的意义,它可以解决绝大部分词语的论元和附属语区分问题。在概率句法分析应用中,子语类框架信息可以大大提高句法分析的准确率,在语义消歧和语义角色标注中子语类框架信息也可提供重要的信息^[1~3]。

在获取具体动词的子语类框架信息之前,建立

收稿日期: 2007-04-05 定稿日期: 2007-06-29

基金项目: 国家自然科学基金资助项目(60503071)

作者简介: 冀铁亮(1982—),男,硕士生,研究方向为自然语言处理;孙薇薇(1983—),女,硕士生,研究方向为自然语言处理;穗志方(1970—),女,副教授,研究方向为自然语言处理。

完整准确的动词子语类框架类型集是至关重要的。动词子语类框架类型集要面向广泛应用的各种自然语言处理系统,所以既要能够解释一般的语言学现象,又要尽可能反映语言的深层本质。

2 国内外研究现状

在国外,有关 SCF 自动获取的研究最早始于 1991 年 Brent 对英语的研究^[4]。目前相关研究在德语、捷克语、西班牙语、葡萄牙语等语言中都有所展开^[5-7]。Anna Korhonen 提出了完整的获取英语子语类框架信息的方法^[8]。而大多数研究所使用的动词子语类框架词典是已经在英语方面达成普遍共识的 Levin 动词子语类框架词典^[9]。

在国内现代汉语研究中,目前尚没有学者就汉语动词 SCF 展开深入的传统语言学研究。但有三个方面的研究——配价理论、题元理论和格语法——与汉语动词 SCF 研究密不可分。这三个方面的研究都涉及对动词的搭配成分的研究,对汉语动词 SCF 的研究很有参考意义。尤其是配价语法,是近些年现代汉语研究的一个热点,也取得了很多很有意义的成果^[10,11]。计算语言学方面,哈尔滨工业大学的韩习武、赵铁军老师正在展开汉语动词 SCF 自动获取方面的研究,目前取得了一些很有意义的进展^[12]。而北京大学计算语言所开发的《现代汉语语法信息词典》收集了 8 万现代汉语常用词的丰富的语法信息。其中关于动词的相关句法信息,可以作为汉语动词 SCF 的原型。

到目前为止,面向自然语言处理的汉语 SCF 类型集,主要是在语料分析的基础上建立的。主要的提取方法都以从语料提取为依据,而没有重视语言学的理论依据。目前,SCF 类型集的设置也没有达成普遍共识。

3 语言学理论与统计学方法相结合建立汉语动词子语类框架类型集

本文提出一种语言学理论与统计学方法相结合获取完整的 SCF 类型集的方法。首先利用语言学方法,建立 SCF 类型的种子集合,然后利用大规模标注语料对建成的种子集进行扩充。利用错误驱动统计方法循环迭代扩充种子集,最终得到的动词子语类框架类型集,是在满足实际语料的基础上根据语言学的规则产生的。

4 利用语言学方法,建立 SCF 类型的种子集

4.1 利用纯概率方法获取 SCF 存在的问题

从语言学角度来看,SCF 位于句法结构和句法语义结构的交合处,是其他句法语义信息的基础。另外,SCF 包括动词的论元而不包括附属语。论元与附属语的区别主要在句法语义层面,例如几乎所有的动作都会涉及动作发生的时间和地点,因此表示动作发生时间和地点的句法成分出现的概率很高,但它们通常不是论元而是附属语。基于以上考虑,我们对于设置 SCF 类型集更加倾向于依靠语言学理论设置,而非机器自动从句法分析结果中获取。

4.2 论元和附属语(包括外围论元)的界定

论旨角色的概念最初由 Gruber 和 Fillmore^[13]提出来的。述语有一些固有的角色,表示述语所涉及的主、客体或动作、行为状态、处所等,这些角色称为论旨角色。论元指带有论旨角色的名词短语。而附属语是和动词联系不是很紧密的短语,不属于动词的搭配成分。如:“3 时 40 分,3 辆中巴车静静地开出医用电子仪器厂的大门。”一句中,“3 辆中巴车”是“开”的论元,而“3 时 40 分”是附属语。

袁毓林老师提出了 17 种常见论元角色^[14]。论元的论旨角色和说话人对动词所激活的语义场景有很大的关系。在手工标注 SCF 的过程中,我们发现环境论元的语法格式相对而言更加固定。考虑到这两点,本次 SCF 设置包括了核心论元和外围论元中的凭借论元及超级论元 13 种,而不考虑环境论元。

在英语中,动词的子类框架不包含关于主语的信息,在英语的研究者看来,动词不单独选择主语,需要主语不是动词本身的属性,而是句法结构的要求。但在汉语之中,句法结构并不限定主语,一个句子不一定要有主语,句子的主语也不一定只有一个。从这个角度看来,主语的选择和述语密切相关。因此我们的 SCF 带有主语信息,这样做的另一个好处是和目前的配价语法保持一致。

4.3 SCF 类型集的设置原则

在设置 SCF 集合过程中,参考文献[15]一文中的原则,我们以下述条件作为判断具体实例中成分是否为论元的测试条件:

原则 1: 候选辨别成分应在句中做主宾语,如:

(1) 钱其琛说,中国和南非正式建交是给两国人民最好的新年礼物。

主语“钱其琛”为施事,是动词“说”的论元,句(1)的 scf 为 np,!,dj。

(2) 水利改革使广大农民对待兴修水利同对待家庭责任田一样,看作是自家的分内事。

主语“水利改革”为致事,是动词“使”的论元,句(2)的 scf 为 np,!,np。

原则 2: 候选辨别成分通过句式变换后,可以在变换之后的句中做主宾语。而变换过程之中,变换前后两个句子中所包含的各个实词性成分之间的语义结构关系一致。如:

(3) 这些舞弊行为长期被野村公司掩盖。

该句可以变换为“野村公司长期掩盖这些舞弊行为”,变换前后,“野村公司”、“掩盖”和“舞弊行为”的语义结构关系一致,故(3)句中“野村公司”为动词“掩盖”的论元,句(3)的 scf 为 np,pp,!。

(4) 考虑到与原有房改政策相衔接。

该句中可以变换为“考虑与原有房改政策相衔接”,变换前后,“考虑”与“与原有房改政策相衔接”语义关系一致,故“与原有房改政策相衔接”为动词“考虑”的论元,句(4)的 scf 为!,pp。

原则 3: 由于场所、源点、终点和范围从意义上较为容易判断,所以当满足上述两个条件的候选辨别成分表现为对动词场所、源点、终点和范围的描述,尤其是以介词“在”引导的体词性成分,不视为论元成分。

原则 4: 候选辨别成分动态语义满足文献[15]中对十三种论元中一种论元的动态语义描述。

原则 5: 候选辨别成分静态语义满足文献[15]中对十三种论元中一种论元的静态语义描述。

以上 SCF 类型集设置原则中,前三种条件是概括性条件,可以简化论元辨别时的工作量,而后两种参考了文献[15]理论的条件则更加细致,遇到模棱两可的问题时,可以做具体参考。

而在 SCF 设置过程中,还遇到了一些典型情况,针对这些具体问题做如下约定:

原则 6: 递系结构的处理

递系结构的特点是,前一个谓词性结构一定是述宾词组,这个述宾词组的宾语,跟后面的谓词性成分在语义上有某种联系。如:

(5) 请你/去一下

(6) 选他/当我们的代表

那么,递系结构中后面的谓词性成分是否为前

一个动词的论元呢?进行如下约定:如果去掉前一个述宾词组的宾语仍然可以说,则将后面的成分视为述宾词组中动词的论元;否则,不将其看成是论元。这样的约定下,“请去一下”可以说,所以(5)句“请”的 SCF 为!,np,vp;而“选当我们的代表”不能说,故(6)句“选”的 SCF 为!,np。

原则 7: 主谓谓语句的处理

主谓谓语句中,谓语由主谓结构充当。如:

(7) 下午我们开会

(8) 我们班一半是南方人

(9) 再大的困难我们也不怕

为叙述方便,将整个主谓结构的主语记为 S,谓语部分的主语记为 s,谓语部分的谓语动词记为 v。这时,S 是否为 v 的论元?进行如下约定:如果 S 不是 v 的十三种论元中的任何一种,则不视为 v 的论元;如果 S 和 s 之间有领属或整体部分关系,则 S 不视为 v 的论元;如果 s 是 S 的指代,将 S 和 s 同时视为论元;除此之外,若 S 为 v 的十三种论元之一,视其为论元。这样的约定下:(7)句中“下午”是附属语,不划为“开”的论元;(8)句中“我们班”和“一半”是领属关系,“我们班”不划为“是”论元;(9)句中“再大的困难”是“怕”的受事,故将“再大的困难”划为“怕”的论元。

4.4 SCF 种子集

从直观上来看,常见的语言现象覆盖率相当大。如 Zipf 定律揭示了这样的现象:前 100 个高频词覆盖了文本的 51%。不仅语言现象如此,自然界中的很多事物都是如此,如渐进均衡性(AEP)所揭示的全集中的典型集的概率近似为 1。根据这一观察,我们手工提取了“哈工大信息检索研究室汉语依存树库”的前 1000 句的 SCF,利用语言学理论规范扩充总共获得 48 个 SCF,高频的 SCF 应在这 48 个 SCF 中,以此作为 SCF 的种子集。

5 错误驱动的迭代算法与种子集的扩充

5.1 依存分析树库

本文采用依存分析树库作为分析语料,与其他语料相比,依存关系分析树可以提供词汇之间的上下位依存关系,词汇之间的附属关系已经准确的标出,在提取子类框架信息过程中可以得到比短语块(Chunk)分析更详细的语法信息,保证子类框架

类型集的分析过程有更充足的判断依据, 比对语料进行块分析的结果准确率更高, 而且在进行人工辅助分析的时候, 更利于分辨被错误分析的句子成分。在分析 SCF 类型过程中, 对于不符合规则的 SCF 类型, 通过词性和依赖关系特征的学习, 可以产生高效的 SCF 类型的过滤规则。

5.2 从依存分析树到子语类框架类型的提取

在依存分析树中找到动词中心语, 以及被动词中心语直接支配的句子成分, 获取备选的 SCF 模式。由于标注短语类型的准确性要远远低于标注词性的准确性, 在实验过程中为了获取高可靠性的 SCF 信息, 在本文中采取了一种多对一的方式来实现提取匹配方式。在提取过程中由于 SCF 类型集中的 SCF 类型都是由短语类型组成的, 本文将短语类型扩展为各种词性之间的组合, 而在对依存分析树库中句子的分析中, 只需要对准确率高的词性标注进行模式提取, 而不需要对正确率相对低而且在语言学上存在争议的短语级别上进行模式提取, 这样只要提取与 SCF 类型具有相同的词性组合的类型之一, 就可以将得到的 SCF 模式对应到 SCF 类型种子集中去。

5.3 错误驱动的迭代算法

首先利用种子集对分组语料进行分析, 对于没有在种子集中匹配上的 SCF 模式类型, 可能是潜在的 SCF 类型, 也可能是不符合语言学规则的, 对于这些 SCF 模型, 人工对其进行分析, 如果符合语言学标准那么就可以得到新的 SCF 类型, 而将同类的词性组合上升为短语类型组合加入到 SCF 类型种子集中。如果不符合就将产生 SCF 类型的句子的词性组合和依赖关系作为负例特征进行学习, 产生淘汰规则。

本文采用机器学习获取淘汰规则。使用机器学习的方法从训练文本中自动学习也可以获得规则的集合, 以节省手工劳动。学习算法的描述如下:

Step1. 根据 SCF 类型种子集对提取的模式进行匹配得到种子集中不存在的类型集合 set;

Step2. 用初始规则 rules 对 set 进行淘汰选择, 得到一个结果集合 result;

Step3. 将结果集合 result 和 SCF 类型种子集作比较, 可以得到未匹配规则类型集合 unmatched;

Step4. 计算出本次规则匹配中评价函数 F 的值;

If 之前没有 F 值, 转 step5;

Else 则将此次新的 F 值与已有值进行比较, 转 step9;

Step5. 将集合 unmatched 分析后, 分别按照规则模板的格式对每一部分作统计, 也就是相应的各部分的词频统计和相应的词性标注记录, 结果存入一定数据结构 word 和 POS;

Step6. 选取候选词性组合加入到规则模板的相应部分, 构成一条新的规则 newrule, 放到候选淘汰规则集;

Step7. 将 F 值最高的规则加入到规则集, 规则达到一定总数, end;

Step8. 如果候选规则集不空, 依次从候选规则中取出一条规则, 使用这条新规则在不存在的类型集合 set 进行淘汰, 得到一个新的结果关系集合 result, 转 step3;

Step9. 选择新规则;

If 新的 F 值大于之前的 F 值, 则记录这条规则, 转 step7;

Else 不选择这条规则, 将该规则从候选规则中删除;

转到 step8, 如果候选规则集空, 转 step7。

整个学习过程结束, 得到一个淘汰规则的最终集合, 这就是经过简单的机器学习得到的结果规则集。在以上的学习过程中, 每学习一条规则, 实际上都有根据评价函数来确定改错效果的过程, 这些函数对每一条候选的规则进行评价, 从中挑出“好”的规则, 加入到最终的规则集中。这样每添加一条规则, 就计算相应的准确率、召回率和 F 综合评价, 然后根据具体数值和评价标准来确定是否保留当前规则。

6 实验

6.1 基础实验

本文使用的是由哈工大信息检索研究室提供的汉语依存树库作为分析语料, 总共 10 000 句(其内容为《人民日报》1998 年 1 月语料), 其中的语料都是经过依存分析和人工校对的树库。将语料分成 10 组, 每组 1 000 句, 分别在实验中通过迭代分析获取子语类框架类型, 最终获得的 SCF 类型集见附录。

6.2 对比试验

对比实验中,我们将完全根据语料分析提取动词子语类框架集的方法(以下简称方法 A)与本论文提出的方法(以下简称方法 B)进行比较,分别依据方法 A 和方法 B 获得框架集。从封闭语料(《人民日报》)和开放语料(《微型计算机基础》和《医学百科全书》)三个不同领域的语料中随机选取 100 句进行分析,对两种方法提取的子语类框架的准确率和查全率进行对比试验。

方法 A——完全根据语料分析提取动词子语类框架集的方法,是利用与本论文相同的训练语料《人民日报》进行分析。查找动词中心语,然后利用词性的依存关系,提取动词中心语相关联的论元,提取词性组合的 pattern,然后将 pattern 根据词性组合到短语类型的对应关系产生子语类框架类型,对产生的子语类框架根据出现频度(大于 5)和语法规则进行过滤,最终提取出含有 63 个子语类框架类型的类型集。该类型集与方法 B 得到的类型集(见附录)的区别与联系如表 1 所示。

表 1 结合语言学方法与完全基于语料的方法产生的类型集的比较

B-A							
!,mp	ap,!,vp	mp,!,mp	np,!,np,mp	np,pp,!	np,pp,pp,!	pp,np,!,np	vp,!,np,np
!,np,mp	ap,np,!,ap	mp,np,!	np,!,np,np	np,pp,!,mp	np,vp,vp,!	pp,np,!,vp	vp,!,pp
!,np,np	ap,np,!,np	mp,pp,!	np,!,pp	np,pp,!,np	pp,!	pp,pp,!	vp,pp,!
!,pp	mp,!	np,!,ap,ap	np,mp,!,np	np,pp,!,np,vp	pp,!,mp	pp,vp,!	vp,pp,!,np
ap,!	mp,!,ap	np,!,mp,mp	np,np,!,mp	np,pp,!,vp	pp,!,np	vp,!,mp	vp,pp,!,vp

两条因素对差别的产生有很大的作用:一、SCF 既表现出句法的性质也表现出语义的性质,特别地 SCF 和句法并不一一对应;二、如前文所分析,附属语的概率并不显性低于论元。因此这种方法不能很好地解决 SCF 类型集设置问题。

对比实验分别从 3 个不同领域的语料中各抽取 100 句子作为测试集,利用两个不同的子语类框架类型集分别从中提取所有动词的子语类框架信息,结果见表 2。

表 2 对比实验结果

语料	《人民日报》		《微型计算机》		《医学百科全书》	
	准确率	查全率	准确率	查全率	准确率	查全率
方法 A	92 %	100 %	66.7 %	92 %	76.9 %	95 %
方法 B	97 %	100 %	95 %	100 %	97 %	100 %

6.3 结果分析

根据分析结果可以看出,利用相同的语料,用两种方法训练得出的子语类框架类型信息中,分析封闭语料时结果正确率差距不大。但是由于方法 B 加入了语言学的方法,在对比试验中正确率有部分提高。而当分析其他领域的语料时,由于完全由语料产生的子语类框架集对语料的依赖性比较大,面对陌生领域的语料时,由于不同领域的语言表达方式的不同,由一个领域的语料训练得到的子语类框架类型,不一定能够涵盖其他领域的子类框架类型的用法,方法 A 正确率下降较多。而方法 B 产生的子语类框架类型集由于加入了语言学的信息,而在

每次扩充时不但根据语料的统计信息,还加入了语言学的特征,对类型进行了泛化调整,不再完全依赖于语料,得到的是更加具有普遍性的类型集。因此在分析不同领域的语料时,依然可以保证较高的正确率。而从查全率的结果可以看出,本文中获取的子语类框架类型集已经可以涵盖绝大多数子语类框架类型。

7 结论及展望

本文通过语言学理论与对语料的统计分析相结合的方法获取了汉语动词子语类框架的类型集,并与单纯基于对语料分析统计的方法进行了对比实

验。实验结果表明,加入了语言学规则信息的类型集,降低了对语料领域的依赖程度,在对不同领域的语料的分析都具有比较理想的效果。初步建立起了既符合统计结果又基本符合语言学本质的汉语动词子语类框架类型集。在此基础上我们将进一步完善和优化,争取形成相对全面和稳定的汉语动词子语类框架类型集。

建立相对完整的汉语动词子语类框架类型集,只是完成了进行动词子语类框架信息提取的一个必要条件。利用类型集提供的子语类框架类型信息,我们将完成常用汉语动词的子语类框架概率信息的提取,在此基础上建立汉语动词子语类框架的概率信息词典。希望能够为概率句法分析以及语义角色标注等研究提供基础资源。

参考文献:

- [1] Briscoe, E. and J. Carroll. Automatic extraction of subcategorization from corpora [A]. In: Proceedings of the 5th ACL Conference on Applied Natural Language Processing [C]. Washington, DC: 1997. 356-363.
- [2] Collins, M. J. Three generative, lexicalised models for statistical parsing [A]. In: Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL-97) [C]. 1997. 16-23.
- [3] Carroll, J., G. Minnen and E. Briscoe. Can subcategorisation probabilities help a statistical parser? [A]. In: Proceedings of the 6th ACL/ SIGDAT Workshop on Very Large Corpora [C]. Montreal, Canada: 1998. 118-126.
- [4] Michael Brent . Automatic acquisition of subcategorization frames from untagged text [A]. In: Proceedings of the 29th Meeting of the ACL [C]. Berkeley, CA: 1991. 209-214.
- [5] Walde SS, Brew C. Inducing German semantic verb classes from purely syntactic subcategorization information[A]. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics [C]. 2002. 223-230.
- [6] Sarkar A, Zeman D. Automatic Extraction of Subcategorization Frames for Czech [A]. In: Proceedings of the 19th International Conference on Computational Linguistics [C]. Aarbrucken, Germany: 2000.
- [7] Eva Esteve Ferrer. Towards a Semantic Classification of Spanish Verbs Based on Subcategorisation Information [A]. In: Proceedings of the ACL 2004 Workshop on Text Summarization[C]. Spain: 2004.
- [8] Anna Korhonen. Subcategorization Acquisition [D]. PhD thesis published as Technical Report UCAM-CL-TR-530. Computer Laboratory, University of Cambridge. 2002.
- [9] Levin, Beth. English Verb Classes and Alternation [M]. Chicago University Press, Chicago: 1993.
- [10] 袁毓林. 汉语动词的配价研究[M]. 江西教育出版社. 1998.
- [11] 郑定欧. 现代汉语配价语法研究[M]. 北京: 北京大学出版社. 1995.
- [12] 韩习武, 赵铁军. 基于子类的汉语动词 SCF 获取回退[J]. 计算机工程与应用, 2005.
- [13] Fillmore, C. Topics in Lexical Semantics [J]. Current Issues in Linguistic Theory, 1977, 76-138.
- [14] 袁毓林. 论元角色的层级关系和语义特征[M]. 世界汉语教学. 2002.
- [15] 袁毓林. 一套汉语动词的论元角色的语法指标[M]. 世界汉语教学. 2003.

附录 子语类框架类型集以及实例

标记短语类型说明

SCF 的框架集采用了如下短语集: {ap、mp、np、pp、vp}。其中: ap 表示形容词性短语, mp 表示数量词性短语, np 为名词性短语, pp 为介词性短语, vp 为动词性短语或主谓结构短语。而中心动词由 ! 表示。

SCF 类型	实 例
!	怎么[抓]!也抓不住
!,ap	[希望]![好起来]ap
!,mp	[新增]![3.12元]mp
!,np	[发展]![两国关系]np
!,np,ap	[嫌]![他]np[年轻]ap
!,np,mp	[生产]![原油]np[1580万吨]mp

续表

SCF 类型	实 例
!,np ,np	[给]![我]np[那本书]np
!,np ,vp	为此[呼吁]![有关部门]np[加强拍卖市场监管]vp
!,pp	[适用]![于中华人民共和国境内已建或在建的电力设施]pp
!,vp	正确[对待]![个人进退留转]vp
!,vp ,vp	[怕]![不听信王书记的忠告]vp[会酿成恶果]vp
ap ,!	[好坏]ap[不易判断]!
ap ,!,np	[警徽闪闪]ap[耀]![国道]np
ap ,!,vp	[形势好]ap也[包括]![各界对“一国两制”方针政策的理解加深]vp
ap ,np ,!,ap	[她对于改革的宣传很突出]ap,[这]np又[说明]![办报人丝毫不迟钝、不守旧]ap
ap ,np ,!,np	[忠诚]ap,[这]np[是]![“绍丝印”在全厂各级干部中倡导的工作精神]np
mp ,!	[10多年]mp[过去]!了
mp ,!,ap	[全镇被他帮扶过的11家]mp都[脱]!了[贫]ap
mp ,!,mp	[改革开放的20年]mp,[是]![我国乡镇企业异军突起的20年]mp
mp ,!,np	[80%]mp[为]![库尔德人]np
mp ,!,np ,vp	[2200多万澳元]mp[帮助]![土著居民]np[预防传染病]vp
mp ,!,vp	[二]mp[是]![硬性规定指标]vp
mp ,np ,!	[这一点]mp,[武汉市的私营企业]np[体会]!特别深
mp ,pp ,!	[这一幕]mp,[被前来检查施工质量的营长]pp[看到]!
np ,!	[两国贸易]np[发展]!迅速
np ,!,ap	[许多企业及投资机构]np[认为]![已经过时]ap
np ,!,ap ,ap	[她]np[怕]![太红]ap[过于鲜艳]ap
np ,!,ap ,vp	[她]np[怕]![太黄]ap[会招来虫子]vp
np ,!,mp	[单面容量]np[为]![15千兆]mp
np ,!,mp ,mp	[近代重要报刊系列]np已经[完成]![《新青年》、《国闻周报》两种]mp[8万页]mp
np ,!,mp ,vp	[理由]np[有]![二]mp:[一是说我们打碎7块地板砖,还打碎一块玻璃]vp
np ,!,np	[各个部门]np想方设法[筹集]![资金]np
np ,!,np ,ap	[江书记]np[问]![张秀财]np[好不好]ap?
np ,!,np ,mp	[集团公司]np就[支付]![当地农民]np[1.8亿元]mp
np ,!,np ,np	[他们]np,[教会]![病人]np[自我综合管理的办法]np
np ,!,np ,vp	[这一经历]np,[引发]![泽州人思考一个问题]vp
np ,!,pp	[他的细心]np正[反映]![在这些小事上]pp
np ,!,vp	[业内人士]np[认为]!,,[需求剧增是主要原因]vp
np ,!,vp ,vp	[王书记]np[担心]![需求增加]vp[会导致工厂断货]vp
np ,mp ,!,np	[江苏队]np[预赛的最后两轮]mp[对]![弱旅浙江队和四川队]np
np ,np ,!	[“海马”牌电热水瓶]np[深圳市海马电器有限公司]np[生产]!
np ,np ,!,mp	这把刀我切了一棵白菜,[那把刀]np[我]np[切]!了[两棵]np

续表

SCF 类型	实 例
np , np , ! , np	[这本书]np[我]np[借]![他]
np , np , ! , vp	[那把刀]np[我]np[用来]![切菜]vp
np , pp , !	[邓小平同志]np[以“务实”]pp[著称]!
np , pp , ! , mp	[瑞典队]np 在全场比赛还剩 2 分钟时 , 方[由替补上场的 14 号阿尔博格]pp[扳回]vp [1 分]mp
np , pp , ! , np	[你]np[给小孩]pp[洗洗]![手]np
np , pp , ! , np , vp	[程维高]np 在会上郑重地[给张家口灾区干部]pp[出]!了[几道自强不息、恢复重建的题目]np:[一是不仅要恢复生产,更要发展生产]vp
np , pp , ! , vp	[中国佛教协会负责人]np[就西藏流亡印度的活佛馈赠释迦佛牙事]pp[答]![记者问]vp
np , pp , pp , !	[各级党委和政治机关]np[把引导部队学习彭楚政同志的先进事迹]pp,[同学雷锋、学英模和学习部队中的先进典型]pp[结合]!起来
np , pp , pp , ! , np	[安南]np[同伊拉克领导人]pp[就和平解决伊拉克武器核查危机问题]pp[达成]![协议]np
pp , !	现[将这部分普查数据的快速汇总结果]pp[公布]!
pp , ! , mp	[用这个喷壶]pp[浇]![两盆]mp
pp , ! , np	[为灾民]pp[送]!上[棉大衣]np
pp , np , ! , np	[就美国对华最惠国待遇问题]pp[外交部发言人]np[回答]![记者]np
pp , np , ! , vp	[就西藏流亡印度的活佛馈赠释迦佛牙事]pp[中国佛教协会负责人]np[答]![记者问]vp
pp , pp , !	[把臭水浜的改造同解决仓库、厂房]pp[和民宅的消防用水]pp 相[结合]!
pp , vp , !	[对这些看似荒唐的通知]pp,[没有人]vp[抗议]!
vp , !	[中国近代史第一次武装反抗西方鸦片商]vp[发生]!在一八八三年
vp , ! , mp	[降雨]vp 还将[持续]![数日]mp
vp , ! , np	[恢复高考]vp,[是]![我们人生道路上的盛大节日]np
vp , ! , np , np	[治瘫]vp[给]![病人]np[站起的希望]np
vp , ! , np , vp	[这些国家的货币汇率和股票价格犹如决堤的洪水狂泻而下]vp,[致使]![整个经济]np[陷入恶性循环状态]vp
vp , ! , pp	[破除迷信]vp 都应[建立]![在科学的基础之上]pp
vp , ! , vp	[奔驰、宝马等一系列豪华车降价]vp,[反映了]![目前豪华车市场暗潮涌动]vp
vp , np , !	[希捷硬盘提供五年质保]vp[用户]np[反映]!良好
vp , np , ! , np	[解决这个矛盾]vp,[两国]np[还需要]![耐心]np
vp , np , ! , vp	[美伊缓和关系]vp[势]np[在]![必行]vp
vp , pp , !	[耐心劝导]vp[和依法强行拆除]pp 相[结合]!
vp , pp , ! , np	[股份制改造]vp[给“中大”]pp[带来]![生机]np
vp , pp , ! , vp	[擅自转让探矿权、采矿权]vp,[由登记管理机关]pp[责令]![改正]vp