

HNC 理论下对“得”的理解和处理

熊亮

(中国科学院声学研究所,北京 100080)

E-mail: bright84@sina.com

摘要: 本文从语句理解处理的角度出发,运用 HNC 理论的语言表述模式,对“得”字的各种意义和用法进行了探讨,考察了各义项在真实语料中的分布情况,并提出了对“得”的处理策略。

关键词: 得;特征语义块;句类;HNC

引言

在现代汉语中“得”是一个出现频率较高的字。多年来,关于句子中“得”的研究,主要仅限于“得”引导述补短语这一最常见的用法,缺乏对“得”进行全面系统的研究分析。而且研究多是阐述语言现象,很少进入语义层面,这就不能适用于汉语信息处理。本文正是基于这两点,运用 HNC 理论对“得”的语义层面进行系统的研究,以期服务于计算机理解自然语言的需要。

HNC (Hierarchical Network of Concepts) 是概念层次网络理论的简称,是面向自然语言理解的理论体系,以概念化、层次化、网络化的语义表达为基础。语义块是句子的语义构成单位,句类可以简单的看成句子的语义类别。HNC 对语句的理解处理是以句类为指导的,对特征语义块的感知和假设是切入点。特征语义块(记作 E) 一般是一个句子最关键的语义构成单位,它或者是述语动词,或者是结构体,可以复合构成。特征语义块的核心部分(记作 E_k) 的前后都可以有说明部分, HNC 把前说明部分叫做上装,记作 q_v ; 后说明部分叫做下装,记作 h_v 。装和下装都有明确的概念符号表示,可以为 E 块感知提供简明的激活信息。特征语义块的构成可以表示为: $E = \text{上装} + E_k + \text{下装}$ 。“得”和特征语义块关系密切。

“得”除了作为 E 块下装这一最常见的用法外,又能作为 E 块上装,还能单独充当 E 块核心动词,以及作为一种重要的构词成分组合许多的复合动词。因此,包含“得”的句类很广泛。笔者从《人民日报》2000 年版的语料中随机抽取了 1000 个含“得”的句子,先剔除掉“得”固定到词组中的情形,然后对余下的句子进行了分类,将“得”作为单字的用法归结为特征语义块下装、特征语义块核心和特征语义块上装三种情况。下面我们就对“得”的这三种情况分别加以研究。

1 特征语义块下装 h_v

h_v 表示特征语义块的下装,是特征语义块中动词的一种后缀成分。现代汉语中典型的 h_v ,如:着,了,过……这是“得”最高频度的用法。此时对应于 HNC 概念语义网络中的语言逻辑概念。作为下装的“得”主要有 3 种用法:

1.1 表示动作行为到达的程度

用 HNC 符号表示为 $l42j60$ 。“得”后面所接的成分是另一种下装 hE ,说明 E 块基本特性。这些用

去主要用于口语,“得”后面所接的多是表示“度”的副词,如“很,非常”一类的概念。

(1)“得”的这种用法经常以两主块句形式出现。当形成两主块句的时候一般都和状态句密切相关,形成的句类有一般状态句SJ、反应-状态混合句X20S*11J等,分别对应于特征语义块是一般动词和反应动词的情形。例如:

他的背|也驼得过早。 S=SB+S
这孩子||饿得厉害。 X20S3*11J=X2B+X20S

当句子中不含E块时,例如简明状态句S041J。我们可以把SC看作“伪E块”,“得”后面的成份依然是修饰“得”前面的词。被修饰词是SC核心,即“伪E块”核心。例如:

他的头发|白得过多。 S041J=SB+SC

(2)当形成的是三个以上主块句时不能用基本格式。如形成三主块句时,只能使用置E块于句末的规范格式或违例格式,如!11、!12、!21、!22。例如:

闭会期间的代表|活动|参加得比较少。 !21X=A+B+X
帝国主义|对新中国|恨得要死。 !11X20J=X2A+XBC+X2

但是当形成参照比较判断句jD021J时可以采用基本格式,这是因为jD021J基本格式就把E块置于句末,例如:

美国的经济实力||远比中国||强大得多。 jD021J=DB1+|h+DB2+jDC

简而言之,“得”在这里用作E块内部的组合符,即“得”把它前面的E块核心和它后面接的程度修饰词组合构成特征语义块E。由于修饰成分被“得”引导而放在了被修饰成分的后面,所以这是一个典型的反偏正结构。反偏正结构在英文中出现很普遍,而在汉语中除了“得”的这种用法之外比较少见。

1.2 表示对象显现出的状态和效应

对象和对象的表现是一个句子所描述内容的两个最基本方面。B是一般对象,A是特殊对象,区别在于A是作用者。E是一般表现,C是特殊表现,区别在于C和对象具有融合性而E没有。在具体句子中,E块表示动作行为,C块则表示这个动作行为使对象B显现出的状态。C块经常由“得”来引导,而且某些情况下C块可以块扩为句子。下面就分别加以阐述。

1.2.1 C非块扩

根据E块核心动词是否同对象B构成作用-对象关系将“得”后面的C分别看成YC和SC,分别形成一般效应句Y0J和一般状态句S0J。状态句中的动作行为是自身产生的;效应句中的动作行为是来自外界的,必有具体的作用者,只不过在句中省略罢了。例如:

火|会烧得|更旺。 S0J=SB+S+SC
元湾村的耕地||变得||千疮百孔。 S0J=SB+S+SC
思想政治工作||做得||更有效果。 Y0J=YB+Y+YC
墙||刷得||光润无比。 Y0J=YB+Y+YC

效应句Y0J可以写成多种等效的格式,以最后一句为例:墙被刷得光润无比;墙被我刷得光润无比;我把墙刷得光润无比。不管转化成哪种新格式,“得”后面接的“光润无比”总是受作用对象“墙”的效应表现。

1.2.2 C块扩

I.当特征语义块动词是来自概念节点0008和0009的动词,如:“使,逼迫”等,句类先验知识约定YC具有块扩性质,形成作用效应句XYJ。例如:

频繁的风灾、水灾、旱灾|逼得|一些村屯的人|[#远走他乡#]。 XYJ=A+XY+B+YC YC=I31T2bJ

II.当特征语义块动词是来自反应型概念的动词时,“得”后面引导的成分表示反应者的后续表现(记作X2C)。句类先验知识约定X2C具有块扩性质,形成的句类有后续反应句X200J和作用反应-后续反应句X2X200J*311J。例如:

在场的医务人员|都感动得|[#流下了热泪#]。 X200J=X2B+X200+X2C

大风雪|呛得|[#我透不上气#]。 X2X200J*311J=XAC+X2X200+X2B+X2C

简单的说，只有出现在“得”前面的动词才可能是全局特征语义块 Eg(g 表示 global)；如果还有出现在“得”后面的动词，那么它肯定是局部特征语义块 El(l 表示 local)，而且必定是块扩。

还有一点注意的是，在表达强烈语气感情时，“得”后面的效应部分常常省略。此时“得”表示的是“无法形容”的意味，如：“看你把他气得！”。这类句子非常口语化，基本不出现于书面语料库中。

1.3 表示许可、可能性

根据“得”后面接不接内容分为两种情形

1.3.1 “得”后面还接有成分

表示结果或趋向得以实现。相当于在复合动词中插入“得”，否定形式把“得”换成“不”。“得”的后面可以是动词或形容词或趋向动词，形式化表示为：动词+得+动词/形容词/趋向动词(上/下/来/出 etc)。完全可以用上装 j112c32 “能够”来代替这里的下装“得”的用法。形成的多是混合句类。例如：

她||看得懂||英文报纸。 她||能看懂||英文报纸。 XD01*21J=A+XD+DBC

你||爬得上||那座山么？ 你||能爬上||那座山么？ PT2b*20J=TA+T2b+TB2

我||睡不着。 我||不能睡着。 YS*11J=YB+YS

1.3.2 “得”后面不再接任何成分

表示情理上能否许可行为进行，或条件能否允许实现某种动作，主要用于口语。如果是用在形容词后则只有否定形式，即“形容词+不得”；而动词后面可以接肯定和否定两种形式，即“动词+得/不得”。例如：

酸葡萄||吃不得。 酸葡萄||吃得。 YJ=YB+Y

开车||快不得。 [开车快得。(有误)] S041J=SB+SC

2 特征语义块核心 Ek

“得”可以单独充当或组合构成特征语义块核心 Ek，此时“得”后面语义块的核心成分必定是静态(记作 g)、值(记作 z)或效应(记作 r)的概念类别。归纳为以下 3 种用法：

2.1 动态概念 v3ae01

意思是“获得”，用 HNC 符号表示为 v3ae01。“得”的这一义项作 Ek，形成一般效应句 Ya0J，句类表示式是 YB+Y+YC。Ek 后常接下装 hv 如“了，过”，YB 一般是表示人或组织结构的具体概念，即 p 或 pe 类概念，YC 即可以是具体概念也可以是抽象概念。例如：

她|得了|第一。 Ya0J=YB+Y+YC

我|像得了|妄想症。 Ya0J=YB+Y+YC

2.2 动态概念(j/vz04, /10, (j3; j41))

相当于“等于、得出结果”之义，用 HNC 符号表示为(j/vz04, /10, (j3; j41))，形成相互比较判断句 jD00J。“得”作此用法有一个显著的特征，那就是两边都是表示数量的语义成分。例如：

二四得八。 jD00J=DB1+jD0+DB2

(六|除以|三)得二。 jD00J=DB1+jD0+DB2 @DB1={XJ}

2.3 “得”的 EK 复合

现代汉语中有很多不登录的、含有“得”的动词，主要有以下两类：

I. “夺得、争得、拾得、求得、征得”等，含有 v3ae01 的意思，后面多接具体概念，形成效应句，例如：

张阜新|夺得|男子20和50公里两枚金牌。 Ya0J=YB+Y+YC

II. “搞得、弄得、引得”等，表示对象显现出的效应，“得”后面也可以形成块扩，相当于前面介绍的用法 1.2。例如：

科技下乡”活动|搞得|热热闹闹。 Y0J=YB+Y+YC

人民网展台|引得|众多观众|[#驻足^观看#]。 XYJ=A+XY+B+YC YC={31[XJA]33T1J}

也有很多已登录的把“得”作为构词成分出现的动词，它们都不再是“得”单字的用法了，不属于本文讨论范围。有以下三类：

I. “分得、得到、赢得、取得、记得、获得、博得、得逞、得出、得胜、得分、得势、得手”等，含有 v3ae01 的意思，形成效应句；

II. “觉得、懂得、认得、值得”等，形成判断句；

III “显得、得宠、得当、得法、得手、得体、得意、得志”等，形成状态句。

3 特征语义块上装 qv

qv 表示特征语义块的上装，和下装 hv 不同，qv 是放在特征语义块中动词前面的修饰语。作为上装的“得”主要用作基本判断逻辑修饰，陈述对势态的判断，相当于情态动词，分为以下两种用法：

3.1 对势态的判断是基于纯客观事实的

用 HNC 符号表示为 jluv129，相当于情态动词“需要”，例如：

{搬起|这块大石头||至少得用||四个壮汉。 Y0J=YB+Y+YC @YB={131T21}

当判断是基于客观事实时，特征语义块前面的 JK1(上句如 YB)经常形成句蜕，而且“得”后面所接的动词多表示“需要，耗用”概念，此时常常可以省略动词，如刚才的例子变为：

{搬起这块大石头||至少得四个壮汉。

所以计算机在作语言理解时一定要把省略的动词补上，否则把“得”理解为特征语义块核心 v3ae01 的概念，就会误解为“石头底下至少压着四个人”，从而弄出笑话来了。

3.2 判断含有主观色彩

用 HNC 符号表示为 jluv13c21，相当于情态动词“应当、应”，多用于口语中。例如：

她||得珍惜||这个来之不易的机会。 X20J =X2B+X2+XBC

4 “得”的处理策略

我们从《人民日报》语料中任意抽取了 200 个含“得”的句子，对其上述各种用法的分布进行了统计，结果如表 1：

各种用法的分布同文体有关，在新闻体语料中用法(6)的比例高；在口语、文学作品等的语料中，用法(1)、(2)、(3)和(8)的比例会高一些。

表 1 “得”各种用法的分布的统计

用法	句子数	百分比
(1)动作行为到达的程度	39	19.5%
(2)对象显现出的状态和效应	64	32%
(3)许可、可能性	21	10.5%
(4)“获得”	13	6.5%
(5)“等于、得出结果”	1	0.5%

(6) “得”的EK复合	40	20%
(7) “需要”	1	0.5%
(8) “应当、应”	21	10.5%
合计	200	100%

基于“得”的各种意义和用法，可将具体处理策略归结如下：

(1) 优先“得”可以同前后的字搭配成词构成E块的情形，然后进行句类检验，直到所有复合成词的句类假设落空为止。以下几步处理“得”的单字用法。

(2) 如果句中没有其他“v”概念，则假设“得”就是E块核心，然后通过检验“得”后面语义块的核心成分是否为静态g、值z或效应r的概念类别来确认。若确认失败，表明是无特征语义块句，例如用法1.1的特例——简明状态句。

(3) 如果“得”出现在E块核心“v”概念之后，则为E块下装hv。

I. “得”后面是一个句子时，表明是效应的块扩，可以根据E块核心的概念来判断包含的是后续反应句还是由作用效用句。

II. “得”后面紧跟“v”概念或“得”后面没有成分时，对应于用法1.3的部分用法，“得”表示可能性或许可。

III “得”后面跟的是修饰“v”概念的词，表示程度，对应于用法1.1。

IV. “得”后面跟的是u、*iu*、*vu*等属性修饰概念时，对应于1.2和1.3两种用法，既能表示状态效应，又能表示可能性。例如“衣服洗得干净”。要想区别开来，还得放到语境中看，前者意在描写，后者意在判断。

(4) 如果“得”出现在E块核心“v”概念之前，则为E块上装qv。如果E块核心表示“需要，耗用”的概念时，“得”为客观判断 $j_{huv}129$ ，否则为主观判断 $j_{huv}13c21$ 。

5 结束语

本文通过对具体语言现象的考察，运用HNC的语言表述模式，总结了“得”的各种意义和用法，并对语言理解小专家级的处理策略和步骤进行了探讨。由于还没来得及在工程上付诸实现，所以对“得”处理策略的处理成功率还不能给出量化的数据。而且，要获得关于“得”字各种用法出现的频率的精确统计数字，有必要在更大规模的真实语料进行考察。这些都是本文接下去要做的工作。

附录：

符号解释：

	全局语义块间隔符		局部语义块间隔符	Λ	复合句连接符号
^	语义块指示符标记	[# #]	块扩标记	{ }	原型句蜕标记
@	语义块内部结构说明引导符				

句类表示式（部分）：

一般状态句	$S=SB+S$	$S0J=SB+S+SC$	一般效应句	$YJ=YB+Y$	$Y0J=YB+Y+YC$
反应—状态句	$X20S*11J=X2B+X20S$		作用效应句	$XYJ=A+XY+B+YC$	
简明状态句	$S041J=SB+SC$		一般作用句	$X=A+X+B$	
一般反应句	$X20J=X2B+X20+XBC$		参照比较判断句	$jD021J=DB1+jh+DB2+jDC$	
后续反应句	$X200J=X2B+X200+X2C$		一般接收句	$T1J=TB+T1+T1C$	
相互比较判断句	$jD00J=DB1+jD0+DB2$		自身转移句	$T2bJ=TA+T2b+TB2$	

作用反应—后续反应句 $X2X200J*311J=XAC+X2X200+X2B+X2C$

三主块句格式代码表（部分）：

规范格式 $!11J=JK1+^JK2+E$ $!12J=JK2+^JK1+E$

违例格式 !21J=JK1+JK2+E !22J=JK2+JK1+E
省略格式 !31J=E+JK2

参考文献

- [1] 黄曾阳. HNC(概念层次网络)理论. 北京:清华大学出版社,1998.
[2] 晋耀红,张全,杜燕玲. HNC 句类分析系统. HNC 与语言学研究. 武汉:武汉理工大学出版社,2001.
[3] 苗传江. HNC 句类知识研究[博士学位论文]. 北京:中国科学院声学研究所,2001.
[4] 吕叔湘. 现代汉语八百词(增订本). 北京:商务印书馆,1999.
[5] 侯学超. 现代汉语虚词词典. 北京大学出版社,1998.
[6] 杨成寅. 现代汉语句型概论. 内蒙古教育出版社,1993.

致谢 中国科学院声学研究所的张全研究员对本文的工作给予了细心的指导,在此表示感谢.

the Understanding and Processing of “得” based on the Theory of HNC

XIONG Liang

(Institute of Acoustics, Chinese Academy of Sciences, Beijing 100080, China);

E-mail: bright84@sina.com

Abstract: This paper discusses the meaning and usage of “得”, from the point of view of understanding and processing the sentence, with the use of expression mode for sentences according to HNC Theory. It investigates the distribution of each meaning item in the real language materials, and brings forward the processing strategy of “得”.

Key words: 得; feature chunk; Sentence-Category; HNC