

# 基于小句关联理论的有标选择 复句层次关系自动识别

尹蔚, 罗进军

(华中师范大学, 湖北 武汉 430079)

[摘要] 以小句关联理论为理论基础, 结合具体事例重点探讨“两句式”“三句式”“四句式”有标选择复句层次关系的自动识别方法。

[关键词] 小句关联理论; 有标选择复句; 层次关系; 自动识别; 表里关联

[中图分类号] H14 [文献标识码] A [文章编号] 1674-117X(2016)06-0094-04

## Automatic Identification of Hierarchy of Marked Selective Compound Clauses based on Clausal Correlation Theory

YIN Wei, LUO Jinjun

(Central China Normal University, Wuhan 430079, China)

**Abstract:** In view of theory of clausal correlation, it analyzes the automatic identification of hierarchy of two - clause, three - clause and four - clause marked selective compound clauses. It is essential to establish a knowledge base for the computer to identify the hierarchy of marked selective compound clauses from its outer sequence.

**Key words:** clausal correlation theory, marked selective compound clauses, hierarchy, automatic identification, correlation between outer form and inner meaning

### 一 引言

有关复句特别是多重复句层次关系的本体研究, 汉语学界做了大量卓有成效的工作。邢福义认为, 多重复句的分析必须善于抓住关联词语, 进行全面观察, 然后逐层深入地剖析。关联词语是形式上的标志。不管是大层次还是小层次, 都要善于根据关联词语来断定关系。没有运用关联词语的, 也

要看能够添上什么样的关联词语, 借以断定其关系。抓住关联词语进行全面观察时, 要善于化繁为简。<sup>[1]</sup>

此外, 沈开木、<sup>[2]</sup> 蔡建中、<sup>[3]</sup> 石安石、<sup>[4]</sup> 肖国政、<sup>[5]</sup> 陆丙甫、金立鑫、<sup>[6]</sup> 汪国胜、刘秀明<sup>[7]</sup> 等也提出了不少真知灼见。鲁松、白硕、李素建等认为, 汉语多重关系复句的句法分析问题主要是由复句中

收稿日期: 2016-09-19

基金项目: 教育部人文社会科学研究青年基金项目“本用结合视野下的有标选择复句研究”(11YJC740137); 国家社会科学基金项目“现代汉语有标复句层次关系信息化研究”(10CYY034); 国家社科基金重大项目“全球华语语法研究”(11&ZD128); 教育部人文社会科学重点研究基地重大项目“汉语复句关系词语的理论解释和实际应用”(13JJD740012); 澳门大学研究委员会立项课题“Grammatical Innovations in Chinese on the Internet and Their Linguistic Implications”(MYRG093-FSH13-XJ)

作者简介: 尹蔚(1980-), 女, 湖南株洲人, 华中师范大学副教授, 博士, 研究方向为现代汉语语法、理论语言学; 罗进军(1978-), 男, 湖南新化人, 华中师范大学副教授, 博士, 研究方向为现代汉语语法、中文信息处理。

的关系分析和层次分析组成。为了有效地形式化地表示多重关系复句的层次结构,该文提出了关系层次树的概念,并以此为基础构造文法,采用部分数据驱动的确定性移进-归约算法实现多重关系复句的关系层次分析。<sup>[8]</sup>

小句关联理论对有标复句层次关系的自动识别做了较为系统的研究。提出了“小句关联体”“辖域”“关联指向”“连用”“合用”一系列概念,并探索了具体的识别流程:计算机先从有标复句中提取语表序列→将所提取的语表序列跟后台表里关联模态库中已有的语表序列进行比较,找到最相似的甚至相同的→给出语表序列相应的层次关系。<sup>[9]</sup>因为有科学实验作为有力支撑,所以该文的很多想法都在实践层面得到了验证。

## 二“两句式”有标选择复句层次关系的自动识别

“两句式”有标选择复句的层次关联模式只有一种,那就是“1”式。所谓“1”式是指在第一小句末尾是第一层,用竖线可表示为:“①|②”。因此,“两句式”有标选择复句的层次关系的自动识别相对来说比较简单,如果是单纯型语里关系,那就更是如此。如:

(1)①白五爷是专程看望我们,②还是另有别的事情要办呢?

(2)①事实上她也抨击到了我,②或者说我也受到了误伤。

计算机提取以上两例的语表序列,得到的结果见表 1。

语表序列的聚类,也就是到我们事先建好的有标选择复句表里关联模态表里去匹配最相似的语表序列模式。结果发现例(1)正好跟 1 号语表序列模式匹配,例(2)跟 2 号语表序列模式匹配。

表 1 有标选择复句表里关联模态表

ID	小句构件数目	语表序列模式	层次关系模式
1	2	①qqrexz②hqrcxz	1xz
2	2	①kb②hqrcxz	1xz

如果是复合型语里关系,如:

(3)①要么是好人,②否则就是坏人。

计算机提取例 3 有标选择复句的语表序列,得

到结果见表 2。

表 2 有标选择复句表里关联模态表

ID	小句构件数目	语表序列模式	层次关系模式
4	2	①qqqcxz②hqqcjz	1xz + jz

两句式有标选择复句的复合型语里关系除了选择关系跟假转关系的复合外,还有选择关系跟假设关系的复合,甚至还有三合型语里关系。但是不管是哪一种复合型语里关系,其自动识别的流程都是类似的。

## 三“三句式”有标选择复句层次关系的自动识别

“三句式”有标选择复句就层次关联而言有三种模式,即“11”式、“12”式、“21”式。所谓“11”式是指在第一小句末尾是第一层,在第二小句末尾也是第一层;所谓“12”式是指在第一小句末尾是第一层,在第二小句末尾是第二层;所谓“21”是指在第一小句末尾是第二层,在第二小句末尾是第一层。如:

(4)①要么是名,②要么是利,③要么是爱。

(5)①是我真的如此自私,②还是我爱他当真不如对世民,③因而不能为他牺牲半分?

(6)①是声音低,②所以大伙没听清,③还是这帮人没来?

我们看计算机如何自动识别(4-6)例有标选择复句的层次关系。让计算机提取它们的语表序列,得到结果见表 3。

将它们放到有标选择复句表里关联模态表中去匹配,结果是,例(4)语表序列跟 6 号语表序列模式可以聚类,例(5)语表序列跟 7 号语表序列模式可以聚类,例(6)语表序列跟 19 号语表序列模式可以聚类。

表 3 有标选择复句表里关联模态表

ID	小句构件数目	语表序列模式	层次关系模式
6	3	①qqqcxz②hqqcxz ③hqqcxz	1xz1xz
7	3	①qqrcxz②hqrcxz · qqcyg ③hqqcyg	1xz2yg
19	3	①qqrcxz · qqcyg ②hqqcyg③hqrcxz	2yglxz

“三句式”有标选择复句的层次关系肯定不止以上所说的三种,我们只是以“三句式”有标选择复句的层次关联模式为标准,每种类型都选取一个代表,具体说明它们层次关系的自动识别过程。

四 “四句式”有标选择复句层次关系的自动识别

从理论上讲,“四句式”有标选择复句的层次关联模式应该有 11 种。如果有三个第一层,那就是“111”式。如果有两个第一层,那就有三种模式,分别是“112”式、“121”式、“211”式。如果是只有一个第一层,那就有七种模式,分别是“122”式、“221”式、“212”式、“123”式、“132”式、“231”式、“321”式。

先看第一种情况,即有三个第一层的。如:

(7)①要么手狠心毒,②要么道行深广,③要么法力无边,④要么树大根深。

例(7)有三个第一层,如果用传统的竖线分析法,都可以表示为:①|②|③|④。计算机提取的语表序列,得到结果见表 4。

将它们放到有标选择复句表里关联模态表中去匹配,发现例(7)可以与 30 号语表序列模式聚类。

表 4 有标选择复句表里关联模态表

ID	小句构件数目	语表序列模式	层次关系模式
30	4	①qqqcxz②hqqcxz	1xz1xz1xz
		③hqqcxz④hqqcxz	

再来看第二种情况,即有两个第一层的。如:

(8)①或者委婉地规劝,②或者从大的方面广泛地论述,③或者从小的方面细细地指点,④知无不言。

(9)①或者用彩色复印机复印彩色盒带封面,②或者将盒带封面翻拍下来,③重新制版印刷,④或者内外勾结从音像出版单位、印刷厂套购乃至盗购盒带封面。

(10)①或者是作品发表了,②论文获奖了;③或者是自学考试最终拿到了文凭;④或者研制的产品最终通过了鉴定。

具体一点讲,例(8)是“112”式,即“①|②|③|

|④”;例(9)是“121”式,即“①|②||③|④”;例(10)是“211”式,即“①||②|③|④”。计算机提取的语表序列,得到的结果见表 5。

计算机将语表序列放到有标选择复句表里关联模态表中去匹配,结果发现,例(8)语表序列可以跟 31 号语表序列模式聚类,例(9)语表序列可以跟 33 号语表序列模式聚类,例(10)语表序列可以跟 34 号语表序列模式聚类。

表 5 有标选择复句表里关联模态表

ID	小句构件数目	语表序列模式	层次关系模式
31	4	①qqqcxz②hqqcxz	1xz1xz2bl
		③hqqcxz④kb	
33	4	①qqqcxz②hqqcxz	1xz2bl1xz
		③kb④hqqcxz	
34	4	①qqqcxz②kb	2bl1xz1xz
		③hqqcxz④hqqcxz	

最后来看第三种情况,也就是只有一个第一层的。如:

(11)①是真法盲,②还是知法犯法,③见利忘法,④无法无天?

(12)①是按传统思路上项目,②争贷款,③铺摊子,④还是另辟蹊径?

(13)①或者武功很高,②却不如楚留香侠义;③或者富可敌国,④却不如楚留香大方。

(14)①是宣泄个人的弗洛伊德所说的各种“情结”,②还是引人向上,③使人更成为人,④而不是沦为兽?

(15)①今年我们很可能没有一个能考上北大,②或者即使有那么一个勉强考上了,③也不值得骄傲,④因为其他学校据说有望考十几个。

(16)①是为了维护党和人民的利益,②不怕吃苦,③不怕牺牲,④还是一事当前先替自己打算?

(17)①他既考不上北大,②也考不上清华,③甚至很有可能连一般本科也考不上,④或者说他根本就不是块读书的料。

(11)-(17)这七例分别为“122”式、“221”式、“212”式、“123”“132”“231”“321”式。计算机提取的语表序列,然后将这些语表序列放到有标选择复句表里关联模态表中去匹配,得到的结果见表 6。

表6 有标选择复句表里关联模态表

ID	小句构件数目	语表序列模式	层次关系模式
41	4	①qqqcxz②hqqcxz ③kb④kb	1xz2bl2bl
42	4	①qqqcxz②kb ③kb④hqqcxz	2bl2bl1xz
163	4	①qqqcxz②hqqczz ③hqqcxz④hqqczz	2zz1xz2zz
755	4	①qqqcxz②hqqcxz ③hqqedj④hqrcbl	1xz2dj3bl
1193	4	①kb②hqqcxz·qqqerb ③hqqerb④hqqcycg	1xz3rb2yg
1488	4	①qqqcxz·qqremd ②kb③kb④hqqcxz	2md3bl1xz
1955	4	①qqrcbl②hqrcbl ③hqqedj④hqqcxz	3bl2dj1xz

“四句式”有标选择复句的层次关系不止上述这些,我们根据层次关联的类型选取代表性的例子。不管是哪种类型的有标选择复句,其层次关系自动识别的原理都是一样的,即在句法关联跟语义关联之间寻求一种对应关系,从而达到“由表识里”的目的。

五 结论

我们根据有标选择复句层次关联类型,结合具体事例示范性地研究了“两句式”“三句式”“四句式”有标选择复句层次关系的自动识别问题。

第一,研究实践表明,基于小句关联理论的有标选择复句层次关系自动识别研究是可行的。句法关联在有标复句中主要表现为由关系标记构成的语表序列,而语义关联则主要体现为层次关系。要想让计算机根据有标复句的语表序列识别其层次关系,就必须事先让计算机“获取”相关知识,其中相当重要的一环就是建立知识库,也就是我们所说的有标复句表里关联模态表,即在语表序列模式跟层次关系模式之间构建对应关系。

第二,我们意识到,与其他有标复句一样,有标选择复句层次关系的自动识别也是一项浩大、繁复的系统工程。其中涉及到多个环节,比如,关系标记的自动识别,非小句的自动识别,语表序列的自动提取,语表序列的聚类,语表序列聚类规则表的构建,表里关联模态表的构建,表里关联框式简图图库的建设,表里关联树型图图库的建设等等。当然,我们也必须认识到,语言现象是丰富多样的,有标复句的运用也是相当灵活的,因此,如何根据新的语言事实不断完善我们现有的工作是今后相当长一段时期内我们必须考虑的问题。我们坚信,只要有正确的理论做指导,有务实的学风做保障,有创新的思想做动力,就一定能够将我们的研究工作不断推向前进。

参考文献:

[1] 邢福义. 谈谈多重复句的分析[J]. 语文教学与研究, 1979(1):48-53.

[2] 沈开木. 句法的层次性、递归性及其在多重复句分析中的利用[J]. 汉语学习,1982(5):22-24.

[3] 蔡建中. 汉语多重复句层次关系的确认[J]. 西北民族大学学报(哲学社会科学版),1982(1):70-74.

[4] 石安石. 怎样确定多重复句的层次[J]. 语文研究,1983(2):39-43.

[5] 肖国政. 关于多重复句的划分[J]. 语文教学与研究, 1983(9):36-37.

[6] 陆丙甫,金立鑫. 关于多重复句的层次问题[J]. 汉语学习,1988(5):4-7.

[7] 汪国胜,刘秀明. 关于多重复句的分析[J]. 华中科技大学学报(社会科学版)2004(6):84-88.

[8] 鲁松,白硕,李素建,等. 汉语多重关系复句的关系层次分析[J]. 软件学报,2001(7):987-995.

责任编辑:李珂