汉英机器翻译中描述型复句的关系识别与处理*

鲁 松[1], 宋 柔[2]

[1](中国科学院 计算技术研究所,北京 100080);

[2](北京工业大学 计算机学院,北京 100044)

E-mail: lusong@ict.ac.cn

摘要: 汉英机器翻译的复句处理不仅要依托于单句的处理,而且还要超越单句的辖域,深入考察复句内分句之间的内在联系.其中,在汉语描述型复句中存在着大量的无特定语言标记的非并列关系复句,为了辨别其中的内在联系,实现英语译文的正确生成,针对不同情况,给出了完整的关系判定规则,并提出采用中心分句动态判定方法来解决部分复句处理规则局部性的问题,最后通过实验系统得以验证.

关键词: 汉英机器翻译;描述型复句;复句内在联系判定;中心分句动态判定

中图法分类号: TP391 **文献标识码**: A

汉语的复句处理是面向篇章的汉英机器翻译中最常见和最困难的问题之一.其中包括汉语复句内分句的内在联系判定、各种回指现象的处理以及相应英语生成等问题.尽管在这些方面已有的研究大多数是从传统语言学角度出发的[1~3],未做深入的形式化、算法化的工作,但其中的探索是极有价值的[4,5].我们的工作是针对汉英机译复句处理中分句间内在联系的问题而进行展开的.

无论汉语还是英语,复句在其语言体系中都占有极其重要的语法和语义地位,是一种表达作者意图的必不可少的语言形式.同时,复句是由单句经过一定加工后组合而成的,而这种加工后的组合不是随心所欲的,是必须基于某种联系的紧密结合[6~8].

为了构建汉英双语的共性框架,我们依据表述类型扩充了汉语复句中隐含的深层联系.如图1所示.

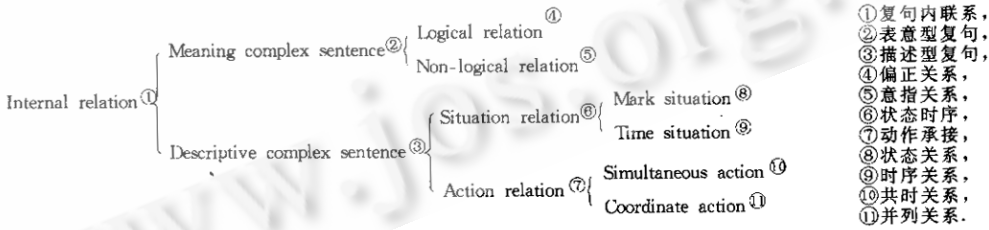


Fig. 1 Internal relation of complex sentence

图1 扩充后的复句内在联系

图1中,复句内在联系分类体系的建立,是根据分句间语义联系的共性和个性确立起来的.其中的两大类是表意型复句和描述型复句.限于篇幅,有关表意型复句的讨论将另文介绍.在此,本文仅对汉英机器翻译中描述型复句的内在联系判定的形式化处理方法进行论述.

* **收稿日期**: 1999-01-13; **修改日期**: 2000-04-19

基金项目: 国家自然科学基金资助项目(69773008);国家863高科技项目基金资助项目(863-306-2D02-01-3);国家973高科技项目基金资助项目(G1998030510)

作者简介: 鲁松(1972—),男,北京人,博士生,主要研究领域为自然语言处理,机器翻译;宋柔(1945—),男,江苏人,教授,博士生导师,主要研究领域为自然语言处理,人工智能.

其中的描述型复句是由强调一定状态下事件发生的状态时序复句和描述对象动作过程的动作承接复句构成,并进一步可将二者划分为图1中的4种关系.必须指出的是,这些关系是构成复句联系的原子关系,即在一个完整复句的构成中,其内在联系可以是一种或多种原子关系的有效组合.

同时,由于汉语自身具有简洁性这一特点,只要使用意合法不会造成汉语理解上的歧义,则所有能省略的分句内语法成分和分句间连接成分就可以或必须省略.这便给汉语复句内在联系的形式化判定工作带来了极大的困难.另外,由于英语中大量依靠形式标记(连词、介词或词形变化)来指明复句关系这一特点,也使汉英复句相应内在联系的直接的一一对应成为不实之想.

针对上述问题,本文在把握各分句中心谓词及其相关词语语义信息的基础上,给出关系匹配规则,判定分句在复句中所处的语义地位,并提出动态判定中心分句的方法以解决不同类型分句特殊语义地位判定的问题.

1 研究对象的界定

为了明确研究目标,解决关键难点,我们对研究对象作如下约定[9]:

(1) 已经通过时间方位词或时间副词等形式化标记,指明时序关系或共时关系的复句不在讨论之列.

(2) 本文讨论的复句中各分句之间不存在表明偏正关系的关联词语.

(3) 复句中各分句主语语义指称一致.在描述型复句中,若述题不一致,则必然存在时间方位词、时间副词或相应连词等形式化标记指示动作时序.

(4) 本文不涉及多个原子关系构成复句内在联系这种复杂情况的处理.

2 概念定义与规则条件的BNF定义

2.1 概念定义

面向机器翻译的汉语描述型复句关系是依据复句内各分句的语义联系而在语义层面上进行定义的,所以,其内部关系的识别工作也必须是基于语义层面来进行的.因此,为了实现复句内部关系的清晰划分,就必须对作为分句内部关键语法成分的词类在语义层面上按照划分要求进一步细分.

以下是本文在对动词、副词和连接成分细分时重新定义的语言学概念:

(1) 痕迹类动词——是指由动词描述的动作完成后,动作造成的结果仍能维持一段时间或成为一种状态的动词.比如,画、涂、漆...这类动词后面邻接虚词"着",并且前面不邻接时间副词"正"作为分句谓语结构时,由它构成的分句就不再是动作句,而是状态句,如:(墙上)正写着(红字)、(嘴上)正涂着(红嘴唇).

(2) 着装类动词——用来描述人们穿衣戴帽的词语,如:头戴、身穿、脚踏...它反映了动作实施者的着装情况,以它作谓语的句子也是非动作的状态句.着装类动词与痕迹类动词存在着交集,这也反映了二者在表现状态上的共性.

(3) 持续动词——所描述的动作在时间上可以持续一段时间,如:吃、跑、研究...

(4) 非持续动词——所描述的动作在时间上不能持续,如:开始、牺牲、丢...

(5) 感官动词——所描述的动作由生物的感觉器官发出或通过感觉器官接受,如:看、听、闻...

(6) 心理动词——表现人头脑中思维过程或内心活动的动词,如:觉得、意识到、察觉...

(7) 承受类动词——所涉及的动词有承受、收受之义,如:受到、遭到、收到...

(8) 能力能愿动词——反映施动者主观能力的能愿动词,如:能、能够、可以...

(9) 需求能愿动词——一类能愿动词,所涉及的动作是施事按照客观需求而进行的,如:必须、应该、不得不、最好...

(10) 以后时间连接成分——以两事相随来表示事件发生先后的连接成分,如:然后、后来、此后、随即...

(11) 共时连接成分——表示两件事同时发生的连接成分,如:同时.

(12) 持续性时间副词——描述动作的发生具有持续性的副词,如:一直、天天...

(13) 强调性副词——带有反问含义,强调不同寻常的副词,如:居然、竟然...

2.2 规则条件的 BNF 定义

描述系统中规则条件的 BNF 形式定义如下:

〈规则条件〉::=〈汉语语法语义信息〉

〈汉语语法语义信息〉::=〈汉语句组信息〉〈汉语分句信息〉{〈汉语分句信息〉}{〈汉语分句间约束关系〉}

〈汉语句组信息〉::=S=C〈分句编号下标〉{+C〈分句编号下标〉}

〈汉语分句信息〉::=[*]C〈分句编号下标〉=〈语法成分结构信息〉{+〈语法成分结构信息〉}[+α〈分句编号下标〉]*[;〈语法成分结构间约束关系〉]

〈分句编号下标〉::=1|2|3|.....

〈语法成分结构信息〉::=[[]〈语法成分结构名〉〈分句编号下标〉[{〈左部约束条件〉;〈中心词约束条件〉;〈右部约束条件〉}][]]

〈语法成分结构名〉::=主语|谓语|连接词|宾语|能愿动词|补语|时量词组

〈左部约束条件〉::=〈结构约束条件〉{〈逻辑连词〉〈结构约束条件〉}

〈中心词约束条件〉::=〈结构约束条件〉{〈逻辑连词〉〈结构约束条件〉}

〈右部约束条件〉::=〈结构约束条件〉{〈逻辑连词〉〈结构约束条件〉}

〈结构约束条件〉::==("〈汉字词语〉"{〈逻辑连词〉"〈汉字词语〉"})|≠("〈汉字词语〉"{〈逻辑连词〉"〈汉字词语〉"})

|=(〈结构类别〉{〈逻辑连词〉〈结构类别〉})|≠(〈结构类别〉{〈逻辑连词〉〈结构类别〉})

〈逻辑连词〉::=∧|∨|-

〈结构类别〉::=痕迹类动词|着装类动词|持续动词|非持续动词|能力能愿动词|需求能愿动词|持续性时间副词|强调性副词|感官动词|心理动词|离合动词|承受类动词|以后时间连接成分|共时连接成分

〈语法成分结构间约束关系〉::=φ(〈语法成分结构名〉〈分句编号下标〉)

〈汉语分句间约束关系〉::=R=〈分句间约束条件〉{〈逻辑连词〉〈分句间约束条件〉}

〈分句间约束条件〉::=SameSub(〈语法成分结构名〉〈分句编号下标〉{,〈语法成分结构名〉〈分

---

* α表示由句中一个或多个语法成分构成,引用它的目的仅是为了描述简便.例如,某分句的"汉语分句信息"是"C_1=主语$_1$+α_1",表示第1分句C_1由主语和其他语法成分构成,这里不关心非主语语法成分的存在与否和约束条件.

句编号下标〉})|HasMainCl*

3 规则与中心分句判定

描述型复句的关系识别工作是4小类描述型复句关系对复句的指称过程,也是从复句形式到其反映的复句关系的一个映射过程.这个过程是采用规则的形式完成的.必须指出,复句中不同内在关系各自的语法语义特点决定了规则策略也必须分门别类地进行判定.因此,我们按照规则对复句管辖约束的范围,将规则分为以下两种类型:

(1) 可以管辖约束整个复句的规则,即规则条件对汉语复句中每个分句的关键语法成分均加以限定.

(2) 仅能管辖约束复句中一个分句的规则,即规则条件仅对汉语复句中一个分句及其相邻分句的关键语法成分加以限定.

为行文方便,我们将可以管辖约束整个复句的规则称为A类规则,将仅能管辖约束复句中一个分句的规则称为B类规则.

A类规则能够管辖约束整个复句,所以识别过程较B类规则简单;B类规则仅能管辖约束复句中的一个分句,无法控制整个复句,故需对满足B类规则的分句在整个复句中的语义地位作出评价.我们将B类规则的这一特性称为B类规则的局部性(localization).也正是由于它的这一特性,给其形式化工作带来一定的困难.

A类规则适用于时序关系复句的识别,B类规则适用于状态关系复句和共时关系复句的识别.既不满足A类规则也不满足B类规则的汉语复句将被判定为并列关系复句.

3.1 时序关系复句的识别

在没有时间方位词或时间副词指示的情况下,复句内同样存在分句间的时序关系,如下面的A类规则(限于篇幅,英译形式的形式化描述略):

[规则A.1].

(1) 汉语的语法语义信息:

$S=C_0+C_1$

C_0=主语$_0$+谓语$_0$[≠(时间介词词组);=(非持续动词);]+宾语$_0$+α_0;φ(能愿动词)

C_1=主语$_1$+谓语$_0$[≠(时间介词词组)∧=("就");=(非持续动词);]+α_1;φ(连接词)∧φ(能愿动词)

R=SameSub(主语$_1$,主语$_2$)

(2) 英译形式:英文中需添加时间连接词'as soon as',以指明前后分句的时序关系.

例1:他收到消息,立刻就起程了.{词类(谓词(分句$_1$))=非持续动词['收到'];词类(谓词(分句$_2$))=非持续动词['启程'];承接词(分句$_2$)='就'}

He started as soon as he received the news. {分句$_1$处理为as soon as引导的从句形式}

[规则A.2].

(1) 汉语的语法语义信息:

$S=C_0+C_1$

---

* 释例:SameSub(主语$_1$,主语$_2$,)表示主语$_1$和主语$_2$在语义上均由一个固定对象指明;HasMainCl表示要求句组中存在中心分句.

C_0=主语$_c$+谓语$_0$[≠(时间介词词组);=(非持续动词);]+宾语$_0$+α_0;φ(能愿动词)

C_1=主语$_1$+谓语$_0$[≠(时间介词词组)∧≠("就");=(非持续动词);]+α_1;φ(连接词)∧φ(能愿动词)

R=SameSub(主语$_1$,主语$_2$)

(2) 英译形式:英文中需添加时间连接词'when',以指明前后分句的时序关系.

例 2:士兵回到村子里,发现自己颇像一个英雄.{词类(谓词(分句$_1$))=非持续动词['回到'];词类(谓词(分句$_2$))=非持续动词['发现'];承接词(分句$_2$)≠'就'}

The soldier found himself something of a hero when he returned to his village.

[规则 A.3].

(1) 汉语的语法语义信息:

$S=C_0+C_1$

C_0=主语$_0$+谓语$_0$[≠(时间介词词组);=(非持续动词);]+宾语$_0$+α_0;φ(能愿动词)

C_1=主语$_1$+谓语$_0$[≠(时间介词词组)∧≠("就");=(承受类动词);]+α_1;φ(连接词)∧φ(能愿动词)

R=SameSub(主语$_1$,主语$_2$)

(2) 英译形式:英文中需添加时间连接词'when',以指明前后分句的时序关系.

例 3:到了那儿,我们受到了热烈欢迎.{词类(谓词(分句$_1$))=非持续动词['到'];词类(谓词(分句$_2$))=承受类动词['受到']}

When we reached there, we were warmly received.{分句$_1$处理为 when 的时间状语从句形式}

上面给出的部分例句表明,在没有时间方位词或时间副词指示的汉语复句中,同样可能存在由各分句动词类型的进一步细分和相关副词和谐搭配表现出来的隐含时序关系.这些汉语复句在机译时必须区别于将复句划分为单句的简单处理方式,辨别其中的隐含时序关系,选用恰当的时间连词引导的英语方式处理.同时,应该注意到 A 类规则管辖约束整个复句的特点.

3.2 共时关系复句和状态关系复句的识别

3.2.1 初步规则

共时关系复句和状态关系复句的识别是由 B 类规则完成的,规则名称如下:

共时关系复句识别规则:

[规则 B.1]. 感官动词谓词规则;

[规则 B.2]. 心理动词谓词规则;

[规则 B.3]. "中心谓词+'着'";

[规则 B.4]. 一般持续动词谓词句.

状态关系复句识别规则:

[规则 B.5]. 着装类动词谓词规则;

[规则 B.6]. 痕迹类动词谓词规则.

这 6 条 B 规则的描述是由在复句中不特定位置的关键分句的句内语法语义约束条件、相邻分句句间约束条件以及相应的译文处理所组成.以[规则 B.2]心理动词谓词规则为例,规则条件形式化定义如下:

[规则 B.2].

(1) 汉语的语法语义信息:

$S=C_0+C_1+\ldots+C_n$

$C_0=\ldots$

$\ldots$

$C_i=$主语$_i$+谓语$_i$[;=(心理动词);≠("了")]+a_i;φ(能愿动词)∧φ(持续性时间副词)∧φ(强调时间副词)

$C_{i+1}=$[[连接词$_{i-1}$[;≠(转折连接词);]]+主语$_{i-1}$+a_{i+1};φ(能愿动词)∧φ(持续性时间副词)∧φ(强调时间副词)

$\ldots$

$C_n=\ldots$

$R=$SameSub(主语$_{i-1}$,主语$_i$,主语$_{i+1}$)

(2) 英译形式:英文中需将分句 C_i 处理为动名词短语形式.

例 5:他觉得不舒服,便下了马.{词类(中心谓词(分句$_1$))=心理持续动词['觉得']}

Feeling uncomfortable in his bowels, the man dismounted. {分句$_1$ 处理为动名词短语形式}

有别于时序关系判定的 A 类规则可以管辖约束整个复句的情况,复句中某分句若满足 B 类规则,我们只能称这条规则能够管辖约束这个复句中的该分句,但不能称此规则可以管辖约束整个复句.正是由于 B 类规则的局部性这一特点,决定了被判定为存在共时关系或状态关系的汉语复句在英文翻译时的不确定性.如下面两个例子:

例 6:他站了起来,感到了热,也感到了自己的年龄.

例 7:他感到了热,也感到了自己的年龄.

规则集管辖约束分句的情况如下:

	分句 1	分句 2	分句 3
例 6	不满足 A/B 类任何一条规则	满足[规则 B.2]心理动词谓词句	满足[规则 B.2]心理动词谓语句
例 7	满足[规则 B.2]心理动词谓词句	满足[规则 B.2]心理动词谓词句	φ

满足 B 类规则的分句,按 B 类规则被处理为动名词短语形式,上述两个例句的英文翻译结果如下:

例 6. He rose up, feeling the heat, and feeling his age. (√)

例 7. He feeling the heat, and feeling his age. (×)

从例 6 和例 7 这两句中文的相似性和英语译文的一正一误,说明了 B 类规则的局部性给形式化带来的困难,也说明了只有强化 B 类规则对整个复句的管辖约束能力才能解决这一问题.

3.2.2 中心分句及其判定算法

3.2.2.1 矛盾与问题

在共时关系复句和状态关系复句中,分句满足规则可能有以下几种事件发生:

事件 1. $(\forall c)(\exists r)$govern(r,c)

事件 2. $(\exists c)(\forall r)\neg$govern$(r,c)\wedge(\exists c)(\exists r)$govern$(r,c)$

事件 3. $(\forall c)(\forall r)\neg$govern$(r,c)$

其中 $c\in C, r\in R$,C 为一个复句中的分句集合,R 为 B 类规则集合;二元谓词 govern(r,c):规则 r 能管辖且是唯一能管辖分句 c 的规则.

在上述3个事件中,当事件3的情况发生时(即复句中的每个分句均不满足A/B规则集合中任一规则),表明该复句内分句间的联系是并列关系.在汉英机译处理时,按一般单句情况用并列连接词处理,在此我们无需讨论.而在事件1和事件2中,涉及到多个规则分别管辖约束复句中几个不同分句的复杂情况,进而也关系到了各分句在复句中的语义地位和分句间的关系.

依据3种事件对例6和例7作进一步分析:例6属于事件2,例7属于事件1.

由此,我们可以确定事件1和事件2的差异,即"$(\exists c)(\forall r)\neg govern(r,c)$"的逻辑真假,将是决定复句中各分句语义地位和得到正确译文的关键.

3.2.2.2 中心分句与非中心分句的概念与特点

为了解决上述问题,我们需要引入中心分句和非中心分句这两个概念来进一步解决B类规则局部性与复句处理需要整体控制之间的矛盾.

概念1. 中心分句:在复句中处于主要语义地位的分句,是话语者着重描述的动作或状态分句.

概念2. 非中心分句:在复句中处于从属语义地位的分句,是话语者为澄清想要着重描述分句的语言环境而给出的分句.

这两个概念属于语义范畴,表明复句中分句所处的语义地位.在复句机译处理中,被处理为动名词短语、介词短语、时间前置词短语或时间状语从句的分句视为非中心分句;支撑复句主干的分句视为中心分句,作为复句的语义重心.

同时,中心分句和非中心分句分别有如下特点:

中心分句	非中心分句
1. 一个复句中至少存在一个中心分句;	1. 非中心分句在复句中可有可无;
2. 中心分句是个相对概念,除极少语言点外,大多数情况都需借助非中心分句的判定来动态判定中心分句.	2. 非中心分句的判定条件是有限的; 3. 部分非中心分句判定的成立条件必须以存在中心分句为前提.

3.2.2.3 中心分句与非中心分句的充分条件

结合中心分句与非中心分句的概念和特点,我们得出它们的判定充分条件:

(a) 中心分句的[充分条件A]=(分句主语承前分句主语省略∨分句主语蒙后分句主语省略∨分句主语未省略)∧(系动词谓语句∨能愿动词句)

(b) 非中心分句的两类[充分条件B]:

(1) 充分条件I=(偏句∨时间句∨分句主语承前分句宾语省略∨分句主语承前分句宾语引出的分句)

(2) 充分条件II=(条件①∧条件②)

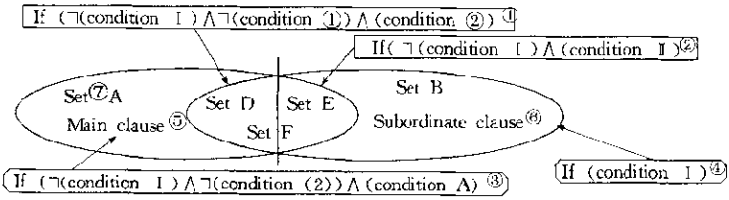
I. 条件①=存在中心分句

II. 条件②=$(\exists r)govern(r,c)$,c为当前待判定分句,$r\in R$={[规则B.1],[规则B.2],[规则B.3],[规则B.4],[规则B.5],[规则B.6]}.

各条件间的关系如图2所示.在图2中,伴随关系复句内所有分句的集合$=A\cup B\cup D\cup E$,其中被处理为中心分句的分句集合$=A\cup D$,被处理为非中心分句的分句集合$=B\cup E$,且集合$F=D\cup E$,其中A,B,E,D这4个集合两两交集为空($\varnothing$).

3.2.2.4 规则修订

根据上面对中心分句与非中心分句的讨论,必须对B类规则的条件判定进行必要的补充.我们还是以[规则B.2]心理动词谓词规则为例进行规则修订.



①满足(¬(充分条件Ⅰ)∧¬(条件①)∧(条件②)),②满足(¬(充分条件Ⅰ)∧(充分条件Ⅱ)),③满足(¬(充分条件Ⅰ)∧¬(条件②))∨(中心分句充分条件),④满足充分条件Ⅰ,⑤中心分句,⑥非中心分句,⑦集合.

Fig. 2 Judgement condition between main clause and subordinate clause
图2 中心分句与非中心分句判定的充分条件

[规则 B. 2. s].

(1) 汉语的语法语义信息:

$S=C_0+C_1+\ldots+C_n$

$C_0=\ldots$

...

$C_i=$主语$_i$+谓语$_i$[;=(心理动词);≠("了")]+α_i;φ(能愿动词)∧φ(持续性时间副词)∧φ(强调时间副词)

$C_{i+1}=$[[连接词$_{i+1}$[;≠(转折连接词);]]]+主语$_{i+1}$+α_{i+1};φ(能愿动词)∧φ(持续性时间副词)∧φ(强调时间副词)

...

$C_n=\ldots$

$R=$SameSub(主语$_{i-1}$,主语$_i$,主语$_{i+1}$)∧HasMainCl

修订的规则与原规则定义的唯一区别是仅在句间关系判定中添加一个逻辑谓语 HasMainCl,即确定该复句中存在非本分句的中心分句.若该谓词为真(true),则该分句被判定为满足规则;否则,被判定为不满足规则.这是判定该分句是否为复句中语义重心的关键,也是整个规则的难点.为了区分原规则和修订规则,在修订规则的规则号后添加字母's',如[规则 B. 2]的重写规则代号为[规则 B. 2. s].

3.2.2.5 规则级别的设定与比较原则

依据图 2 可知,复句中分句隶属集合 A 或隶属集合 B 是可局部判定的,即仅需考察本分句的 B 类规则的满足情况即可;而在集合 F 中分句对集合 D 和集合 E 的隶属问题,则不仅需要考察本分句,而且还需考察复句中其他分句的情况.进一步细化问题得到以下 3 种情况(设 c_1 为当前分句,判定 c_1 的集合隶属问题):

情况 1. subjection$(c_1,F)\wedge(\exists\ c)$subjection$(c,A)$;

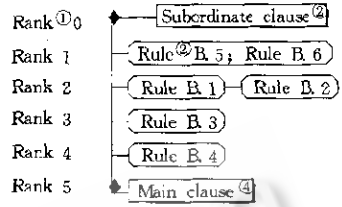
情况 2. subjection$(c_1,F)\wedge(\forall\ c)$subjection$(c,B)$;

情况 3. subjection$(c_1,F)\wedge(\forall\ c)\neg$subjection$(c,A)\wedge(\exists\ c)$subjection$(c,F)$.

其中,二元谓词 subjection(a,S):分句 a 隶属于图 2 中的集合 S,c_1 为当前分句,$c\in$ {复句中分句集合} $-\{c_1\}$.

相应于上述 3 种情况,只有在情况 1 和情况 2 下,分句 c_1 的隶属才是可判定的,即情况 1 中的 $c_1\in E$(非中心分句);情况 2 中的 $c_1\in D$(中心分句).而在情况 3 中,复句中至少存在两个分句在集合 F 中,换言之,它们都需要彼此判定对方是否为中心分句.此时,这个递归问题就成为整个中心分句判定问题的关键.我们结合客观实际和语义背景,对 B 类规则进行了级别设定,通过级别比

较,实现对情况3的可判定.为了使下节的算法描述更加简洁,将中心分句和非中心分句同时列入规则级别表,参加级别高低比较,如图3所示.



①级别,②非中心分句,③规则,④中心分句.

Fig.3 Rank list of rule B

图3 B类规则级别表

规则级别表的使用原则:

(1) 同一原则:若参加级别比较的分句均满足级别表中同一规则,则它们均被判定为中心分句.

(2) 同级共处原则:若参加级别比较的分句分别满足不同规则,但在级别表中共处同一单元块,则它们均被判定为中心分句,如图3'级别1'中的[规则B.5]和[规则B.6].

(3) 同级先后原则:若参加级别比较的分句满足级别表中同级的不同规则,且分处不同单元块,则需根据这些分句在现实句组中的前后位置来判定两个句子的隶属:在先的分句为非中心分句,置后的分句为中心分句.如图3'级别2'中的[规则B.1]和[规则B.2].

(4) 级别高低原则:若参加级别比较的分句满足级别表中不同级别的规则,则满足高级别的规则被判定为非中心分句,满足低级别的规则被判定为中心分句(级别0是最高级,级别5是最低级).

例8:她$_i$身穿西服,φ_i涂着红嘴唇(She was dressed in western-style clothes and printed lipstick).

例9:他$_i$身穿西服,φ_i感到很不自在(He felt uncomfortable in the western style clothes).

例10:我$_i$感到很奇怪,φ_i仔细地看着它(Feeling it rather strange, I looked at it carefully).

示例分析表为:

	第1分句C_1	第2分句C_2	级别判定	结果
例8	满足[规则B.5]	满足[规则B.6]	级别$_{\text{肯装动词句}}$=级别$_{\text{感官动词句}}$;	C_1,C_2均被判为中心分句,二分句均不满足规
	$C_1\in F$	$C_2\in F$	同级原则;	则s5s和规则s6s,$C_1\in D$,$C_2\in D$
例9	满足[规则B.5]	满足[规则B.2]	级别$_{\text{肯装动词句}}$<级别$_{\text{感官动词句}}$;	C_1为非中心分句,C_2为中心分句;C_1满足规
	$C_1\in F$	$C_2\in F$	级别高低原则;	则s5s,C_2不满足规则s6s;$C_1\in E$,$C_2\in D$
例10	满足[规则B.2]	满足[规则B.1]	级别$_{\text{感官动词句}}$=级别$_{\text{心理动词句}}$;	C_1为非中心分句,C_2为中心分句;C_1满足规
	$C_1\in F$	$C_2\in F$	同级先后原则;	则s2s,C_2不满足规则s1s;$C_1\in E$,$C_2\in D$

通过原B类规则的级别设定,使情况3(即$\text{subjection}(c_1,F)\wedge(\forall c)\neg\text{subjection}(c,A)\wedge(\exists c)\text{subjection}(c,F)$)的中心分句判定成为可能,进而也完成了此类汉语复句内在联系的判定.

3.2.2.6 中心分句判定算法

'条件①',即中心分句的存在判定核心算法,由下面两部分构成:

```
Part I:{对伴随关系复句中所有分句进行扫描}
    IF (分句 C_i 满足充分条件1中的子条件)   THEN C_i:=0;{非中心分句的级别号}
    IF (分句 C_i 满足'伴随关系复句规则')     THEN C_i:=条件②规则的等级号;
                                            ELSE C_i:=5;    {中心分句的级别号}
Part II:{判断一个分句 C_i 是否为非中心分句}
    IF (复句中每个分句的级别号相同)
      THEN IF ((满足级别判定'同一规则')∨(满足级别判定'同级共处原则'))
             THEN (分句 C_i 被判定为中心分句){即相应B类规则不能管辖分句 C_i}
             ELSE IF (分句 C_i 是最后一个分句)
                    THEN (分句 C_i 被判定为中心分句){即相应B类规则不能管辖分句 C_i}
                    ELSE (分句 C_i 被判定为非中心分句){即相应B类规则能够管辖分句 C_i}
```

ELSE IF (C_j的级别号最大)

THEN (分句C_j被判定为中心分句){即相应B规则不能管辖分句C_j}

ELSE (分句C_j被判定为非中心分句);{即相应B类规则能够管辖分句C_j}

依据中心分句与非中心分句的判定结果,B类规则能够管辖约束的分句,按规则翻译操作处理;不能管辖约束的分句,按一般单句作复句主干句处理.至此,通过中心分句与非中心分句的动态判定,解决B类规则局部性所带来的问题,进而实现了汉英机译中共时关系复句和状态关系复句的形式化处理工作.

4 实验系统与结论

为了验证本文中提取规则和算法的正确性与可行性,我们构造了一个基于LISP方言系统DCLISP平台的汉英翻译计算机实验系统.该系统主要由复句关系规则库、中心分句判定级别表和词典库(汉语语法词典、汉语语义词典和汉英对照词典)构成.系统以人工预处理的汉语格式模板为输入,通过复句关系规则库驱动,结合词典库和中心分句级别表完成汉语复句内在联系的自动分析和判定工作,并依据规则满足处理模块和一般分句处理模块构成的复句处理,实现英语译文的生成.

按照研究对象的界定要求,我们对426句复句进行测试,本系统均能在本文讨论范围内得到正确处理.但由于该系统正在发展之中,且汉英复句机译研究工作还处于起步阶段,无验证的标准语料库,故正确率尚难确定.

尽管对多个原子关系构成的复杂复句作了一定的限制,工作还需进一步完善,但该系统已充分表明了凭借完备的各级语义信息划分,实现面向篇章的汉英机器翻译的可能性.但同时工作也十分艰巨,其中最核心的问题就是如何构建一个完整的和可行的基于深层语法驱动的大型汉语语言知识库,并依靠它把具有语言分析本质意义的语义识别工作深入到汉英机器翻译的各个分析阶段.这也正是我们下一步的努力方向.

References:

[1] Li, Lin-ding. Chinese Verbs in Modern Times. Beijing: Chinese Social Science Publishing Company, 1990 (in Chinese).

[2] Chen, Ping. Referent introducing and tracking in Chinese narratives[Ph. D. Thesis]. California: University of California, 1986.

[3] Huang, C. T. J. Pro-Drop in Chinese: a generalized control approach [Ph. D. Thesis]. Ithaca, NY: Cornell University, 1986.

[4] Feng, Zhi-wei. New Review of Machine Translation. Beijing: Chinese Publishing Company, 1994 (in Chinese).

[5] Song, Rou. The delesion of the fronts of clauses in Chinese narratives. Journal of Chinese Information Processing, 1992,6 (3):62~68 (in Chinese).

[6] Strube, M. Processing complex sentences in the centering framework. In: Hajicova, Eva, ed. Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics. Santa Cruz: Association for Computational Linguistics, 1996. 34~41.

[7] Hideo, Watanabe, Koichi, Takeda. A pattern-based machine translation system extended by example-based processing. In: Martin, Kay, ed. Proceedings of the 17th Conference on Computational Linguistics. New Brunswick: Association for Computational Linguistics, 1998. 1369~1374.

[8] Takehiko, Yoshimi, Toshiyuki, Okunishi, Takahiro, Yamaji, *et al*. Evaluation of importance of sentences based on connectivity to title. In: Kay, Martin, ed. Proceedings of the 17th Conference on Computational Linguistics. New Brunswick: Association for Computational Linguistics, 1998. 1443~1447.

[9] Lu, Song. The approaches of processing the appearance of the zero anaphoric subjects in Chinese-English machine translation [MS. Thesis]. Beijing: Computer Institute of Beijing Polytechnic University, 1998.

附中文参考文献:

[1] 李临定.现代汉语动词.北京:中国社会科学出版社,1990.

[4] 冯志伟. 自然语言机器翻译新论. 北京:语文出版社,1994.
[5] 宋柔. 汉语叙述文中的小句前部省略现象初析. 中文信息学报,1992,6(3):62~68.
[9] 鲁松. 汉英机器翻译中主语省略现象的处理方法[硕士学位论文]. 北京:北京工业大学,1998.

Distinction and Treatment of the Internal Relation of Descriptive Complex Sentences in Chinese-English Machine Translation

LU Song[1], SONG Rou[2]

[1](*Institute of Computing Sciences and Technology, The Chinese Academy of Sciences, Beijing* 100080, *China*);
[2](*Computer Institute, Beijing Polytechnic University, Beijing* 100044, *China*)
E-mail: lusong@ict.ac.cn

Received January 13, 1999; accepted April 19, 2000

Abstract: In Chinese-English machine translation, the treatment of a complex sentence must not only depend on the individual clauses, but also review deeply the internal relation among the clauses in the complex sentence. In Chinese, there are lots of descriptive sentences with non-coordinating relation without certain language marks. To distinguish the internal relation among them and get the right English output from Chinese-English Machine Translation System, the integrated rules to distinguishing them are presented, and the method of judging dynamically the main clauses is put forward to solve the problem of localization of some rules. Finally the rules are verified in virtue of the experimental system.

Key words: Chinese-English machine translation; descriptive complex sentence; distinguish the internal relation; dynamic judgement on the main clauses