

用句节间系受关系的日语分析

王启祥 王锡江 雍殿书 陈家骏

(南京大学计算机科学与技术系)

摘 要

本文介绍用句节间关系对日语作表层结构分析的方法,通过得到的句节间系受分析树可以转换生成为汉语,而且采用各阶段允许人工介入,消除有歧义的结果。用此方法不仅积累了语言知识,还得到较好的译文,与生成部分结合可以构成一个会话方式的日汉翻译系统。

一、前言

机器翻译实用化的前景已无用置疑,而现今推出的全自动翻译系统内部都存在着不同程度的问题,有待进一步解决。为了进行研究,我们开发了一个允许用户在翻译处理各阶段人工介入的对话式辅助翻译系统。其目的是通过用户的使用,在交互过程中使问题显露积累语言处理的知识,找出解决的方法。为进一步改进系统,提供知识,使机译系统日臻完善,以获得高质量的翻译结果。

我们研究的是人工辅助的日汉翻译系统,利用日语句节间关系对日语句法结构及表层的语义分析。得到句节间系受关系的树形结构,经转换生成得到汉语。本文介绍句节间系受关系的分析方法。

二、日语的句节

自然语言的句子是由概念和概念间的关系、发话者的态度视点等构成的,不同的语言对这些概念、关系、样态采用不同的具体表现形式,生成不同的语言的表层结构。对日语而言:

(1) 表示概念的叫独立词,如名、动、形、形动、副词等;

(2) 表示独立词之间关系的词指名词与动词间的格助词、系助词,动词(或句子)和动词(或句子)间关系的接续助词,名词与名词间的连体格助词,及用言和助动词的活用型所表示的关系;

(3) 对独立词(主要是名词)添加信息的有系助词、副助词、准体助词、接头词和接尾词等;

(4) 与动词连接表示发话者样态的是助动词。

由这些构成的日语句子,可以利用词性及其变化切分成句节,所以日语的句子可在看作是句节的有限序列。

〈日本语句〉 = 〈句节〉 〈句节〉 … 〈句节〉

〈句节〉 = {接头词} 〈独立词〉 {接尾词} {附属词} •

〈 〉 表示一定有

{ } 表示可有可无

• 表示可在0~10以内重复

由句节构成句,每个句节内可以由接头词、独立词、接尾词和附属词构成。一个句节中一定存在一个独立词,对独立词添加意思的接头词、接尾词,表示句节间关系或表示样态的附属词是可以有也可以没有。

例: 古い车を买いました

この本はフランス语の本です

明天私は学校へ行く

例句中,古い、この、明天,没有附属词在日语中也算作句节,日语的句节是日语的重要句法特征。

三. 句节间的系受关系 (Kakeriuke Relation of two Bunsetau)

日语句子由若干个句节组成, 一个句子若有二个以上句节组成时, 前一个句节总以某一种关系关联到后面一个句节, 这种关系日语里称为'系'的关系(即与后面一个句节有联系), 后面一个句节总是受到前面句节的关联, 这种关系称为'受'的关系, 两方面结合起来就是系受关系。(係り受け) 句子中第一个句节只有对后面句节的系的关系, 句子中最后一个句节只对前面句节有受的关系, 句子中间的句节既承受前面句节的关联, 又必然关联到后面的句节, 这种句节就起到系受两方面的作用。特别应指出日语中系受关系不能简单的理解为修饰关系, 因为它还包含着主、谓等其他关系。

例如 桜の花が咲いた

桜のは句首的句节, 只有对后面系的关系, 咲いた是句末的句节, 只有受的关系。花が这一句节花既受桜の的修饰, 又因花が的作为主语句节, 关联到咲いた这个谓语句节。一个句节受两方面的作用, 可以根据该句节的独立词和附属词的构造、形式加以判断, 即一个句节与前面句节(不一定指紧前一个句节)的关系取决于该句节的独立词; 该句节与后面的句节构成什么系受关系取决于该句节的最后的附属词。用言的连体型、连用型以及副词、连体词、感叹词、接续词之类可以不必借助附属词也可以明确表示出与后面句节的系受关系, 对用言有助动词时也同样可以定出其关系。

四. 用句节间系受关系分析日语语法结构

我们将一个日语句子看成由若干个句节组成。如果分析出句节间的系受关系, 实际上就理解了句子的表层结构。

例如 私は 8時に 学校に 行く

私は句节因は是提示助词与行く有系受关系, 其他句节也是与行く有关系:

即 私は 行く
8時 に 行く
学校 に 行く

这样该句实际上给出了: (谁)が (何时)に (场所)に 行く

同样 私は 7時の 电車で 行きます

而且日语中句节系受关系有如下性质:

- (1) 句子中各句节除句尾的句节, 在句尾前总会与某一个句节存在系的关系;
- (2) 句中各句节除句头的句节, 总存在着0个以上的受的关系;
- (3) 连体、连用型具有就近系受(修饰)的原则;
- (4) 同一句中句节间的系受关系不相交原则。

由上可见, 根据词典内提供的词性, 邻接信息可以将句子切分为句节, 再根据独立词、附属词等日语语法特征即能求得句节间的系受关系。

日语中句节间的系受关系, 通常表示表层构造的关系, 可分类如下:

(1) 体言与用言间的系受关系

如	动作主体1	が	花が咲く
	动作主体2	は	私は行く
	时间的起始	から	朝から晩まで動く
	时间的终点	まで	朝から晩まで遊ぶ

时间	NIL	5 年间
时间	まで	5 时で归る
场所的起始	から	京都から东京 まで行く
场所的终点	まで	学校から家 まで歩く
动作的场所	で	工場で作る
动作的同伴	と	彼と归る
比较	より	彼より速い
数量	NIL	5グラム入れる

(2) 体言和体言的系受关系

数量	の	5グラムの水
用连结的体言	の	太郎の
并列	と や か	太郎と(や, か)花子

(3) 用言和用言的系受关系

连用接续	ご飯を食べて行きます 花が開き, 実を結んだ
连用并列	計算機は四則运算を行ったり 方程式を解いたりすることができる

(4) 用言和体言的系受关系

连体形	黄色い结晶 稀金属を含む特殊鋼
-----	--------------------

(5) 副词、接续词和用言、体言、副词间的系受关系

副词与用言	ゆっくり歩く
副词与体言	この合金はかなりの耐蚀性をもっている
副词与副词	もつとゆっくり歩く
接续词与用言	しかし行かなかった

系受关系的分类可粗可细,粗分仅十几类,而细分可达六、七十类;分得越细,生成时越方便,生成的译文质量也越高。

五. 用句节间系受关系分析日语句子

为了说明方法,今举例如下,对一给定的日语句子S,切分为句节后,从首句节开始顺序编号;

S: 私は 山に 登って 景色を 見た

1 2 3 4 5

$$1. \Rightarrow i \Rightarrow j (K(i, j)) \rightarrow K\{(i, j)\} \quad i, j=1, 2, \dots, 5$$

即求各句节两间有系受关系的对 (i, j) , 若句节 i 与句节 j 有系受关系, 将这些有系受关系的对作元素组集合 K

$$K = \begin{pmatrix} (1, 1) & (1, 3) & (1, 5) \\ (2, 2) & (2, 3) & (2, 5) \\ (3, 3) & (3, 5) \\ (4, 4) & (4, 5) \\ (5, 5) \end{pmatrix}$$

该例中当 $i=1$ 时, 句节 $私は$ 只能与 $j=3$ 登って, $j=5$ 見た有主语的系受关系; 当 $i=2$ 时句节 $山に$ 可与 $j=3, j=5$ 有に格系受关系, 其他类推。这里把 $i=j, i=1 \dots 5$ 也作为 K 的元素, 是从仅有一个句节的句子来, 放入不矛盾, 而且对今后的讨论和表达反而方便。把该句所有可能的系受关系 (i, j) , 也可以用矩阵 $M(S)$ 表示, a_{ij} 表示 i 句节与 j 句节有无系受关系, 有系受关系为 $a_{ij}=1$, 无系受关系为 $a_{ij}=0$ 。

$$M(S) = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad M(S) \text{ 称为系受矩阵}$$

任何一个句节只有自身和该句节句末一侧的某一个句节有系的关系, 因为日语的这个性质, 系受矩的对角线以下的元素均为 0。

2. 系受关系不相交的定义

在日语中系受关系不允许相交, 对系受矩阵中两系受关系不相交的判定原则定义如下:

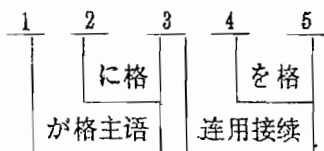
$$\text{在 } M(S) = [a_{ij}] \quad i, j \approx 1, 2, \dots, n$$

若 $a_{ij}=a_{kl}=1$ 不相交应满足下列条件之一

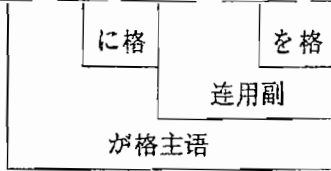
- (1) $i < j \quad K < L \quad i < K$ 而且 $K < L < j$
- (2) $i < j \quad K < L \quad i < K$ 而且 $j \leq K$

删除相交的系受关系, 并不是说就能得到该句子的唯一的系受关系, 与其他分析方法一样, 系受关系分析也存在歧义问题, 就本例而言, 它存在三个系受关系分析的结果。

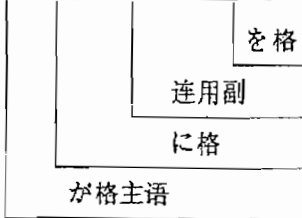
(1) 私は 山に 登って 景色を 見た



(2) 私は 山に 登って 景色を 見た



(3) 私は 山に 登って 景色を 見た



上述的系受关系分析根据系受关系分类可以求出一个句子中那个句节与那个句节有什么样的关系，有了这些表层格的格关系和连体、连用接续、连用副(连用修饰)等就能理解句子的表层及部分深层结构的含义，分析过程中可以通过得到的M(S) 求出歧义结构的个数和各结构中的系受关系，有些句子的歧义结构可以很多甚至成百成千，对这种歧义结构利用消歧规则也难以得到唯一的结果，从效率及翻译质量出发允许人工介入，用对话方式选择一个结果进一步分析与生成。

六. 结束语

利用日语的句节与系受关系分析对日语进行句法结构的分析，在必要时人工介入，用会话方式消除歧义结构，能够达到积累信息，为改进分析系统提供知识的目的。用此分析结果生成的汉语时也得可得较好的结果，在分析阶段排除了分析不良的结果，在分析器还不十分成熟的阶段，是一种积累知识、提供改进的手段。

参考文献

1. 王启祥, 王锡江等, 日汉机器翻译中词的自动切分技术, 《中文信息学报》 Vol. 2 No. 3 1988
2. 池田尚志, 日本語の構文意味解析システム, 日本电总研技报871号 1986
3. 荻野紫穂, 日英机械翻译システムJETSにおける日本語解析,
《日本自然语言处理》 Vol. 91 No. 67 1991

Syntax Analysis of Japanese Sentences Based on Kakariuke Relations Between Bunsetsu's

Wang Qixiang, Wang Xijiang, Yong Dianshu, Chen Jiajun

Nanjing University, Nanjing, 210008

Abstract

In this paper, we present a method of analyzing the surface structures of Japanese sentences. Using the kakariuke relations between Bunsetsu's in a Japanese sentence, we can generate a parsing tree reflecting these relations. Then, this tree can be used to generate the corresponding Chinese sentence. In the process, we permit user intervention to eliminate the ambiguities. By using this method, we can generate pretty sentences as well as accumulating the knowledge about languages. It can be combined with Chinese generation to make an interactive Japanese-Chinese machine translation system.