一个汉英机器翻译系统中的语义处理框架*

常宝宝* 詹卫东**

*北京大学计算语言学研究所, 北京, 100871
**北京大学中文系，北京, 100871

摘要: 在汉英机器翻译中，由于汉语本身的一些特点以及汉语和英语的巨大差异，因而语义的处理显得十分重要。本文介绍在我们研制的一个汉英机器翻译系统中所采用的语义处理框架，主要介绍我们在这个问题上的指导思想和我们的做法。

关键词: 机器翻译，语义处理，配价语法，格语法

A Framework of Semantic Processing in a Chinese-English Machine Translation System

Chang Baobao* Zhan Weidong**

*Institute of Computational Linguistics, Peking University
**Department of Chinese Language and Literature, Peking University
Beijing, 100871

Abstract: Because the characteristics of Chinese and the divergency between Chinese and English, Semantic processing is very important in Chinese-English machine translation.In this paper, we will introduce a framework of semantic processing adopted in a Chinese-English machine translation system which is being developed by us. some opinions, views and approaches on semantic processing which we take are also disscussed in this paper.

Keywords: Machine Translation, Semantic Processing, Valency Grammar, Case Grammar

一、前言

综观我国的机器翻译研究，尤其是英汉、汉英机译研究。从机器翻译的译文质量来看，英汉系统要远优于汉英系统[1] 。究其原因，一般认为，主要是由于汉语的特点造成的，一方面，汉语的形式标记不发达，分析难度高于英语，另一方面，由于英语形态丰富，较难生成，在汉语到英语的翻译过程中，一些汉语中缺乏的语言范畴在英语生成过程中难以处理。这些特点都必然决定了汉英机器翻译系统必须加强语义处理工作，汉英转换不仅要在语法层面上进行、也要在语义层面上进行。

语义处理至少有下面两个显著作用：(*1*)语义处理有助于得到句子正确的句法结构(排歧)。(*2*)语义分析所建立的句子的语义结构使得转换不仅可以在句法层次上进行，还可以在语义结构层次上进行。同时，在当前语义处理技术还远不成熟的背景下，构筑一个实用机器翻译系统，要求我们必须应用工程的观点看问题，既要吸收语义研究的新成果，也要考

* 本文工作得到了国家 863 计划的资助

虑它们的可行性及可操作性。本文介绍我们目前研制的一个汉英机器翻译系统* 中的语义处理框架。

应当特别指出，在我们目前的系统中，基于句法信息的处理仍然占据着中心位置，利用语义信息处理的主要目的是在系统中起辅助作用。

二、语义处理框架的设计

在语义处理领域，*Fillmore* 的格语法理论目前已获得广泛应用[2] ，在格语法中，主要描述了中心动词和句中与之共现的名词性成分之间的语义关系，但对名词与名词、名词与形容词之间的语义关系则较少涉及，这同机器翻译对语义处理的要求相比还是不够的。由法国语言学家特思尼耶尔创立的配价语法近年来在汉语学界得到了很多的讨论，并有所拓展[3]， 尤其是学者们对于汉语名词和形容词的配价研究[4] [5]。

从研究思路上说，配价语法、格语法有很强的一致性。二者均把句法结构中的主要成分区分为支配成分和从属成分。但二者在研究内容上各有侧重，并且配价语法除对动词的配价关系进行探讨以外，还对名词、形容词的配价关系进行了有益的探讨。可见，综合运用格语法、配价语法的理论，将句法结构中的主要成分区分为支配成分和从属成分，根据支配成分和从属成分之间的语义关系，由支配成分给从属成分指派语义角色，有助于建立面向机器翻译的系统的语义处理模式。

基于上述原则，我们首先建立了一个汉语词语的语义分类体系，然后在分类的基础上，描述中心动词与受其支配的名词、名词与受其支配的其它名词以及形容词之间的语义关系。

三、语义分类体系

在机器翻译中，词的语义分类是标明一个词的语义属性的常用手段，也是我们基于分类原则描述语义结构的基础，词语的语义分类的原则和标准应是受应用目标驱动的。以下简要介绍我们的分类原则和分类体系。

3.1 分类原则

我们在分类对象、分类的深度、广度以及分类的目的方面做了如下的规定：

(1)分类对象：语义分类建立在词的语法分类的基础之上，主要对汉语动词、名词和形容词三大类实词进行了语义分类。

(2)分类深度、广度：由于我们以动词为中心描述动词、名词之间的语义关系，以名词为中心描述名词和名词间的语义关系，因而我们对名词采用了较细的分类，对动词、形容词采用了相对较粗的分类。同时在分类中，由于抽象事物的分类标准不易确定，而且具体实践也证明，对抽象事物做过细分类也不利于实际操作，因而我们在分类中只对具体事物进行了层次较多、较细的分类，而对抽象事物采取相对较粗的分类。

(3)分类的目的：确定语义分类应该考虑分类的应用目标，分类应该从为句法分析、确定语义关系服务的角度确立分类原则和分类基础。

在具体确定一个词的归类时，允许一个词兼属多个语义类。

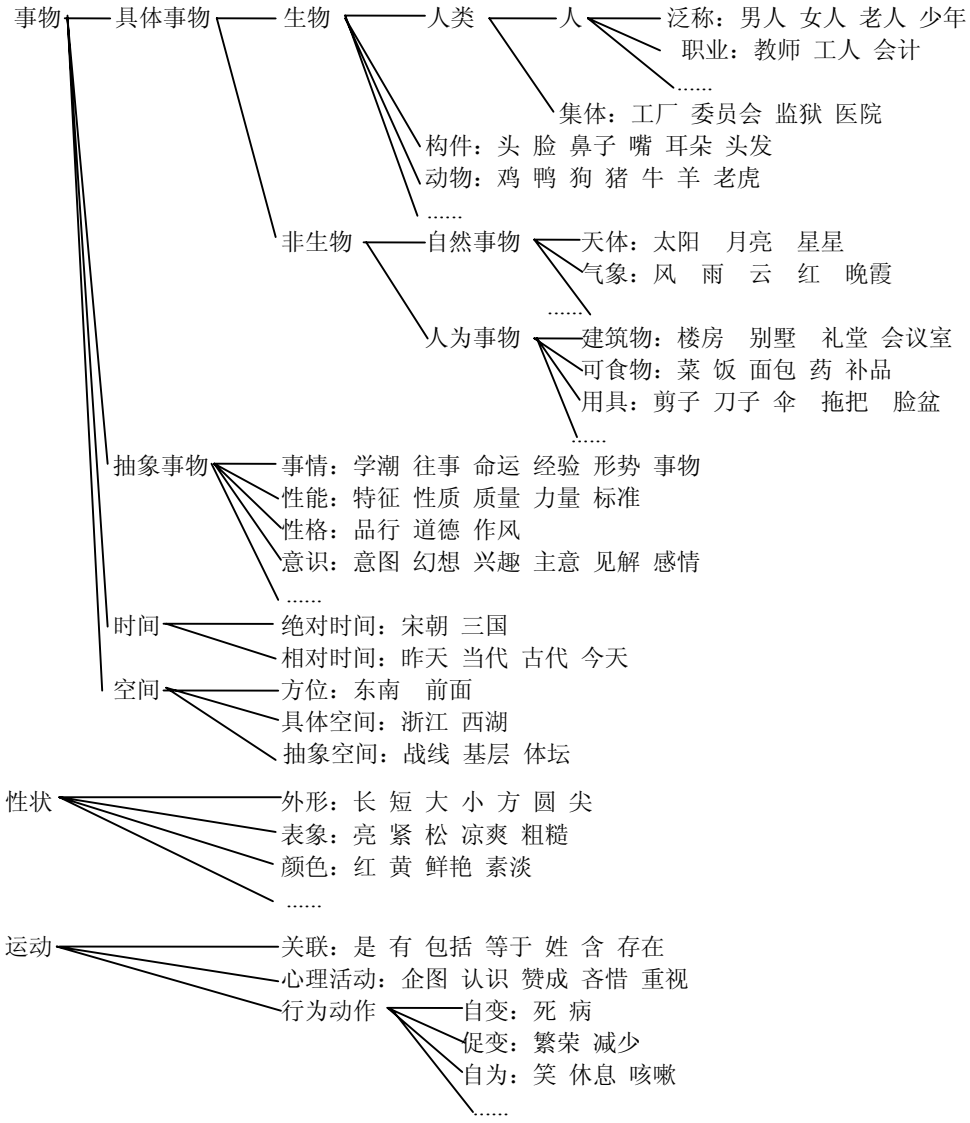
3.2 分类体系

我们将汉语词语总分为事物、性状与运动三个大类。其中事物类又分为具体事物、抽

---

* 中国科学院计算所二室和北京大学计算语言学研究所目前正在联合研制的一个汉英机器翻译系统

象事物、时间和空间等小类。下图为语义分类的一个概貌。



四、语义关系

有了上述语义分类，就可以以此为基础，刻画词与词之间的语义关系。下面介绍我们关于语义关系划分的一些做法。

4.1 原则

(1)描述对象：主要描述三大类实词之间的语义关系，描述名词和动词、名词和名词、形容词和名词之间的语义搭配关系。

(2)描述深度：在语义关系的分析实践中，异类词之间的语义关系的描述起来相对较易，同类词间语义关系描述起来要难一些。例如：动词与名词之间的语义关系易于描述，而动词和动词之间的语义关系就较难描述。为了便于操作，我们对异类词采用了较为深入的语义关系划分，同类词之间采用相对较浅的语义关系划分。

4.2 语义关系设置

我们综合应用配价语法和格语法的观点，对一个词语，描述它的配价数，以及该词对其各个配项的语义要求。

(1)配价数

用配价数来描述一个实词跟其他实词之间发生语义联系的能力。例如：

1)形容词“大”可以跟一个事物类的词发生语义关联。配价数为 1。表现为：

大树

形容词“热情”可以跟两个事物类的词发生语义关联。配价数为 2。表现为：

老王对我很热情

2)名词“儿子”可以跟一个事物类的词发生语义关联。配价数为 1。表现为：

老王的儿子

名词“意见”可以跟两个事物类的词发生语义关联。配价数为 2。表现为：

老王对你的意见

3)动词“咳嗽”可以跟一个事物类的词发生语义关联。配价数为 1。表现为：

老王咳嗽得厉害

动词“吃”可以跟两个事物类的词发生语义关联。配价数为 2。表现为：

孩子在吃苹果

动词“给”可以跟三个事物类的词发生语义关联。配价数为 3。表现为：

老师给了学生一本书

(2)词的语义框架描述

对于动词、名词和形容词的配价要求在词典中进行详细描述。主要有如下设置：

• 主体: 动作或性状的发出者或承当者；事物的参照者，例如：

他把书给我。　　动词“给”的主体是“他”，语义要求是“人”

花蔫了。　　形容词“蔫”的主体是“花”，主体语义类要求是“植物”

老王的妻子　　名词“妻子”的主体是“老王”，语义要求是“人”

• 客体: 动作或变化的影响者；事物的关联对象，例如：

老师正在指导学生。 动词“指导”客体是“学生”，客体语义要求是“人”

他对象棋的兴趣淡薄了。形容词“淡薄”的客体是“兴趣”，语义要求是“抽象事物”

老王对油画的兴趣　　名词“兴趣”的客体是“油画”，语义要求“领域”

• 与事: 事件中的受益者或受损者，例如：

他把书给我　　动词“给”的与事是“我”，与事语义要求是“人”

能跟动词发生语义联系的名词性成分还有时间、处所、方式等等多种情况。这里就不一一列举了。

五、分析实例

本节，我们用一个具体例子来说明上述语义处理机制在实际的翻译中所起的作用。考虑将下面的句子提交翻译系统进行翻译：

老王听取了朋友的意见。

句中出现的词语的词典内容如下，词条不仅描述了词语的语法属性[6] ，对于动词、有价名词还描述了其配价数以及各配项的语义要求。

```
$$ 朋友
    **{n} n $=[名词子类:na,数量名:是,个体量词:个|位|名,前名:是,前动:否,后名:是,名状语:否,临
时量词:否,语义类:人的社会属性|人,配价数:1]{主体:[语义类:人]}
        => N<friend> $=[NSUBC:NCONT,NMORF:REGU]
$$ 意见
```

**{n} n $=[名词子类:ne,数量名:是,个体量词:条,前名:是,前动:是,后名:是,名状语:否,临时量词:否,语义类:意识,配价数:2]{主体:[语义类:人],客体:[语义类:事物]}

=> N<opinion> $=[NSUBC:NCONT,NMORF:REGU]

$$ 听取

**{v} v $=[谓词性主语:否,系词:否,助动词:否,趋向动词:否,补助动词:否,形式动词:否,准谓宾:否,前名:否,后名:否,体谓准:体,双宾:否,兼语句:否,后动量词:是,后时量词:是,不:是,没:是,很:否,单作谓语:是,单作补语:否,动宾:否,小句宾:否,语义类:感知,配价数:2]{主体:[语义类:人],客体:[语义类:事理]}

=> V<heed> $=[VSUBC:VO,VMORF:REGU,OBJTYPE:NP,COMPTYPE:VG|NF]

$$ 王

**{g} g $=[语义类:姓名]

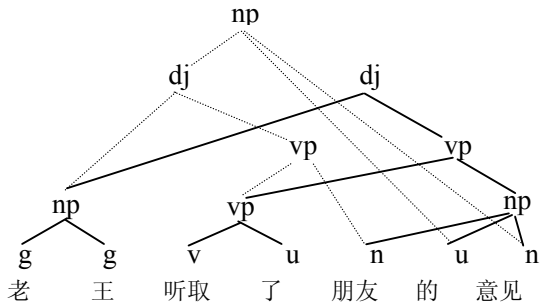
=> N<Wang>

$$ 老

**{a} a $=[形容词子类:aa,谓词性主语:否,准宾语:是,形谓语:是,形补语:粘|组,带补:是,形趋:是,形状语:是,准谓宾:否,有宾:否,不:是,很:是,前名:否,重叠词性:z,兼类:d,语义类:生理属性|性状]{主体:[语义类:具体物]}=> A<old> $=[ASUBC:ANORM,AMORF:SUFF]

**{g} g => N<Lao>

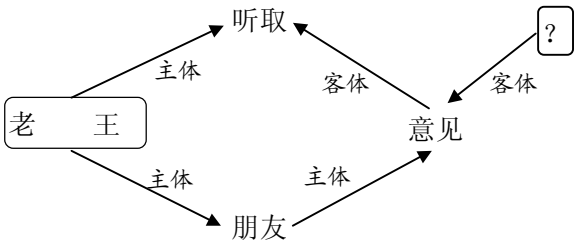
(1)由于语义信息的参与，我们就可以从下面有歧义的句法树中挑选出正确的句法树。（下图中实线标记的句子结构树是正确的）



这是因为，二价动词“听取”的客体语义要求是“事理”而不能是“人”，因而“听取了朋友”为不合法结构，而“听取了朋友的意见”为合法结构。

(2)分析得到的配价关系网络可以有效地指导转换，以生成更为自然的英语。

经过分析，可以得到下面的配价关系网络：



根据句中主要词汇之间的配价关系，该句可以翻译为“*Lao Wang heeded his/her friend's opinion*”，该译法比“*Lao Wang heeded friend's opinion*”要自然一些。

六、结束语

本文介绍了一个用于汉英机器翻译的汉语语义处理框架，主要是在汉语配价理论和格

语法的指导下，对动词、形容词以及名词三类主要实词进行基于语义分类的框架描写，使得机器能够分析出词语之间的语义关系，从而提高了译文质量。

可以看出，我们目前所引入的语义处理框架还比较简单，对词语之间语义关系的描述还不是很细。并且在确定词语之间的配价关系时，多数是描述处于同一句法层次上的两个直接成分(IC)之间的语义关系，对处于不同句法层次上的非直接成分之间的语义关系较少兼顾。经过实际测试和分析，我们发现该处理模式使译文质量有了明显改善，但同时也存在一定的局限性。[7] 在我们目前的系统中，基于句法层面的转换仍然占据中心位置。我们正在对这一体系进行改进，以期获得更好的翻译效果。

参考文献

[1] 段慧明，俞士汶，关于 *1995* 年度机器翻译评测的总结报告，《计算机世界报》评测版，*1996* 年 *3* 月 *25* 日

[2] 俞士汶，自然语言语义分析技术，中国计算机用户，*1988* 年第 *5* 期

[3] 沈阳，郑定欧主编，《现代汉语配价语法研究》，北京大学出版社，*1995*

[4] 袁毓林，一价名词的认知研究，中国语文，*1994* 年第 *4* 期，*pp241-253*

[5] 袁毓林，现代汉语名词的配价研究，中国社会科学，*1992* 年第 *3* 期

[6] 俞士汶等，现代汉语语法信息词典规格说明书，中文信息学报，*1996* 年第 *2* 期

[7] 詹卫东，刘群，语义分类在汉英机器翻译中的作用及其存在的问题，已投稿