

逻辑语义及其在机译中的应用

董振东

问题的提出

对逻辑语义进行探讨，是一个涉及语言本质的深刻的理论问题，这也是从语言学研究的视角对机器翻译将产生重大影响的实践问题。

从事翻译的研究，无论是人工翻译或是机器翻译，都必须对这样一个问题做出回答。即把一种语言转换成另一种语言，究竟是换什么？翻译包含两个相互联系、相互制约的过程：理解和表达，这是公认的。然而问题在于：理解，究竟理解的是什么？表达，究竟表达的是什么？机器翻译是否也必须对这些问题做出回答？还是它根本上就不同于人工翻译，只是象译码或数学公式转换似的符号对应？机器翻译是不是一种对人工翻译的模拟？两者在本质上的区别何在？我们认为所有这些重大的理论性问题，仍然有必要进行探讨和研究。这些问题没有在理论上和实践上做出科学的回答之前，我们的研究必然存在着较大的盲目性，也不可能制定一个长远的为最终实现机器翻译真正投产应用的有效规划。

就我们现有的研究水平，我们对上述种种问题的回答是：在本质上，机器翻译是对人工翻译的模拟。语言学——包括翻译学——的研究应科学地和精密地揭示和描写人在翻译自然语言中的全部过程，把这一工程加以提炼，以机器可接受的水平教给机器。这是解决机器翻译的基本途径。无论是人工翻译还是机器翻译，两种语言转换的基本内容归根结底是逻辑语义的问题。翻译工程中的理解（实际并不仅限于翻译过程），是理解源语言的逻辑语义，而表达，则是把同一逻辑语义用目标语言的表层结构加以表现，实际上这是生成目标语言的过程。我们感到机器翻译迄今为止所面临的重大关键问题——语义问题——的核心是逻辑语义的研究。我们相信：逻辑语义的研究，将会为机器翻译的研究开辟新的前景，从而取得新的突破。

关于逻辑语义和逻辑语义结构

逻辑语义，诸如施事、动作、受事、空间、材料、方式、新质等，并不是什么新的概念。无论是古典语言学派或现代语言学派都承认和提出过这个概念。他们在各种著作中也多根据需要利用这个概念。但是他们常常有一个共同的特点，这就是一种实用主义的态度，而不是系统地研究它，把它作为描写语言和揭示语言本质的一个重要方面对待它。语法学认为它缺乏形式的标志，容易导致虚玄臆断，而语义学又认为它似乎应该是语法关系。因此，对于逻辑语义的揭示不系统、不科学、不精密。

我们认为语言是一个无限的复杂的符号系统，使人们交际的工具。对于这样一个系统应从各种相互区别又相互关联的方面对它进行描写。逻辑语义则是其中的一个方面，而且是一个重要的、不可忽视的方面。

语言之所以能够成为交际的工具，是由于它的表义性，就根本上说，是表达逻辑语义。所谓理解了某一个语言片断，就是理解了该语言片断所传达的种种逻辑关系，即逻辑语义。当某人对我说“把窗子打开”这句话时，我之所以会去打开窗子，是因为我明确了下面各种关系：①对方命令我发出一个动作；②这动作是“打开”；③这个“打开”的动作的对象是“窗子”；④现时“窗子”是“关着”的。当一个人不理解某一个语言片断时，可以表现为多种因素如某个词不懂，不懂语言片断的语法，甚至是由于缺乏背

景知识等，但这都是属于表层的，归根结底是由于他不明确该语言片断所表达的各种逻辑关系，即逻辑语义。经过这番讨论之后，我们可以进一步探讨究竟什么是逻辑语义了。

1. 什么是逻辑语义

逻辑语义是一种语义关系。在特定的交际环境下，某一语言片断的各个基本单元之间必然存在着某种逻辑关系，这些逻辑关系称为逻辑语义。逻辑语义的集合称为逻辑语义结构。例如：

students read in the room.

此句逻辑语义为：

students ——施事

read ——动作

room ——空间

此句逻辑语义结构为：

施事——动作——空间

所谓逻辑关系，是由于在客观世界上，包括语言者（语言传输者和语言接受者）在内的各种事物、动作、现象等都存在着一定的逻辑关系。这种关系是客观的，是高度抽象概括的。这里必须指出，语言的逻辑语义结构所包含的逻辑关系并不一定也并不要求符合客观的逻辑关系，因为语言作为一种符号是客观的，但是表达的意义是主观的。不能用某一语言片断的逻辑关系是否符合客观的逻辑关系来判定该语言片断的正确或不正确，成立或不成立。一定要把语言本身的正确与否同语言所表达的思想的正确与否分开。例如：

太阳是一块长方形的金属。

这个句子的语言本身是正确的，它有正确的逻辑关系结构，即：施事——施事判断体。这个句子的语言所表达的思想是谬误的，正是因为这句子的语言的逻辑语义结构是正确的。假设句子改成下面这样：

的金属—太阳长方形块是。

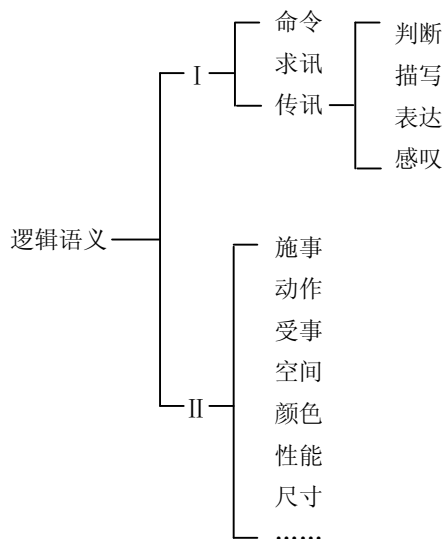
就无法判定它所表达的思想正确与否，因为句子的语言本身是不正确的，它没有正确的逻辑语义结构。只有具有正确的逻辑语义结构的语言，才能被人理解。

2. 逻辑语义的种类

每个语言片断（通常指句子）都蕴含三种关系：

- (1) 语言传输者和语言接受者之间的关系；
- (2) 语言者和语言片断内部其他实体之间的关系；
- (3) 语言内部除语言者外的各实体之间的关系。

我们把前两种关系归成一大类（I）；把最后一种关系列为另一大类（II）。这样，逻辑语义的种类可以列表如下：



绝大多数的逻辑语义项是为大家所熟悉的，不再赘述，但是，仍有几点似应稍加说明。

第一，确定某种逻辑语义时，语法结构和形式是重要的依据，但是又不能单凭结构和形式。以“求讯”为例，不能把疑问句笼统地归为“求讯”。这种误断，是由于没有把语法形式和逻辑语义加以区分的缘故。在语法上可以把一切以“？”结尾或具有其他疑问句特征（如汉语中句尾的“吗”，英语中主谓倒置，带疑问词等）的句子都定为疑问句，但是它们并非都表达“求讯”这一逻辑语义。例如：下面的句子是“命令”而不是“求讯”：

你给我把收音机关掉不行吗？

第二，属于第II类的逻辑语义，可以分为主要逻辑语义项和次要逻辑语义项。前者包括施事、动作和受事，除此均属于后者。

第三，在逻辑语义结构中，“动作”是核心。这一点对于逻辑语义结构及其层次的分析具有特殊重要的意义。

3. 逻辑语义的层次分析

在理论上，语言是无限的，每个句子都是一个可以无限扩展的符号链。这个扩展的最重要的特点在于它是一个依层次循环的过程，它的基础是逻辑语义，赖以实现的手段是语法结构。

层次对于逻辑语义结构是极端重要的。各种语言有其体现层次的结构或方式，但层次是都具备的。

在讨论逻辑语义结构及其层次的分析之前，必须先搞清楚几个重要的概念。

第一，通常，我们说对某一个语言片断进行逻辑语义结构及其层次的分析，是指一个句子而言的，即以一个句子为单位的。正是由于进行的是逻辑语义结构的分析，我们认为，才有可能打破以一个句子作分析的单位。而进行句法结构分析的时候，是很难突破以句子作为单位的。

第二，对一个句子进行逻辑语义结构及其层次分析的时候，基本的单元是词。但是必须说明，并不是句中的每一个词都承担逻辑语义，有的词仅仅是帮助相关的词表现逻辑语义，而自己并不承担任何逻辑语义，如汉语中的“着”、“的”、“所以”，英语中的介词、连词等。我们把能够承担逻辑语义的词叫做实体词；把没有能力承担逻辑语义的

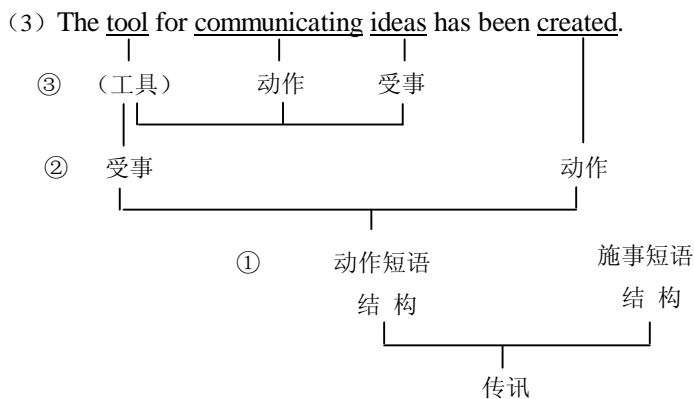
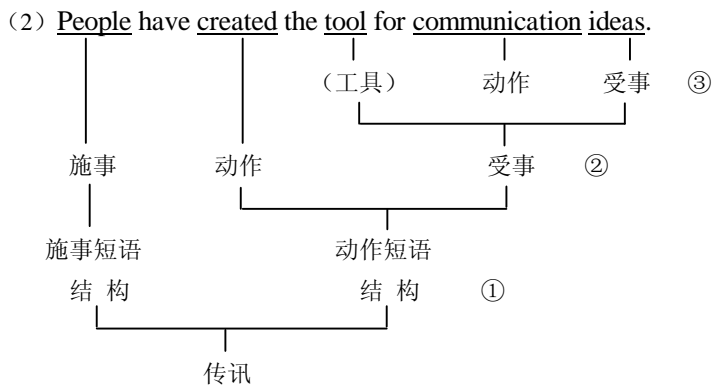
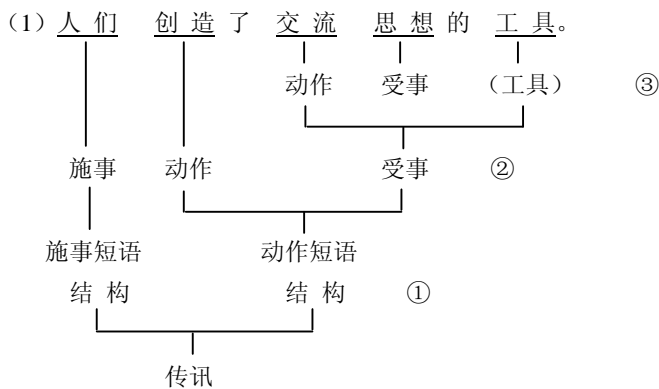
词叫做关系词。在逻辑语义结构的分析中，承担了逻辑语义的词或词组叫逻辑语义核；没有承担逻辑语义的词或词组叫非逻辑语义核。

第三，对一个句子或一个短语结构进行逻辑语义结构及其层次的分析，是求解这个句子或这个短语结构中各个逻辑语义核的逻辑语义及其所处的层次。

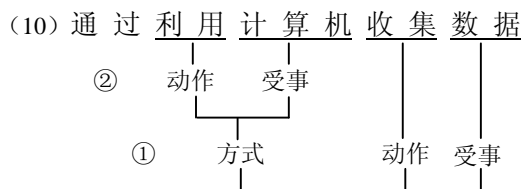
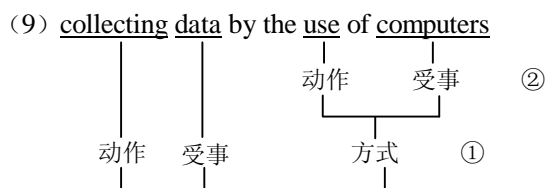
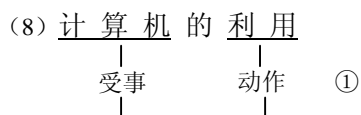
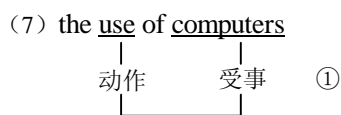
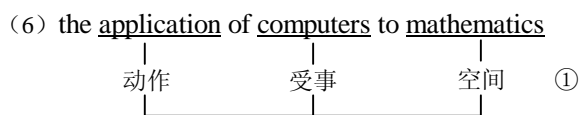
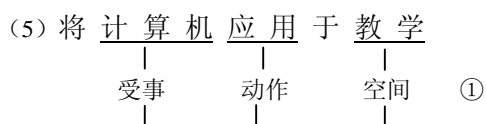
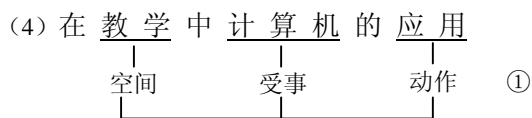
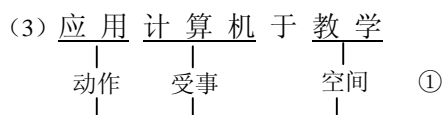
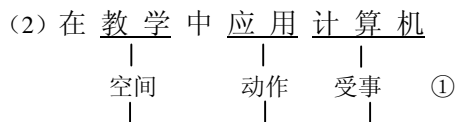
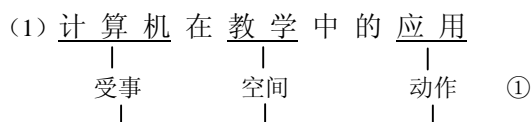
如前所述，“动作”是中心，不仅是句子或短语结构的中心，也是每一层次的中心。“动作”可具有双重逻辑语义，即它同时兼低层与高层的两个不同的逻辑语义。如果其他非“动作”的词也同时兼两层不同的逻辑语义时，可用低层消除法消除它在低层的逻辑语义（我们在图解中用括弧表示）。

这里还必须明确一点，我们进行的是逻辑语义结构分析，而不是句法结构分析，因此所谓“动作”并不等于“动词”，动词是语法的概念，而“动作”才是逻辑语义的概念。下面举例说明逻辑语义结构及其层次分析：

句子分析：



短语结构分析:



从以上的分析看: 对于逻辑语义结构, 重要的不在于结构内各逻辑语义的分布, 而在于它所包含的逻辑语义及其所处的层次。这也是逻辑语义结构同句法结构的重要区别点。

4. 逻辑语义与语法和词汇的关系

首先，逻辑语义和语法是两种不同的范畴，前者表义，是语言的内容，逻辑语义结构是一种深层结构。后者表形，是语言的形式，语法结构是表层结构。把逻辑语义同句法关系或称句子成分混为一谈，是一种常见的错误，这是有害的。

语法分析中，不考虑意义，而仅仅考虑形式，这是正确的。下面举一个在汉语语法中有争议的例子：

台上坐着主席团

在语法分析中，我们认为既可以是“主语—谓语—宾语”，也可以是“状语—谓语—主语”。只要语法上有规定，就可以成立。但是，这句话的逻辑语义结构只能是这样的：“空间—动作（表示存现）—存在体”。正因为如此，我们可以将上面一句话转换为：

台上是主席团

台上有主席团

三句话的内容是相同的，因为他们有着相同的逻辑语义结构，并且层次也相同。

其次，句法结构同逻辑语义结构的区别还在于它们并无一一对应的关系，如：

- (1) Impurity effects on sintering. 杂质对烧结的影响。
施事 动作
- (2) Micro wave absorption of hardened cement. 水泥石对微波的吸收。
受事 动作
- (3) SEM study of clinker. 用扫描电子显微镜研究水泥熟料。
方式 动作

相同的句法结构可以体现不同的逻辑语义结构，反之，不同的句法结构也可以体现相同的逻辑语义结构。

由此可见，绝对不能用句法分析代替逻辑语义分析。逻辑语义分析和判定不能单凭语法形式，它们在更大的程度上取决于词汇的语义因子。

最后，句法结构可能有二义性，这是说同一种句法结构可能体现两个或更多逻辑语义结构。但是，逻辑语义结构是没有二义性的。因为逻辑语义一方面受词汇的各种语义因子及其搭配的影响，一方面又必定处于特定交际环境中。逻辑语义结构一旦出现了二义性，只有两种可能，一是语言传输者特意而为的，一是不规范的语言。特定的交际环境是不局限于一个句子的范围的，因此，在正常的绝大多数情况下，二义性是不存在的，否则人们的交际将是不可想象的。我们要研究的语言是人们实际应用的真实的语言，而不是语法学家头脑中为了某种目的而造出来的只言片语。

逻辑语义在机译中的应用

由黑龙江大学和黑龙江应用教学研究所共同设计、联合中国科学技术情报研究所作试验的英汉机译规则系统的语言方案是以逻辑语义结构及其转换理论为基础的。从对译文质量的要求、英语词类的分类、句法结构分析和汉语生成的各个方面，都体现了关于逻辑语义的思想。关于这个系统的程序设计，语言方案和机器试验的详尽情况，我们将在实验报告中介绍，这里我们仅抽出几个重要的应用问题，加以讨论。

1. 逻辑语义对译文质量的作用

关于机译文的质量问题，是一个普遍关心的问题。总的来说，现有的机译文质量不甚令人满意。鉴于现有的语言研究的深度和计算机人工智能发展的高度都还有限，想在

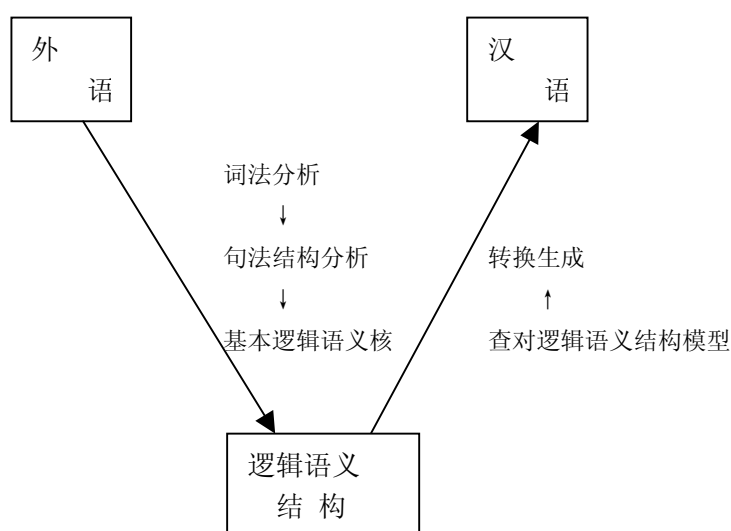
短时间内提高译文质量是困难的，达到人译水平更是不可能的。在研究中，由于过高地追求译文水平，将严重影响真正能够付诸实用的系统的设计和试验，而常常陷于顾此失彼，捉襟见肘的困境。如果采用大量的人工干预，进行译前编辑和译后加工，结构恐怕也是得不偿失，或是根本失去了机译的唯一优点——速度。

因此，我们有必要为机译文质量规定一个比较明确的标准。我们的意见是：这个标准就是逻辑语义及其层次的正确与否。在确定这一标准的基础上尽可能的改善译文质量。

我们认为机译文和其他的译文一样也应该有自己的风格和文体。

2. 逻辑语义时语言转换的基础

逻辑语义属于语言的深层结构的范畴。由一种语言翻译为另一种语言，将以逻辑语义不变为先决条件。逻辑语义是各种语言共通的。基于这样的理论，我们的机译系统语言方案的基本思想可以图示如下：



在机译中采用“外语→逻辑语义结构→汉语”的公式，实质上是提高计算机的“理解力”。我们认为这也是机译中实现人工智能的必要条件。

采用这个公式，才有可能建立一种语言方案与程序设计相对独立的机译规则系统。至少是后半部即逻辑语义结构→汉语可以做到语言方案与程序设计相对独立。

我们可以以逻辑语义结构为一种媒介语，实现多对一的转换生成。也就是说，逻辑语义结构→汉语这个部分可以用各种外语对汉语的翻译。例如：

- (1) [英]Effect of the mode on casting soundness.
模对铸件完整性的影响。
- (2) [英]Stain rate effects on the corrosion-resisting properties.
应变率对抗腐蚀性能的影响。
- (3) [德]Einfluß von Versteifungen auf Restfestigkeitsverhalten von Strukturen.
强化对结构的残余程度的影响。
- (4) [德]Einfluß des Spannung zustandes auf den Verlauf der R-Kurve.
应力状态对 R 曲线演变过程的影响。
- (5) [俄]Воздействие помех на фазовые системы.
干扰对构位系统的影响。

从上面的例子中，我们可以看到：各种语言在表达同一种逻辑语义结构式的表层结

构有所不同，甚至在同一种语言中也可以用不同的表层结构表示同一种逻辑语义结构。但只要我们能够求出它们是“施事—动作—受事”逻辑语义结构及所处的层次，就可以转换为汉语表层结构：

N1 对 N2 的 V

3. 逻辑语义是生成译文的依据

逻辑语义作为生成译文的依据，是基于它是语言转换的基础的。机器翻译毕竟不同于语言转换的理论研究，在转换中，或说是在生成译文中，实际的工作是调整词序以及进行必要的增删和润饰。因此在这个意义上说，逻辑语义就是调整词序的依据，同时又是提高译文质量的关键之一。这里是举三种情况加以说明：

第一种情况，关于英语被动结构的翻译。英语的被动结构的使用频率很高，尤其是在科技作品中，而在文摘中则比例更高。这里列举颇为常见的句型：“受事—动作”型，如：

(1) A new approach was reported.

这里报道了一种新的方法。

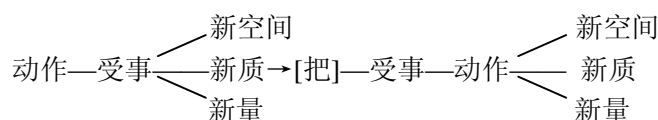
(2) The results have been examined.

这里检查了结果。

根据我们调查，这样一种结构在汉译时，绝大多数采用了以主动代替被动的办法，即改“受事—动作”为“动作—受事”。我们的规则系统也因此这样处理。

为什么人译中会建立这样一种对应模型呢？这里涉及到逻辑语义重心。在“受事—动作”的结构中，逻辑语义重心落到了“动作”上。在汉语中则以“动作—受事”来体现逻辑语义重心为“动作”的状态。与此类似的还有“受事—动作—目的”和“受事—动作—方式”。

第二种情况，关于类似英译时动词词义前加“把”的问题。英语动词中不存在与汉语的“把”——对应的词，我们在汉译时不得不根据需要予以增添。问题是这一增添的根据是什么呢？我们仍然是应用逻辑语义来决定增添与否的。这一转换在我们的规则系统中是这样的：



这样的转换模型可以保证任何一个应加“把”的情况都可以加上“把”，同时又可保证在该动词处于被动结构时不应加“把”的情况也得到正确的处理，因为后置受事和所列举的逻辑语义项是加“把”的依据和条件。

同样的，我们利用逻辑语义可以方便而正确地解决下面的情况：

(1) We combine retrieval with MT.

我们把检索与机译结合。

(2) Retrieval is combined with MT.

检索与机译结合。

在我们的系统中建立的模型是：

“动作—受事—关联体”→[把]—受事—关联体—动作。这里后置受事和“关联体”是依据和条件。这样，我们就没有必要给英语中的某些动词的汉义中，加“把”的信息，也可以使被动结构的处理大大简化。

第三种情况，关于译文的词序排列。词序排列的准确性，不仅影响译文的通顺与否，更重要的是可能会引起逻辑语义结构的破坏，进而影响译文的正确性。例如：

Chomsky's treatment of meanings

这个英语短语的翻译如果用传统的方法，即以为后置定语置于前置定语之前，是肯定行不通的。用什么方法来决定这类问题呢？只有利用逻辑语义，既准确也方便。判定这是短语层次上的“施事—动作—受事”，那么它可以同样转化成汉语的表层结构：

N1 对 N2 的 V

与第二节中的例子相同（“对”这个词，是汉语用以提前受事的手段，一般用于短语层）。又如：

As previously pointed out in Chap.5...

这个句子在机译中有两种可能性：

(1) 正如以前在第 5 章中指出的……

(2) 正如在第 5 章中以前指出的……

正确的译文应为前者，这是无疑的。用什么可以保证这一点呢？我们仍然是利用逻辑语义，建立与动作相关的逻辑语义项的排列顺序，即最大逻辑语义链。与上例类同的如：

As repeatedly pointed out in the book...

可以肯定译为：

正如在书中反复指出的……

因为“动作”的前附逻辑语义项在逻辑语义链中的排列是：时间—空间—……—方式—动作

我们确立的汉语的最大逻辑语义链是 MT-H78 系统的汉语生成的基本依据。由于 MT-H78 系统主要是用于英汉全文机器翻译的，但是同时可兼承英汉题录机器翻译，因此同时系统附有题录汉译——生成汉语题录的模型。题录的逻辑语义模型是以最大逻辑语义链为基础的。只是它特别强调由于层次的不同和问题的不同所产生的差异。试以我们前面曾经举过的两条题录为例：

(1) the use of computers

计算机的应用

(2) Collecting data by the use of computers

通过应用计算机收集数据

这里例（1）中和例（2）中划——部分的逻辑语义同为“动作—受事”，为什么汉语的表层结构不同呢？再看例（2）中划~~~~~部分，其逻辑语义也与例（1）中的相同，为什么汉语的表层结构又不同呢？关键在于①逻辑语义结构不同；②同样的逻辑语义结构所处的层次不同。例（2）中划~~~~~部分的逻辑语义结构与例（1）相同，但是整个逻辑语义结构是不同的，即：

(1) 动作—受事

(2) 动作—受事—方式

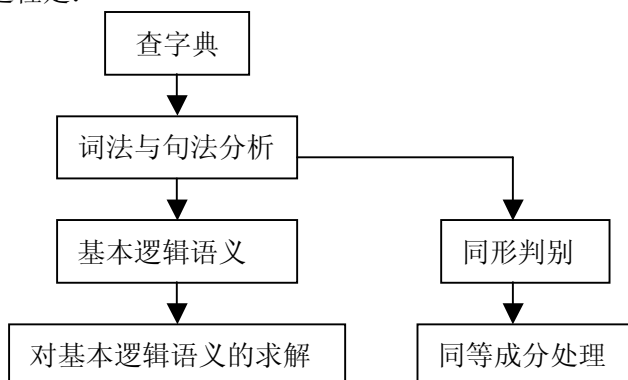
例（1）中划——部分的逻辑语义结构与例（2）中划——部分的逻辑语义结构完全相同，但是例（1）中的处于①层次，而例（2）中的处于②层次。

根据上述原理，我们可以推导出的一套相当完整的题录机译逻辑语义结构模型。这里要特别说明一点：并不是每一种题录逻辑语义链就是一种逻辑语义模型，完全没有必要这样做。我们只是根据动态逻辑语义核的多寡及其所处的层次建立一套以最大逻辑语义链为基础的有限的几个模型，便可以在极广的范围内适应题录的机器翻译。

4. 逻辑语义在机译中的求解

在机译系统中，逻辑语义的求解是一个关键性的实践问题。如果不能使机器正确地、顺利地求解逻辑语义，那么关于逻辑语义的理论只能是纸上谈兵了。

我们的实践证明，求解逻辑语义在机译中是可以做到的。MT-H78 系统求解逻辑语义的基本过程是：



我们求解的逻辑语义是基本逻辑语义。首先是对句子进行语法分析，将一切非基本逻辑语义放过（包括诸如冠词等非逻辑语义核）暴露出基本逻辑语义框架，然后采用以动态逻辑语义核为中心，依层次递归的循环判定过程求解这一框架中的各逻辑语义。下面的句子：

The Russian-English mathematical glossary and text tapes have been corrected and updated.

经过句法分析后的基本逻辑语义框架是：

tapes updated

这一框架中的逻辑语义是：

tapes updated
受事 动作

根据汉语生成逻辑语义模型，汉译文字先是：

改进 磁带

复原曾放过的逻辑语义核及非逻辑语义核：

曾经修改过和改进俄英教学词汇表和原文磁带。

再例如，下面的句子：

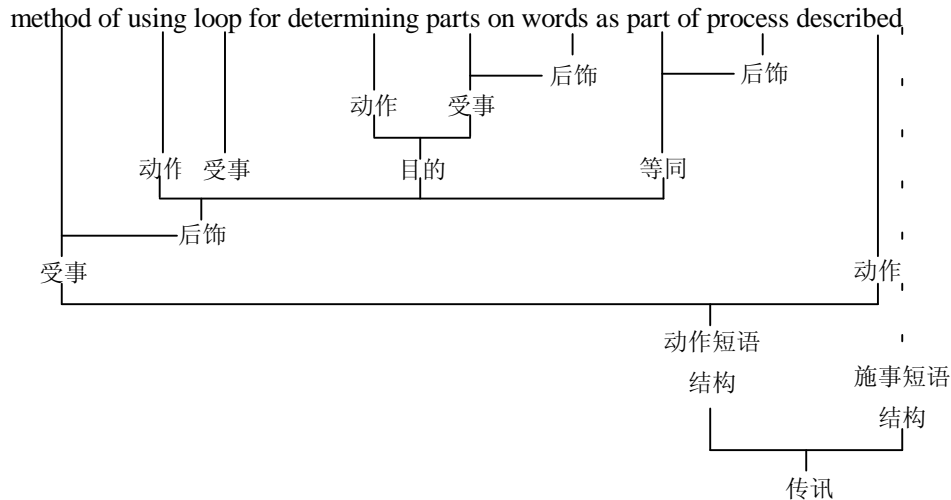
A method of using a feedback loop for determining parts of speech on multi-parts-of-speech words as part of the machine translation process is described.

经过句法分析后的基本逻辑语义框架是：

method of using loop for determining parts on words as part of process described

这一框架的逻辑语义结构和层次如图所示。

求解逻辑语义一个十分重要的前提是对英语词汇进行科学语义分类，即为英语词汇提供足够的语义信息。为此，我们必须对英语词汇进行语义分解，研究语义因子的搭配关系。就英语来说，逻辑语义主要是通过词序和语义的抽象框架来体现的。这些抽象的逻辑语义框架为逻辑语义规定了确定的解。



附：MT-H78 试验结果（选录 11 条）

(001) A TRANSLATOR IS A PROGRAM WHICH TRANSLATES A SOURCE PROGRAM INTO AN EQUIVALENT OBJECT PROGRAM.

翻译程序是把源程序翻译成等价的结果程序的程序。

(003) THE EXECUTION OF THE TRANSLATOR ITSELF OCCURS AT TRANSLATION TIME.

执行翻译程序本身在翻译时间发生。

(011) THE USUAL METHOD IS TO PROGRAM THE INTERPRETER IN TWO PHASES.

通常的方法是要按二遍编写解释程序。

(012) THE FIRST ANALYZES THE COMPLETE SOURCE PROGRAM, MUCH THE WAY A COMPLIER DOES, AND TRANSLATES IT INTO AN INTERNAL FORM.

第一分析完全的源程序，很类似于编译程序做，以及把它翻译成内部形式。

(016) THIS WILL ADD COMPARATIVELY LITTLE TO THE BOOK SINCE MOST TECHNIQUES USED IN COMPILER CONSTRUCTION ARE ALSO USED IN WRITING INTERPRETERS.

这将较为少地添加至书由于所用于编译程序构造的大多数技术所用于也是编写解释程序。

(025) THESE SYMBOLS ARE THEN PASSED ON TO THE ACTUAL ANALYZER.

这些符号传递至真正的分析程序。

(032) THE FIRST STEP IN THE METHOD LIMITS THE NUMBER OF PARTS OF SPEECH THE WORDS CAN TAKE FROM THEIR CONTEXTS.

方法中的第一步骤限制词汇能够从它们上下文取出的词类的数量。

(037) PROGRAMMERS AND DATA STRUCTURE DESIGNERS ARE OFTER FORCED TO CHOOSE BETWEEN ALTERNATIVE STRUCTURES.

程序员和数据结构设计者经常不得不在可供选择的结构之间选择。

(039) HOW MUCH DEBUGGING FACILITIES SHOULD THE OBJECT PROGRAM PROVIDE.

结果程序应该提供多少调试功能。

(043) WHEN INCONSISTENCIES ARISE AT THIS STAGE OF SENTENCE-STRUCTURE ANALYSIS, THE NEXT LIKELY CHOICES ARE SELECTED FROM THE MULTI-PART-OF-SPEECH WORD COMBINATIONS, AND A SIMILAR PROCESS, IS CARRIED OUT.

当矛盾在句子结构分析的这段出现时，下一个可能的选择从多词性词汇组合选择，以及实现类似的过程。

(045) THE COMBINATORIAL PROBLEM OF STORING ARRAYS AS VARIOUS KINDS OF LIST STRUCTURES IS EXAMINED.

检查作为各种各类的表结构的存放的数组的结合的问题。

中国的机器翻译 刘涌泉编 知识出版社 1984, pp. 25-45