心 理 学 报
第3期 ACTA PSYCHOLOGICA SINICA 1988年

机器理解古汉语的实验和若干问题[1)]

陈永明 李家治 蔡 山

中国科学院 心理研究所

摘 要

本文介绍了机器理解古汉语的一个实验系统(ACLUS)。该系统由一个知识库和三个子程序组成。知识库包含有一部机器词典和若干集用来加工句子的规则。三个子程序是:句法分析子程序、翻译子程序和问答子程序,它们分别执行三种不同的功能。本文还讨论了有关机器理解古汉语的若干重要问题。

一、前 言

两年前,我们曾对机器理解古汉语的问题进行了初步探索[1]。在这个基础上,我们又调整了程序的结构,扩大了理解的范围,对机器理解古汉语的某些问题作了进一步的探讨。我们希望,通过计算机理解古汉语的研究,可以达到以下几个目的。

1. 使机器表现出较为通顺地理解一些古汉语小故事的能力;

2. 使机器具有通过人机对话回答有限范围问题的能力,如有关句法、语义或情节的问题,以便于我们了解设计古汉语的计算机辅助教学系统的可行性;

3. 加深对理解现代汉语的研究,考察汉语的理解与思维、推理的关系,为制订适合计算机的汉语语法积累资料。

因而,从总体上看,对机器理解古汉语的研究,既有它的文化价值,也有它的实用价值。当然,本系统是一个实验性的,而不是一个实用性的系统。

二、系统的结构及功能

这个古汉语理解系统(简称ACLUS)以十个古汉语小故事作为计算机理解的对象。

1) 本文于1987年11月2日收到。

整个系统分成四块：知识库和三个子程序(图1)。知识库是系统处理古汉语的基础；三个子程序分别执行三种不同的功能，即句法分析、翻译和人机对话。

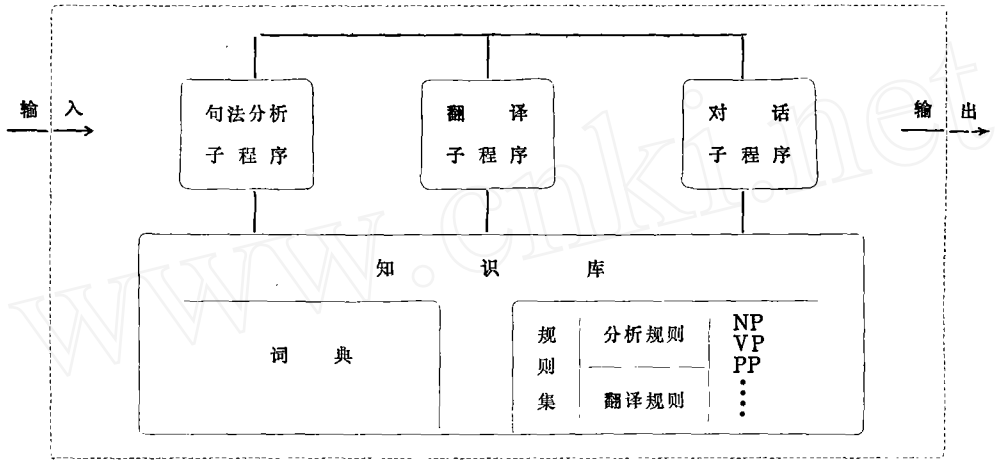

图1 系统的总体结构

(一)知识库：

知识库包括两部分内容：一部机器用的小词典以及各种规则集。

在机器词典中，包含了所有小故事中出现的词。在每一词条中，列出了该词的句法特征、语义特征，以及动词和名词之间的语义搭配信息和相应的现代汉语译词；有些词在不同语境下有不同的释义，故在词条中也列出了几种不同的译词，供机器翻译时选取。在某些名词的词条中，也列出了翻译时应该加上去的量词。此外，有少数词(主要是虚词)的用法比较特殊，在这类词的词条中，存有一个特殊处理的函数名。机器在查阅词典时，若遇到这类函数名，就直接访问该名称所包含的规则。这样，就减少了搜索的时间。

规则分为分析规则和翻译规则两类，每类中又分别包括：

(1)名词短语规则集； (2)动词短语规则集；

(3)介词短语规则集； (4)虚词规则集。

这些规则都是用产生式构成的。当规则的条件部分与语句的某种情境相匹配时，即规则的条件部分得到满足时，规则的行动部分就执行，对语句施行某种操作。以句子分析的一条规则为例：

名词短语规则n：

若第一个词是名词或代词，并且，若第二个词是名词；

那么，在第一个词前加上"限定词"的符号；在第二个词前加上"名词"的符号；并且，把两者组合在一起，前面加上"名词短语"的符号，存入临时记忆；从原句中剥掉这两个词；然后退出该规则集。

由此可见，每一规则的行动部分，实际上包括许多个动作的执行。

由词典和规则集构成的系统的知识库，是机器对古汉语小故事进行句法分析、翻译和人机对话的基础。

（二）句法分析：

理解课文的第一步，是对课文中的句子进行分析。这是由句子分析子程序来执行的。句子分析子程序针对特定课文中的句子，查阅词典；根据词典中提供的信息，去访问相应的规则集；通过那些合适的规则的激活，把句子分解成各种短语结构或语块，建立句子的分析树。图 2 是故事“鹬蚌相争”中“蚌方出曝而鹬啄其肉”一句经过分析后建立起来的分析树。

人们在实际处理语句时，是否象短语结构分析方法那样，把一个句子中的各个词分成不同的组或语块？心理学家从知觉和记忆的角度，对这个问题进行过探讨[2],[3]。例如，约翰逊(Johnson)曾采用“成对联想”的方法，对此问题作过实验验证。结果表明，

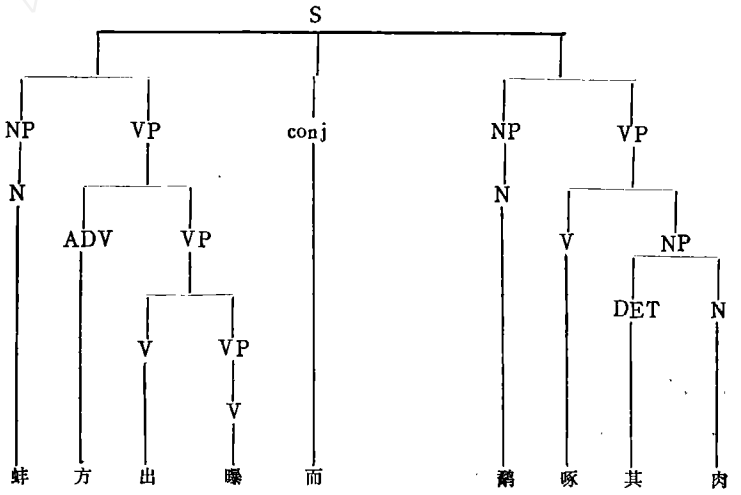

图2　句子的分析树

在短语交界的地方，其“转移错误概率”最高，即遗忘率高；而在短语结构内部，其转移错误概率较低。前者的遗忘概率要比后者大数倍。这说明，处理句子的短语结构分析方法是有一定的心理学根据的。它们确实抓住了人们如何分析和组织句子以达到知觉和记忆的特点。事实上，人们对句子的理解过程，在一定程度上可以说，就是对句子的分解和组块的过程。因此，一旦人们正确地把句子作了分解和组块，他也就基本上理解了这个句子。

同时，我们采用短语结构分析方法的另一目的，是试一试各种自然语言的处理方法在处理汉语时的有效性。句子分析的作业流程见图 3 。

对句子的分析，除了名词短语(NP)、动词短语(VP)和介词短语(PP)以外，我们还特设了一项虚词(FW)，古汉语的虚词数量多，出现次数也多。它们在一句话的不同位置，具有不同的句法功能。这是古汉语独特之处。在我们的十个古汉语小故事中，有十二个虚词，即：之、者、所、而、其、也、以、则、乃、耳、矣、就。仅以知识库中虚词规则中的“之”字为例，说明它在不同情况下的句法功能。

(1) 如果“之”在动词后面(V)(之)，或在介词后面(PRE)(之)，则作为代词，代表人、物或事。例如：

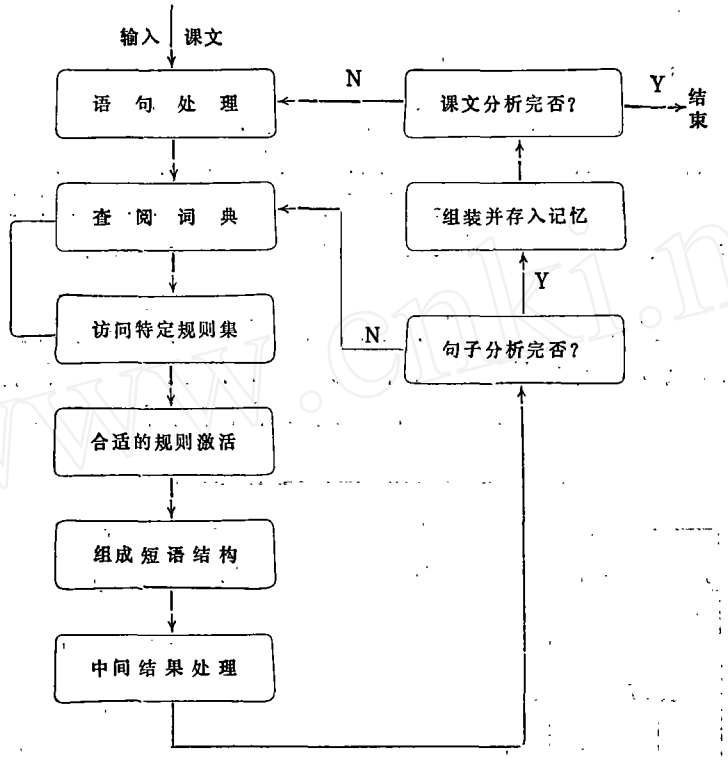

图3 句子分析作业流程

数人饮"之"不足。"之"代表酒。

遂与"之"行。"之"代表狐。

(2)如果"之"的前面无动词，"之"的后面为名词(之)(N)，则作为动词，是"去"、"到"的意思，例如

至"之"市而忘操之。

(3)如果"之"字在两个名词之间(N)(之)(N)，则"之"为助词，相当于"的"字。例如

以子之矛，陷子之盾

(4)如果"之"字前面是名词，后面是形容词(N)(之)(ADJ)，则"之"字是副词，相当于"很"、"非常"。例如：

"吾盾之坚，吾矛之利"译作："我的盾很坚固，我的矛很锐利"。

(5)如果"之"字在两个动词之间(V)(之)(V)，或两个动词短语之间(VP)(之)(VP)则"之"字为连接词。它不具有任何词义，只是连接它的前后两部分成为一句。例如：计之曰。

以上是以"之"字为例，说明古汉语的句法分析如何处理虚词。

分析过程主要是自底向上进行的。

(三)翻译：

句子经过分析后，就划分出了各种短语结构或语块。机器对课文的翻译，也是以短语为单位进行的。在作业过程中，翻译子程序根据不同类型的短语，去访问特定的翻译

规则集，寻找合适的规则；一旦发现某规则的条件部分得到满足，规则的行动部分立即被激活，执行一系列动作，其中包括某词是否需要译出；取词典中哪个词义；是否需要添加什么词；词序是否需要作适当的调整等等。

1. 添加：

为了使译文通顺，古汉语译成现代汉语时，有时需作词的适当的添加，例如：

(1)添补虚词"的"。若名词短语第一个词的句法特征为"限定词"或"形容词"，后面为名词，那么，在译出第一个词后，插入一个虚词"的"。如"吾盾"译为"我的盾"。

(2)添补量词。古汉语一般不用量词，现代汉语却需要。我们在词典中，对于人和动物的名词标明惯用的量词：人的量词为"个"，马、鹿、驴为"匹"，虎、狐、鸟、蚌为"只"，蛇为"条"。但我们规定，只有在"有"字后面出现这些词，才添加量词。例如在"鹬蚌相争"这个故事中，"蚌方出曝，而鹬啄其肉"译文中的蚌和鹬前面都不加量词，但"有死蚌"、"有死鹬"则译作"有只死蚌""有只死鹬"。

2. 省译：

古汉语小故事中的虚词，在某些情境下往往不必译出。就"之"而言，它在两个动词(短语)之间，起一"连词"的作用，象"誉之曰"、"计之曰"和"……饮之有余"等，在译成现代汉语时，不译出来反而显得顺口。根据规则的规定，凡满足这种条件时，中间的"之"就省略不译，只译前后两个动词。因此，机器就把它们分别译为"夸口说"、"心理盘算说"等。

3．选取不同的译词：

有些词有几种不同的译词，需要根据不同的上下文来选取合适的译词。如动词"为"，在不同的情况下，分别有"成了"("身为宋国笑"译为"……成了宋国人的笑柄")和"给……添上"("吾能为之足"译为"我能够给它添上脚")等多种译词。这些不同的译词，在词典中均予以列出。但究竟选取哪个译词，则由规则根据具体的上下文来确定。

4．词序调整：

古汉语的词序和现代汉语一样，起着重要的作用；而且，两者基本上是对应的，一般情况下，不需要词序的调整。但有时也需要把句子的结构调整一下，翻译起来才顺当。如"一人蛇先成"，它的结构为$NP_1 + NP_2 + VP$。机器在译此句之前，先用一条调序规则，进行预处理，把此句的结构调整为$NP_1 + VP + NP_2$，然后再进行翻译。至于短语内部词序的调整，那是在翻译过程中处理的。

5．上下文参照：

词的翻译或增删，不但需要参照短语内部或句内的上下文关系，有时还要参照句外的上下文关系。例如，在"画蛇添足"一文中，前面提到"一人蛇先成"，后面又出现"一人之蛇成"。人们懂得，这两个"一人"并不是指同一个人，根据上下文来看，后者是"另一个人"，所以，机器在处理时，也同样地往回搜索一下，看看前面是否已经提到过"一人"。如果已经出现过，并且带有表示时间先后的副词，那么，在译后一个"一人"时，前面应添加个"另"字；并且，根据量词添加规则，在"一"与"人"之间，添加量词"个"。这样，该短语就被译为"另一个人"，看起比较通顺和达意。

句子的结构可能是不同的。有的是简单句，有的是复合句；有的是并列复合句，有

的属于主从复合句。所以，在翻译时，要把整个句子分解成各子句，送到不同的层次去处理。当然，最终还是以一个短语结构为单位，去访问翻译规则集。通过规则的引导，去查阅词典，再通过规则生成译文。一个较长的短语，往往需要连续激活一条以上的规则才能被译出。翻译作业的流程见图 4

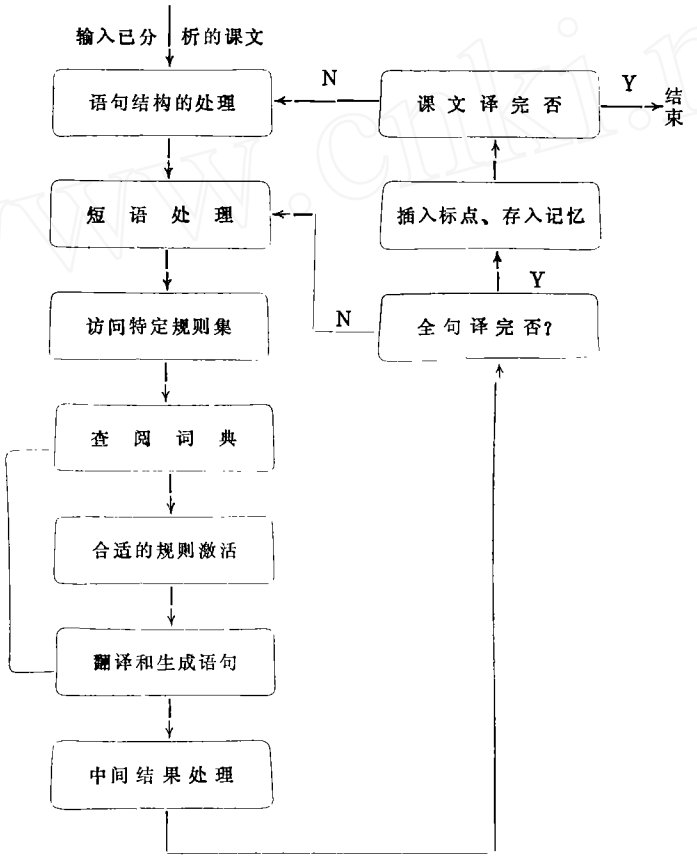

图4 翻译作业流程

(四)人机对话：

我们还设置了人—机对话部分，使机器在一定范围内能处理现代汉语，通过现代汉语与人进行对话，回答有关的问题。

对机器提的问题分两类：一是“是什么”，二是“为什么”。前一类问题包括句法知识、词、句的解释等。例如：

问：“以子之矛，陷子之盾”中的“之”字是什么意思？

答：这两个“之”是助词，译为“的”。

对“为什么”问题的回答，往往需要另外一些知识。例如：

问：“数人饮之不足”、“一人饮之有余”中的“之”字指什么？

答：“之”是代词，指的是“酒”。

问：为什么？

答：因为动词“饮”后面的宾词必须是“液体饮料”，在课文中只有酒是液体饮料。

问：在“鹬蚌相争”这个故事中，为什么“今日不雨、明日不雨，即有死蚌？”

答：因为蚌若缺水时间久了就会死。

人机对话需要许多并非课文提供的知识。在我们的人机对话程序中，这种知识还不多。因此，它所能回答的问题是十分有限的。

三、需要讨论和进一步探索的若干问题

前面叙述了我们对机器理解古汉语所做的一些探索性工作。有些问题我们或许处理不当，甚至并没有找到合适的解决办法。

(一)关于格分析：

我们对句子的结构人工地进行了格分析，但是在系统中没有应用。古汉语和现代汉语的结构有较强的对应关系，除虚词以外，基本相同，因而，不标明词的格，对译成现代汉语影响不大。但是，如果将古汉语译成其它民族语言，标明词的格是必要的。

我们根据十个小故事，把古汉语的格暂定为以下七个，即主格、宾格、宾主格、受格、工具格、处所格(处所始、处所终)和状态格。当然，这七个格是以有限的材料为依据的，不一定适当，有待进一步讨论。

(二)关于动词的时态问题：

对句子的分析，我们也未标明动词的时态。这是因为古汉语和现代汉语的动词，并无严谨的时态规则。古汉语往往以动词前面的副词表示动词的时态。例如：

蚌方出曝，舟已行矣，引酒且饮之。

但更多的情况是，整个故事中的动词都没有任何时态标志。例如：故事“自相矛盾”就是如此。若单独分析其中一句话，如“誉之曰”，我们很难确定它的动词的时态。但是，把这一句话置于故事之中，我们就知道，它们都是过去的行动，因为故事本身就是过去发生的事。机器译成的现代汉语，虽无时态标志，我们看起来还通顺，是因为它符合我们的思维和语言习惯。

(三)关于词的增删：

古汉语非常简练，往往省略一些词，译成现代汉语时，需要予以添补。前面已谈到量词和虚词“的”的添补。此外，有时还需添补介词、动词及其他的词。

“放之山下”和“置之其坐”这类句子，相当于现代汉语的“把”字句。在译第一句时，需加上介词“把”和“在”，译为“把它放在山下”。这种介词的添补，根据动词的语义特征和上下文关系，有时还是有规则可循的。

有些古汉语句子缺少动词，这种情况，只能针对具体问题具体解决，因为很难找到普遍可应用的规则。另外，动词的过去式一律加或不加“了”字，都不能使译文符合人们

的语言习惯。我们对动词宾语的过去时态加“了”字，其他情况一般不加。如“折颈而死”译为“折断脖子死了”；“兽见之皆走”译作“野兽看见它们都跑了”。可是“了”的用法很多，有些用法人很容易理解和使用，但对机器来说却非常困难。在这种情况下，不得不作一些具体规定来处理。

关于名词的复数怎么处理和表示，也是一个困难的问题。“狐假虎威”中的“百兽”和“兽”，指的都是复数。机器把前者译作“各种野兽”，而把后者译为“野兽”，这是根据前面的数词而作的具体规定。在古汉语句子中，经常省略主语。译文在什么情况下应该添加主语，并无规律可循。我们让机器“忠于”原文，不添补主语，但有时不见得很通顺。

此外，如前面已谈到的，某些词的省略不译，可以使译文更加通顺。但是，什么情况下应省略不译，要找到普遍的规律.这是非常复杂的问题。

总之，由古汉语译成现代汉语，为了使译文接近人们的语言习惯，必须对一些词进行增删。但是，许多情况下，很难找到可普遍应用的规则。这是需要进一步加以探索的问题。

四、结 束 语

我们以十个古汉语小故事为材料，对机器理解古汉语的问题进行了实验研究。

机器对古汉语的处理是以短语为单位进行的，并特设了一项虚词短语。古汉语的一大特点是经常使用虚词。除虚词以外，古汉语的句法结构与现代汉语基本上类似。因此，基本上依照古汉语的原有形式，可以译出能为人接受的现代汉语。这也是因为，我国人民习惯于依靠较大的知识背景、依靠思维和推理来理解语言，而单独一句话的形式居于次要地位。

由于古汉语非常简炼，译成现代汉语时往往需要添补一些词。有些词的添补，只能根据具体情况做些具体规定，因而还不能称之为规则，也就是不能称之为语法。如何寻求既符合我国人民的语言习惯，又适用于计算机的语法，是语言学家、心理学家和计算机科学家应该共同努力解决的问题。

参 考 文 献

[1] 李家治、陈永明，机器理解古汉语的初步探索，《心理学报》，1985年，第3期。

[2] Johnson, N., The psychological reality of phrase-structure rules. Journal of Verbal Learning and Verbal Behavior, 1965, 4, p. 469—475.

[3] Marks, L. E., Some structural and sequential factors in the processing of sentence. Journal of Verbal Learning and Verbal Behavior, 1967, 6, p. 707—713.

A STUDY ON COMPUTERIZED INTERPRETATION OF ANCIENT CHINESE AND RELATED PROBLEMS

Chen Yong-ming　Li Jia-zhi　Cai Shan
Institute of Psychology, Academia Sinica, Beijing

Abstract

This paper describes an ancient Chinese language understanding system (ACLUS) in processing ten proverbial stories written in ancient Chinese. The system consists of a knowledge base and three subprograms. The knowledge base includes two parts, one is a dictionary, the other consists of a few sets of rules for processing sentences. The words in the dictionary are annotated with their lexical meanings and syntatic and semantic features, and also with information on semantic match between some verbs and nouns.

The syntatic analysis subsystem makes sentence parsing and provides special treatment of function words (FW),The translation subsystem translates stories from ancient into modern Chinese, The question-and-answer subsystem behaves like a dialogue between teacher and student about the meanings of words and the meanings or grammatical construction of some sentences.

The paper discusses the problems about case analysis, the tense of the verb and the addition or deletion of words in a sentence in ancient Chinese so as to facilitate interpretation and translation. The authors point out that there is a remarkable difference between Chinese and Western languages, and that the Chinese have their own way of understanding their mother tongue.To identify the case of a noun or the tense of a verb, the Chinese do not depend on the morphological signs of the word, but on a broad understanding of the context.