心 理 学 报
第1期 ACTA PSYCHOLOGICA SINICA 1982年

机器理解汉语——实验 I[1)]

李家治* 郭荣江** 陈永明*

一、问 答 片 断

本实验模拟一次动物常识课的师生对话。由学生(人)提问，老师(机器)回答。下面是问答的片断。

人：麻雀、燕子和鸵鸟都是鸟吗？机：它们都是鸟。人：麻雀、燕子和鸵鸟都有翅膀吗？机：它们都有翅膀。人：它们都会飞吗？机：麻雀和燕子都会飞，鸵鸟不会飞。人：既然鸵鸟有翅膀，为什么它不会飞？机：虽然鸵鸟有翅膀，可是它的翅膀是退化的，没有飞的功能，所以它不会飞。人：蝙蝠有翅膀，又会飞，它也是鸟，对吗？机：不，蝙蝠不是鸟。人：它是什么呢？机：它是兽。人：什么是鸟呢？机：鸟有羽毛，有翅膀，有两条腿，会下蛋，是卵生动物。人：什么是兽呢？机：兽有毛，有四条腿，会生崽儿，是胎生动物。人：蝙蝠不是有翅膀吗，为什么它是兽呢？机：蝙蝠是有翅膀，也有四条腿，可是它的腿是特化的，前肢特化成翅膀的骨架，后肢短小，翼膜连接前后肢，构成蝙蝠的翅膀。它还有毛，会生崽儿，是胎生动物，所以它是兽。

二、模 型

本实验采用语义网络模型。这种模型是M. R. Quillian于1968年首次提出来的[6]。他认为无论是理解别人的话，或是回答别人的话，都不能离开记忆。人们的记忆中有概念以及概念与概念之间的关系。由于许多概念是相互联系的，在人们的记忆中似乎形成了一种网络，叫语义网络。人们依靠搜寻和选择记忆中适当的语义网络路线来理解和回答别人的话。十多年来，这种模型经过多人的修改和发展[6][8]，在人工智能中形成了机器理解自然语言的主要途径之一。

我们的模型所用的概念包括鱼类、爬行动物类、鸟类和兽类(哺乳动物)中的一些动物，以及它们的主要特征和行为。图1是部分语义网络示意图。我们仅绘出了麻雀和鸵的各种关系，因而也只是鸟的语义网络的一部分。图中X和Y叫做"待填项"。X可以代入任何会飞的鸟的名称，Y可以代入除鸵鸟外任何不会飞的鸟的名称。

三、词、句法和问话句型

本文对于词和句型没有严格的依照文法教科书的规定来分类和命名。我们力图做到两点，即便于表示本实验的意图和符合汉语的语言习惯。

1) 本文于1981年7月14日收到。
* 中国科学院心理研究所 ** 中国科学院北京自动化研究所

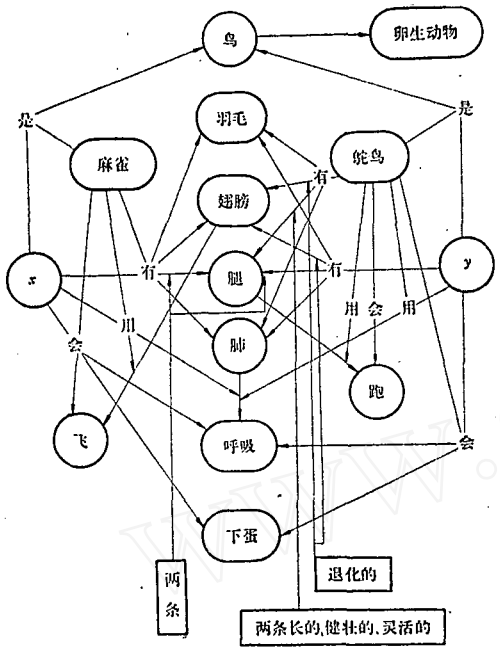

图 1 语义网络示意图

(一) 词

名称类别词(即动物的名称和类别)：鲤鱼、黄鱼、鳝鱼、乌龟、麻雀、燕子、鸵鸟、狗、马、鲸鱼、蝙蝠、鱼、爬行动物、鸟、兽、卵生动物、胎生动物、动物。器官词：鳞、鳍、鳃、须、肺、甲、羽毛、毛、翅膀、腿、蹄、爪。功能词：游水、爬、飞、跑、呼吸、产卵、下蛋、生崽儿。代词：它、它们。关系词：有、会、用、是。连接词：和、既然、虽然、可是、因为、所以。副词：还、又、也。疑问词：吗、对吗、对不对、是吗、是不是、呢。数量词：大多数、多半、少数、两对、两条、四条。助词：的。注解词：长的、短的、笨拙的、灵活的、健壮的、退化的、特化的。

注解：对于注解词必要时可以进一步注解。如对于蝙蝠的腿的注解词"特化的"的注解是：前肢特化成翅膀的骨架、后肢短小，翼膜连接前后肢构成蝙蝠的翅膀。

(二) 句法

本实验以词序分析进行句法分析；实验中的语句都是用基本陈述句的词序为基础的，即：〈主词〉〈关系词〉〈宾词〉

1. 主词：主词可以是单数、复数或待填主词。

〈主词〉：：＝〈单数主词〉｜〈复数主词〉｜〈待填主词〉

〈单数主词〉：：＝〈名称类别词〉｜〈名称类别词〉〈助词"的"〉〈器官词〉

例如：卵生动物、鸟、鲤鱼等都可作单数主词。再如：鸟的翅膀、乌龟的腿、鲤鱼的鳞，其中翅膀、腿、鳞都可用作单数主词。

〈复数主词〉：：＝〈几个并列的名称类别词〉｜〈几个并列的名称类别词〉〈助词"的"〉〈器官词〉｜〈凡是〉〈名称类别词〉｜〈凡〉〈除"是"以外的关系词〉〈的〉〈名称类别词〉

例如：〈马、狗和鲸鱼〉都是兽。

麻雀、燕子和鸵鸟的〈翅膀〉都是由羽毛构成的，

凡是〈兽〉都用肺呼吸，

凡有肺的〈动物〉；凡会飞的〈动物〉。

以上的例子中，〈 〉内的名称类别词和器官词都用作复数主词。

〈待填主词〉：由"凡"构成的短语中，最后的名称类别词可以省略，成为待填主词。在我们的实验中多半可以用"动物"来填补。例如：凡有鳍的都会游水，凡会飞的都有翅膀。

2. 宾词：宾词也可以是单数、复数或待填宾词。宾词所用的词由关系词来决定。

(1) 在关系词"有"后面是器官词，器官词之前可加数量词和注解词。例如：鸵鸟有腿，鸵鸟有两条腿，鸵鸟有两条长的、健壮、灵活的腿。最后这句话的词序是：

〈主词〉〈有〉〈数量词〉〈注解词〉〈宾词〉

(2) 在"会"的后面是功能词，如：鸵鸟会跑；蝙蝠会飞。

(3) 在"用"的后面是器官词和功能词联用，其条件是，器官和功能必须在语义网络中有关系，如：鸵鸟用腿跑；蝙蝠用翅膀飞。

(4) 在"是"的后面有以下几种情况：

Ⅰ. 表示归类

〈名称类别词〉 〈是〉 〈名称类别词〉：乌龟是爬行动物。

Ⅱ. 表示加重语气，"是"与其它关系词短语和助词"的"并用，最后是〈待填宾词〉，它们的词序是：

〈主词〉 〈是〉 〈其它关系词短语〉 〈的〉 〈待填宾词〉

乌龟是有四条腿的，乌龟是不会跑的，乌龟是用腿爬的。

以上三句的待填宾词都是"动物"。

Ⅲ. 表示对器官词的注解

〈名称类别词〉 〈的〉 〈器官词〉 〈是〉 〈注解词〉 〈待填宾词〉 | 〈名称类别词〉 〈的〉 〈数量词〉 〈器官词〉 〈是〉 〈注解词〉 〈待填宾词〉

鸵鸟的腿是长的、健壮、灵活的。鸵鸟的两条腿是长的、健壮、灵活的。

以上两句话的待填宾词是"腿"。

上面的一些例子是说明主词和宾词的用法，但也涉及到其它词的用法。现在对助词"的"的用法再进一步明确一下。

3. 助词"的"的用法：

（1） 在名称类别词之后，表示从属关系，如：鱼的鳞，马的腿，等等。

（2） 〈是〉 〈其他关系词短语〉 〈的〉 表示加重语气。如…是会飞的；…是不会跑的。

（3） 构成注解词，如笨拙的，灵活的，……等等。

4. 关系词。本实验只采用四个关系词，即有、会、用、是。每个关系词有单数、复数、肯定关系、否定关系、疑问关系等形式，共二十四种形式，如下表：

	肯定关系	否定关系	疑问关系
单数	有	没有	有没有
复数	都有	都没有	都有没有
单数	会	不会	会不会
复数	都会	都不会	都会不会
单数	用	不用	用不用
复数	都用	都不用	都用不用
单数	是	不是	是不是
复数	都是	都不是	都是不是

例如：蝙蝠有没有翅膀？

5. 连接词：

（1） "和"用来连接并列的主词或并列的宾词，构成复数主词或复数宾词。如：麻雀、燕子和鸵鸟都是鸟，凡是鸟都有羽毛、翅膀和两条腿。

（2） 〈因为〉〈陈述句〉，〈所以〉〈陈述句〉用来回答〈为什么〉疑问句。〈所以〉〈陈述句〉带有结论性质，但有时不说自明，因而可以省略。例如：问：为什么马不会飞？答：因为马没有翅膀，所以它不会飞。或答：因为它没有翅膀。

（3） 〈虽然〉〈陈述句〉，〈可是〉〈陈述句〉，〈所以〉〈陈述句〉用来回答〈既然〉〈陈述句〉〈为什么疑问句〉的问题。〈虽然〉同〈既然〉相对应，〈可是〉〈陈述句〉是对〈虽然〉〈陈述句〉的进一步说明或注解，表示特殊，例外等情况。〈可是〉后面的陈述句包括注解词，甚至还包括注解。例如：问：既然鸵鸟有翅膀，为什么它不会飞？答：虽然鸵鸟有翅膀，可是它的翅膀是退化的，没有飞的功能，所以它不会飞。

（4） 对于〈陈述句〉〈为什么疑问句〉，答话时可不用〈虽然〉一词，只对应进一步解释的词句

用〈可是〉〈注解词〉〈注解〉……〈所以〉〈陈述句〉来回答。

6. 副词：还、也、又

（1）"还"连接并列的关系词短语，表示有所增添的意思。例如：

蝙蝠有翅膀、有四条腿、有毛、还会生崽儿。

语句中某关系词短语中的词(本实验中都是器官词)有注解词和注解者，就先插入注解词和注解，注解之后，另起一句，并用副词"还"连接随后的关系词短语。例如：

蝙蝠有翅膀、有四条腿，可是它的腿是特化的，前肢……，它还有毛、会生崽儿，……。

（2）"也"在关系词短语之前，表示和其它对象有相同的条件，例如：凡是兽都有四条腿，蝙蝠也有四条腿。

如果同其他对象比较，有几种或多种条件相同，副词"也"只加在第一个关系词短语之前，例如：蝙蝠也有四条腿，有毛，……等等。

（3）"又"，两个以上关系词短语并列，最后的关系词短语之前加"又"字，表示更进一层的意思。例如：

蝙蝠有翅膀，又会飞，……

以上是本实验所用的句法。关于疑问词的用法将在问话句型中附带说明。

（三）句 型

本实验所用的句型有以下数种

Ⅰ．基本陈述句：句型是

〈主词〉〈关系词〉〈宾词〉

主词在关系词之前，宾词在关系词之后。如"鸵鸟有翅膀"

Ⅱ．一般疑问句：

1. 〈基本陈述句〉〈除"呢"以外的其他疑问词〉，如：鸵鸟有翅膀吗？
2. 〈主词〉〈疑问关系词〉〈宾词〉，|
   〈主词〉〈疑问关系词〉〈宾词〉〈呢〉，如：鸵鸟有没有翅膀？鸵鸟有没有翅膀呢？
3. 〈是不是〉〈主词〉〈关系词〉〈宾词〉|
   〈主词〉〈是不是〉〈关系词〉〈宾词〉|
   〈主词〉〈关系词〉〈宾词〉〈是不是〉，
   语尾可加、可不加疑问词"呢"，如：是不是鸵鸟有翅膀？是不是鸵鸟有翅膀呢？| 鸵鸟是不是有翅膀？鸵鸟是不是有翅膀呢？鸵鸟有翅膀是不是？鸵鸟有翅膀是不是呢？

例外：如果基本陈述句中的关系词为"是"，要避免用"是不是"在主词前后造成疑问句。

Ⅲ．加重疑问句。

1. 〈是〉〈主词〉〈关系词〉〈宾词〉〈疑问词"吗"〉|
   〈主词〉〈是〉〈关系词〉〈宾词〉〈疑问词"吗"〉|
2. 〈不是〉〈主词〉〈关系词〉〈宾词〉〈疑问词"吗"〉|
   〈主词〉〈不是〉〈关系词〉〈宾词〉〈疑问词"吗"〉如：是鸵鸟有翅膀吗？鸵鸟是有翅膀吗？不是鸵鸟有翅膀吗？鸵鸟不是有翅膀吗？

Ⅳ．"什么"疑问句：

1. 〈什么动物〉〈关系词〉〈宾词〉
   这里的关系词只限于用"有"、"会"、"用"，如：什么动物用鳃呼吸？
2. 〈什么〉〈是〉〈宾词〉
   这里的宾词只限于用名称类别词，〈什么〉后面有一待填主词〈动物〉，如：什么是鸟？什么是兽？

3. 〈主词〉 〈是〉 〈什么动物〉 | 〈主词〉 〈是〉 〈什么〉，如：蝙蝠是什么动物？蝙蝠是什么？

4. 〈主词〉 〈用〉 〈什么〉 〈功能词〉，如：鲸鱼用什么呼吸呢？

"什么"疑问句的句尾都可加、可不加疑问词"呢"

Ⅴ. "为什么"疑问句：

1. 〈为什么〉 〈主词〉 〈关系词〉 〈宾词〉 |
〈主词〉 〈为什么〉 〈关系词〉 〈宾词〉 |
〈主词〉 〈关系词〉 〈宾词〉 〈为什么〉，如：为什么鸵鸟不会飞？鸵鸟为什么不会飞？鸵鸟不会飞为什么？

2. 〈陈述句〉 〈为什么疑问句〉

3. 〈既然〉 〈陈述句〉 〈为什么疑问句〉

4. 〈加重疑问句〉 〈为什么疑问句〉，如：鸵鸟有翅膀，为什么它不会飞？既然鸵鸟有翅膀，为什么它不会飞？鸵鸟不是有翅膀吗？为什么它不会飞？"为什么"疑问句尾都可加、可不加疑问词"呢"。

Ⅵ. "凡是"、"凡……的"组成的疑问句：

1. 〈凡是〉 〈主词〉 〈复数关系词〉 〈宾词〉 〈疑问词"吗"〉 这里的复数关系词只用"都有""都会""都用"，如：凡是鸟都会飞吗？

2. 〈凡……的〉 〈复数关系词〉 〈宾词〉 〈疑问词"吗"〉，〈凡……的〉后面有一待填主词〈动物〉如：凡会飞的都是鸟吗？

四、程序及其功能

在我们的实验中，机器理解问话和组织答话要通过以下四个步骤：

1. 判别句型，2. 分析语句，3. 查问知识库，4. 组织答话。

图2是判别句型的流程图。问句输入机器后，依次检查句中有无"为什么"、"凡"、"凡是"、"什么"等等。如果有"为什么"这个词，就可以确定输入问句属于"为什么疑问句型"。但这一句型有四个亚型，即 V_1、V_2、V_3、V_4。根据这些亚型的结构特点，我们可以先让机器以有无"吗"、"既然"、"，"来判别它们，待分析语句时再进一步分析。

上述四个步骤中比较复杂的工作是分析语句，而最复杂的是组织答话，这两种工作，对于人来说，都需要智能来完成。对于机器来说，则需要人工地给予它某些智能。

我们的程序使机器具有一点推理能力和一点正确运用汉语句法的能力。兹简述如下：

推理：

1. 概括能力，即从具体事例中概括出共同特征，得出一般概念。例如问话：什么是鸟？这里，鸟是一个一般概念，在机器的记忆中并没有现成的答案。如果仅依靠信息的提取来答话，结果将是：麻雀是鸟，燕子是鸟，鸵鸟是鸟，鸟是卵生动物。

这样的答话未能概括鸟的全貌，没有给出鸟的一般概念。我们的程序使机器从麻雀、燕子和鸵鸟少数几个例子中抽出它们的共同特征，所做的答话可以说是对鸟下了一个粗略的定义：鸟有羽毛，有翅膀，有两条腿，会下蛋，是卵生动物。答话没有提鸟会飞，因为有的鸟不会飞。

2. 依靠部分信息进行推理。在前面的语义网络示意图中，有X和Y两个待填项。X可以代入任何会飞的鸟的名称，Y可以代入任何不会飞的鸟的名称。依靠部分信息进行推理，可以适当地回答一些问题。例如问话：鸽子是鸟，它有翅膀吗？

在机器的记忆中，并没有"鸽子"这个名称。可是既然"鸽子"是鸟，它就可以代入X和Y项中，而且

无论代入哪一项，都有翅膀。因此机器回答：鸽子有翅膀。再问：鸽子是鸟，它会飞吗？在机器的记忆中，大多数鸟都会飞，因此回答：鸽子多半会飞。这一答话是推测来的，而且是正确的。可是，依靠部分信息进行推理有时也会犯错误。例如：问：企鹅是鸟，它有翅膀吗？答：企鹅有翅膀。再问：它会飞吗？答：它多半会飞。

第二句答话当然是错误的。这个例子只是为了说明依靠问话所提供的部分信息进行推理虽然是可能的，但是在提供的信息中不可缺少关键性的信息。在我们的模型中有两种鸟，即会飞的鸟和不会飞的鸟。关键性信息是"会飞"或"不会飞"。缺少这种关键性信息，即便是人，如果对企鹅毫无所知，也会发生同样错误。

3. 依靠间接信息进行推理。如果对人提出一个问题："为什么马不会飞？"人们立刻可以回答："因为马没有翅膀"。把"翅膀"和"飞"两个概念联系在一起，是人们从儿童年代起经过长期的经验和学习

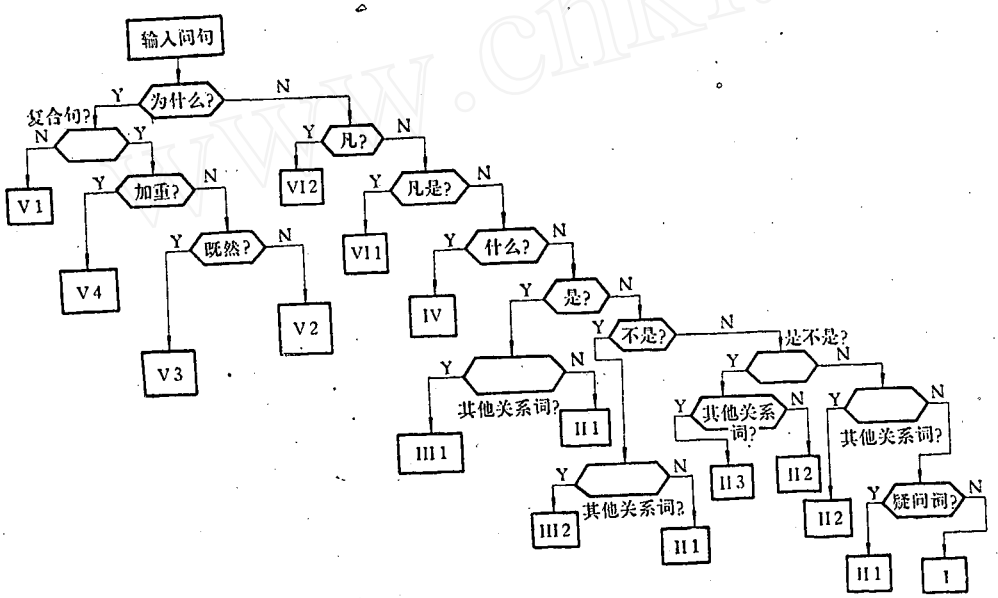

图 2 句型判别示意图

的结果。我们的机器不是学习机，不能建立这样的联系。可是我们的机器能够间接地从会飞的动物中得到信息，做出回答。它发现凡有"飞"的功能的动物都有"翅膀"这种器官，而且在语义网络中"翅膀"和"飞"是有关系的，即都是用"翅膀飞"，没有"翅膀"的动物就没有"飞"的功能。机器再查询知识库，看马是否有翅膀，马没有翅膀，因而也没有飞的功能，于是回答：因为马没有翅膀，所以它不会飞。

凡会飞的动物都有翅膀，可是有翅膀的动物不一定都会飞。因为有些动物的翅膀是退化的，没有飞的功能。例如，问：为什么鸵鸟不会飞？

机器通过与上述类似的查寻，发现鸵鸟有翅膀，可是它不会飞，它的翅膀也注明是"退化的"，可以说鸵鸟的翅膀无飞的功能。因而机器回答：因为鸵鸟的翅膀是退化的、没有飞的功能，所以它不会飞。

以上是我们的程序给予机器的一点推理能力。

句法的分析和运用：

正确运用句法本身就是一种高级的智能，一个白痴运用句法的能力是很差的。我们的机器除具有一点正确运用句法的能力以外，在程序中还包含有少数几条规则，使机器在分析语句和组织答话时有所遵循。

1. 分解和填补规则：对于包含复数主词的输入问句，机器在分析语句时要把复数主词分解为几

个单数主词，并对每一单数主词填补适当的关系词、宾词和疑问词。例如：输入问句为："麻雀、燕子和鸵鸟都是鸟吗？"分解为：麻雀是鸟吗？燕子是鸟吗？鸵鸟是鸟吗？

再如，输入问句为："凡有翅膀的都是鸟吗？"这句话的主词待填补，填补后的问句是："凡有翅膀的动物都是鸟吗？"机器首先要查明哪些是有翅膀的动物，然后再进行分解和填补，得出四句问话：麻雀是鸟吗？燕子是鸟吗？鸵鸟是鸟吗？蝙蝠是鸟吗？机器依照分解和填补规则来理解输入问句、查问知识库和寻找答话。

2. 合并和删除规则。根据查问知识库的结果组织答话时，要把分散答话合并为一句，并删除重复的词是，合并后的答话是：

〈主词1、主词2、…和主词n〉　〈复数关系词〉　〈宾词〉。

最后一个主词之前加连接词"和"。例如通过分解填补、查问知识库得到的答案是：麻雀是鸟、燕子是鸟、鸵鸟是鸟。合并后得出答话：麻雀、燕子和鸵鸟都是鸟。

3. 代替规则。单数主词与代词"它"、复数主词与代词"它们"可以相互代替。在分析语句和查问知识库时，以单数或复数主词代替"它"或"它们"；在组织答话时相反。代替规则有三条：

（1） 在同一句话中，重复出现的主词以代词代替。（2）答话的主词如果与问话的主词相同，就用代词代替。（3）在连续谈话的情况下，如果主词不变，除第一句话以外，以后每句话，无论是问话或是答话，都可用代词代替。

4. 转换规则。关系词"有"的句子和关系词"是"的句子有时是需要相互转换的，而且是可以转换的。转换规则是：

〈名称类别词〉　〈有〉　〈数量词〉　〈注解词〉　〈器官词〉

〈名称类别词〉　〈的〉　〈数量词〉　〈器官词〉　〈是〉　〈注解词〉

例如："鸵鸟有两条长的、健壮、灵活的腿"，可转换为"鸵鸟的两条腿是长的、健壮、灵活的"。

5. 镶嵌规则：机器组织答话，有时需要比较复杂的句子。在这种情形下，机器要适当地选用有关的句法及上述规则，把不同的词安排在一句话中的适当位置，例如，对于问话"蝙蝠不是有翅膀吗，为什么它是兽？"机器查问知识库，得："蝙蝠有翅膀、有四条腿、有毛、会生崽儿，是胎生动物"。

仅这几点就可以回答蝙蝠是兽，可是程序使机器通过以下几个步骤，组成更合适的句子：（1）以加重语气回答加重语气的问话，（2）比较蝙蝠和兽的特征，适应地应用副词"也"，（3）对于蝙蝠的腿加上注解词和注解，（4）适当地应用副词"还"。

最后，机器回答：蝙蝠是有翅膀，也有四条腿，可是它的腿是特化的、前肢特化成翅膀的骨架，后肢短小，翼膜连接前后肢，构成蝙蝠的翅膀；它还有毛，会生崽儿，是胎生动物，所以它是兽。

五、讨　　论

我们的语义网络模型是参考了A. M. Collins 和 M. R. Quillian[3] (1969), P. H. Lindsay 和 P. A. Norman[5] (1972)以及R. F. Simmons[8] (1973) 等人的模型制定的，但和他们的模型不尽相同。我们的模型有三个特点：

1. 应用关系词：在我们的模型中，如果概念与概念之间存在关系，就不仅标明有关系、而且标明有什么关系；在语义网络示意图中，不仅用有向弧 (directed arc) 表示有关系，而且在有向弧上用关系词表示有什么关系：在资料结构中也有相应的安排，这一点是和 Collins 与 Quillian 的模型不同的。在他们的模型中，概念与概念之间仅由有向弧来表示存在关系。

我们所以在有向弧上标明关系词，是因为这样才符合人们的语言习惯。一句话如果

没有关系词多半就没有意义：同时，只有在关系词的指引下，人们才可以较快地沿着记忆中的语义网络路线寻求适当的答话。这样做也是节省机器探寻活动的一种措施。因为，一个概念可以同许多其他概念有关系，可以从一个概念伸出许多有向弧。如果不在有向弧上标明关系的性质，机器就会进行许多无用的探寻活动、浪费很多时间。

我们所以用"关系词"这个名称，是由于概念与概念之间有各式各样的关系，这些关系不能用文法教科书中任何单独一个词类所概括。在我们的实验中，仅运用了"有"、"会"、"用"、"是"这四个词作为关系词。在文法书中，"是"和"有"都是动词，"会"是助动词，"用"是介词。

以上四个关系词有肯定、否定、疑问三种形式，每种形式又有单数、复数的区别。我们用这些形式的关系词提问、让机器造句、答话，并未遇到困难或出现生硬、不通的句子。

我们对吕淑湘主编的"现代汉语八百词"[1]做了粗略的统计，发现约有一百个词可以依照上述形式用作关系词，其中包括动词、助词、连词、介词等。可见，关系词的这种用法是有一定普遍意义的。

但是，我们的关系词的用法，也存在一些问题：

（1） 我们所用的四个关系词有肯定、否定、疑问三种形式。可是具有这三种形式的词不一定都是关系词。例如：

这张画画得好、(肯定)

这张画画得不好，(否定)

这张画画得好不好？(疑问)

在这三句话里"好"、"不好"、"好不好"是对绘画技术的评价，是对画的形容词。

（2） 在汉语八百词中虽然约有一百个词可以依照肯定、否定、疑问三种形式用作关系词，可是其中百分之八十以上是动词。这些动词还有其他形式的用法，如过去、现在、将来三种时态的形式。

（3） 我们的四个关系词实际上是对我们模型中涉及的关系所做的分类。与此相应也对概念进行了分类。在我们的问答系统的狭窄谈话范围内，这样的分类是可行的。可是这只是权宜之计。当谈话范围扩大，所涉及的概念和关系增多的情况下，怎样对概念和关系进行分类，是机器理解汉语需要解决的问题。

2. 应用词序分析：我们的模型与Lindsay, Norman以及 Simmons 等人的模型的区别在于，我们的模型是依靠词义和词序来确定一句话的含意，他们的模型是依靠"格分析"。

在拉丁语系统的语言中，有所谓名词变格，即对于一个名词，要看它的词尾来确定它是处于主格、宾格、与格、有格等等，进而确定一句话的含意。C. J. Filmore[4] (1968)把这种思想应用于格变化不明显的英语，借以找出一句话的真实含义，如在一句话中，要找出谁是"施动者"，"受动者"和"受益者"等等，Lindsay Norman. Simmons 等人就是在格分析的基础上建立语义网络模型的。

我们所以没有用他们的格分析方法，是由于我们在初步实验中不打算采用复杂的、容易产生歧义的句型；另一方面，也想考查一下"词序"在机器理解汉语中的作用。汉语没有词尾变化，也不用字形的其他变化来确定一个词是什么格。汉语的词序在很多情况下发

挥一定作用，一个词在一句话中的位置往往就决定了它是处于什么格。我们的做法也可以说是利用词序进行格分析。

我们的实验仅应用了汉语的一种基本词序，即基本陈述句的词序：

〈主词〉〈关系词〉〈宾词〉

这一基本陈述句的词序加上疑问词就成为疑问句。在我们的问话句型中可以看出，有些疑问词只能放在一句话的末尾，如"吗"、"呢"、"对吗"等；有些疑问词可以放在一句话的句首、句中或句尾，如"为什么"、"是不是"等。无论疑问词放在什么位置，主词、关系词、宾词的前后关系仍保持不变，问话的含义也不变。我们的机器对于十几种问话句型的理解没有发生困难。

机器组织答话用陈述句。我们的机器可以依照问话的要求，在主词、关系词、宾词之前，正确地加上形容词、数量词、副词、一些短语或注解词句，但基本陈述句的词序不变，答话的主旨也不变。

汉语的介词如"被"、"把"、动词"是"或助词"的"与动词"是"联用，可以改变基本陈述句的词序，句子的主旨也不变，仅语气有所改变。例如，"宝宝吃了苹果"这一基本陈述句的词序可以有如下改变：苹果被宝宝吃了；宝宝把苹果吃了；吃了苹果的是宝宝。

以上这几句话虽然改变了原来的词序，可是根据新添加的介词，助词"的"和动词"是"，我们仍能清楚地指出谁是"施动者"，什么是"受动者"。这是由于，添加了新词，形成了新的词序。正是新的词序使我们能够正确理解一句话的含义。

上述这类句型，我们虽然还没有在机器上试，但估计机器对它们的理解也不会遇到什么困难。

在汉语中，介词"被"可以把宾词调到句子的首位，介词"把"可以把宾词调动到句子的中间，我们根据介词的词义和词序就可以理解一句话的含义。可是，在日常谈话中，人们往往又把介词省略掉。例如，对于"我吃过早饭了"这句话，人们往往说成：早饭我吃过了；我早饭吃过了。如果添上介词，人们反而觉得不顺耳：早饭被我吃过了；我把早饭吃过了。

省略掉介词，把两个名词留在句首，对于这类句子，仅凭前面所说的那样词序分析就有困难了。可是，人对这类句子的理解并不感到困难。这是因为人们知道早饭不会吃东西，而是被吃的东西，只有人，或者扩大到动物园的高等动物才会吃早饭。

可是机器怎样处理这一问题？词序分析遇到的这种困难，也是格分析的困难。要解决这种困难，恐怕只有一条共同的途径，就是把人理解这类句子的依据和方法，化为规则，让机器依照这些规则来判断谁是施动者，谁是受动者，而不是由人替机器做这样的判断。

把人理解上述句子的依据和方法化为规则，就涉及名词和动词的分类问题。"我"是人，是高等动物，动物是有主动行动的，因此动词总是同动物的名词联在一起：自然能源也有动词直接联系，如阳光照耀、微风吹动等等；人造的有动力的机器也有动词直接联系，如6104班机在飞翔，汽车奔驰，计算机运转等。虽然上述各类名词都可以直接连接动词，但是"吃"这个动词只可以同动物或人直接联系，而"说"、"读"、"画"、"写"、"思考"、"欣赏"这类动词只能同人直接联系。

非动物、非能源、非机械的无生命物质，如果有动词直接连在后面，也只表明这些物质

是受动者而非施动者。例如：窗子破了，玻璃碎了。

窗子不会自动破，玻璃不会自动碎，一定是窗子被打破了，玻璃被打碎了。窗子和玻璃都是受动者。至于谁是施动者，就要看上下文。可能是大风、地震、玩皮孩子投的石子，或是擦窗子的粗心笨拙的主人。

对名词和动词进行适当的分类，制定一些规则，说明在什么情况下哪类名词应该同哪类动词相联。再加上其他词类、词序的应用规则，可以使机器对困难的句子进行正确分析和组织答话。但是怎样对名词和动词进行分类，是一个重大课题，是我们今后的艰巨任务。

3. 推理的应用。我们的模型使机器具有归纳概括能力，依靠部分信息进行推理和依靠间接信息进行推理的能力。这也是与国外的一些作者不同的。

Quillian, Lindsay 和 Norman 等人的模型可以由类特征来推论个体特征，即由共性求个性，这种做法是有缺点的，它不能推求出个体间的差别。此外，如Collins和Quillian 1969年的模型，把"会飞"作为"鸟"的一种特征，显然是忽略了还有很多不会飞的鸟。这样会导致推理错误。

目前，机器理解自然语言的研究，无论是语义网络理论或其他理论。都是在很狭窄的知识范围内进行实验，推理的应用是很有限的。而且不同作者有不同的目的，一些实验虽然涉及到推理，但并不一定能解决推理的问题。对于理解自然语言需要一些什么推理，以及怎样实现这些推理，似乎还处于多途径的探索阶段。

例如，R. C. Schank 等人的 MARGIE 程序[7]，对机器输入一句话，机器能够输出几句有关而且相当合理的话，例如，

输入：约翰给了玛丽一些阿司匹灵，

输出：（1）玛丽病了，（2）玛丽需要阿司匹灵，（3）约翰相信玛丽需要阿司匹灵，（4）玛丽想要好受一点。

他们的做法是对一些原始行动和受行动支配的物体做了一些注解，指出与这些行动和物体有关而且可能产生的含义。如"给"的原始行动是把物体从甲转移到乙。如果甲、乙都是人，则含义是乙需要这种物体，或甲相信乙需要这种物体。这种物体如果是好吃的东西，则乙可能很高兴；如果是药，则乙可能是病了，药会减轻他的痛苦。等等。

再如，E. Charniak 利用一些"小鬼"(demon)解决推理问题[2]。例如，给机器一句话，以类似的中国话来说："小胖摇晃了几下璞满，拿了两个五分硬币跑到对门商店去了。"然后问：小胖哪儿来的钱？他跑到对门商店去干什么？

人回答这两个问题并不困难。因为人根据生活经验知道璞满是儿童们存钱(硬币)用的。摇晃几下会蹦出钱来。对门商店卖什么东西？看故事的全文就可知道，不过可以推知是个食品店。小胖跑到那里去可能是要买冰棍、糖葫芦、果丹皮或其他好吃的东西。Charniak 把这些知识分别交给一些"小鬼"。这些小鬼各守岗位，一遇到同自己的知识有关的问话就跳出来答话。

我们实验中的推理也是局限在狭窄的知识范围。可是我们着眼于使这些推理能应用于理解一般自然语言，即常识性的，不需要专业知识的谈话。可是，这种一般性的谈话究竟需要一些什么样的推理，还不清楚。我们只能把心理学和形式逻辑中所讲的归纳推理

和演绎推理编进程序。在我们的程序中，归纳概括能力属于归纳推理，依靠部分信息和依靠间接信息进行推理属于演绎推理。借助于这些推理，机器可以回答一些仅凭记忆不能回答的问题。

我们所用的三种推理，仅是归纳推理和演绎推理中的一部分。今后我们要逐步增加推理的种类，考察一下这些推理对于理解自然语言是不是有益的，必要的。

六、总　　结

本实验采用语义网络模型，模拟一次动物常识的师生对话，内容包括鱼类、爬虫类、鸟类及兽类十几种动物的特征和行为。

机器分析问话和组织答话，除了要遵循各词类所规定的用法以外，还要适当地应用下列五条规则：（1）分解和补充规则，（2）合并和删除规则，（3）代替规则，（4）转换规则，（5）镶嵌规则。前三条规则是为了正确应用代词；第四条规则是为了把用关系词"有"组成的句子转换成用"是"组成的句子；第五条是为了把不同的词安排在适当的位置以便组成较复杂的句子。

各词类的用法以及分析和组织语句的规则都是根据汉语的一个重要特点—"词序"进行的。这就是说，机器理解一句话的真实含义是依靠"词序分析"。本实验仅应用了一个基本陈述句的词序，即〈主词〉〈关系词〉〈宾词〉。

基本陈述句加上疑问词就成为疑问句。在句首、句中或句尾加上适当的疑问词，只要主词、宾词的前后顺序不变，问话的含义就不变。同样，在主词、关系词或宾词之前加上形容词、副词、或一些注解性的词、短语或句，可以组成比较复杂的陈述句。只要基本词序不变，陈述句的主旨也保持不变。我们的机器对于以这一基本词序为基础建造的十几种问话句型，并没有发生理解困难，并且能够以这一基本词序为基础，组织比较复杂的陈述句来回答问话。

我们注意到，汉语的一些词类，特别是介词，可以改变上述词序而一句话的含义不变，甚至因添加介词而被改变了的词序，在日常谈话中又往往把介词取消而一句话的含义仍保持不变。机器怎样处理这些情况，是我们今后需要研究的课题。

正确理解问话和组织答话，不仅需要句法分析（在我们的实验中即词序分析），而且需要推理。我们的程序使机器具有少量推理能力，即归纳概括能力，依靠部分信息进行推理和依靠间接信息进行推理的能力。机器依靠这些能力，可以回答一些仅凭记忆系统中的知识不能回答的问话。但是，在一般常识性的谈话中，究竟需要一些什么推理能力，也是今后需要深入探索的问题。

*　　*　　*

本实验在设计过程中，曾得到社会科学院语言研究所范继淹同志的许多好的建议，并使我们得以参考"现代汉语八百词"（当时尚未出版）的部分底稿，对我们的实验设计很有帮助，特此鸣谢。

参 考 文 献

〔1〕 吕叔湘主编：现代汉语八百词，商务印书馆，1980
〔2〕 Charniak, E.: Towards a Model of Children's Story Comprehension. AI TR—266, MIT Artificial Intelligence Laboratory, Cambridge, Mass., 1972
〔3〕 Collins, A. M. and Quillian, M. R. Journal of Verbal Learning and Verbal Behavior, 1969, 8. 240—247
〔4〕 Fillmore, C.: The Case. for Case. in E. Bach, R. T. Harms (eds) Universals in Linguistic Theory, Holt, Rinehart and Wiston, New York, 1968
〔5〕 Lindsay, P. H. and Norman, D. A. Human Information Processing. Academic Press, New York, 1972
〔6〕 Quillian, M. R.: Semantic Memory. in M. Minsky (ed) Semantic Information Processing, MIT Press, Cambridge, Mass. 1968
〔7〕 Schank, R. C.: Conceptual Information Processing. North-Holland Publishing Company, Amsterdam, 1975
〔8〕 Simmons, R. F.: Semantic Networks, Their Computation and Use for Understanding English Sentences, in R. Schank and K. Colby (eds) Computer Models of Thought and Languages. Freeman, San Francisco, 1973

A CHINESE LANGUAGE UNDERSTANDING SYSTEM ——CLUS I.

Li Jia-zhi, Guo Rong-jiang, Chen Yong-ming

Abstract

This is the first experiment of a Chinese language understanding system constructed on the principle of semantic network. The model simulates a dialog between a pupil and a teacher on the common sense about animals.

This system has limited abilities of reasoning, including induction, making inferences from partial information and from indirect information so that the machine is capable of answering questions which could not be answered depending only upon its knowledge base.

Owing to the fact that word order is an important characteristic of Chinese language, our system, in understanding questions and making answers, adopted "word order analysis" instead of "case analysis" which is used by P. H. Lindsay and P. A. Norman, R. H. Simmons and others.

The program is run on a China-made computer TQ16, supported by BCY programming language, and about 200 Chinese characters written in alphabetic way have been used for the input and output of the experiment.