限定领域中汉语语义求解的方法

——类型逻辑语义学应用初探

高峰 陆汝占

上海交通大学计算机系，上海，200030

fgao913@mail1.sjtu.edu.cn

摘要：类型逻辑语义学是当前计算语义学研究中的热点之一，本文通过对人机对话领域中例句的分析，显示了通常情况下疑问语句、带有补语动词的疑问语句以及带有多义项词的语句获得语义解释的过程。

关键词： 类型逻辑语义学，对话，疑问句，义项

A Solution to Chinese Semantics Acquisition In Specific Domain——An Application of Type Logical Semantic

Gao Feng　　Lu Ruzhan

Department of Computer Science and Engineering

Shanghai Jiao Tong University , Shanghai , 200030

ABSTRACT: Type Logical Semantics is one of the hot spot in computational semantics currently. This paper shows the processing of semantics acquisition of the question clauses and question clauses with complementary verb in usual condition, and gives the solution to determining the interpretation of polysemantic word.

Keywords: Type Logical Semantics, Dialogue, Question clause, sememe

1.引言

在人机交互时采用自然语言作为交互手段是目前研究领域的热点问题之一，这样自然语言处理的最终目的就是了解语句在语言环境中的含义，使计算机能够以此为依据进行反馈。从目前的情况看，计算机对自然语言语句理解的准确程度距离实际应用中的需要还有相当大的差距，在中文信息处理中这个问题尤其突出。一方面，现代计算机技术是面向以英文为母语的群体建立起来的，使用汉语的用户与计算机之间多了一层交互时的障碍；另一方面，汉语的句法系统尚未形成完整的理论，有关词的定义、词性的划分等问题长期得

不到彻底解决，这使得传统的句法分析——语义分析——语用分析的解决方法没有坚实的基础。本文把以类型/范畴理论为基础的类型逻辑语义学作为汉语词汇-短语语义分析的手段，通过在限定领域中的应用，希望能在解决中文语义形式化的问题上做出有益的尝试。文中第一节将简单地介绍背景知识（有关类型逻辑语义学详细的介绍见文献[1][3][4]），并结合交通查询领域对其中的词汇类型进行分析，第二节用一些例句展示如何用通过 Lambek 演算求取查询领域中汉语疑问句的语义。

2.介绍

2.1.背景

类型逻辑语义学是目前计算语义学领域中的热点之一，其主要部分为范畴语法。最早的工作是由 Lambek 提出的 Lambek 演算，为范畴语法提供了一个演绎系统，Montague 建立起来的 Universal Grammar 理论，首次把句法范畴和语义类型联系起来，并且在 Montague 语法中演示了如何利用数学和逻辑对自然语言进行形式化的工作[1]。但目前将范畴语法应用到实际的自然语言处理中，还存在着不少困难，主要是范畴的自动生成和演绎系统的演算规则方面[2]，而在汉语处理领域，就目前所知还没有人进行这方面的具体工作，主要的原因（同时也是中文信息处理中的难点）是词的界定和词性的划分问题。本文所作的工作是在限定的领域中用类型逻辑语义学作为求解语义的手段，由于汉语是意合的语言，因此这种类型驱动（type-driven）自然语言分析对解决汉语处理遇到的问题更具有特殊的意义。

2.2.人机对话领域中的语句

在人机对话过程中，用户的目的是获得某一领域中的信息，这样它必须先向计算机描述他所关心的领域的具体特点，这些特点的具体内容并不一定能够在对话的初始有用户一次就完整地给出，按照心理学的规律，用户最初描述的是他所最关心的信息，其它的信息往往需要计算机对用户进行询问才能得到答案。因此整个对话的过程实际上就是将用户最初的不确定的信息内容，通过自然语言处理的手段，求解出其具体的确定的内容的过程。如用户在查询出行的交通路线信息时，需要提供给计算机的信息包括：出发地点、目的地点、希望使用的交通工具以及他所认同的路线是否最佳的判别标准等等。让用户一次描述如此多的内容显然不符合汉语的习惯，正确的方式是由用户——计算机——用户之间的交互进行，计算机的作用一是诱导用户描述计算机必需的信息，二是对用户描述的信息进行理解，产生出新的诱导（询问）语句。整个过程结束的时候，也就是所用为确定的信息（在自然语言中就是词的意义）得到确定的时候（即每个多个义项词都得到了该系统中的唯一解释）。

以交通路线查询为例，当用户的表述为："从地点 A 到地点 B 怎么走"时，"怎么"一次的意思是指：使用什么样的交通工具，交通路线，该路线的起终点（如果是公交线路的话）等等；而当用户的表述为："从地点 A 到地点 B 怎么乘公交车"时，"怎么"的意思就已经限制在公交车的路线名称和起终点上面了。

仍以上面说的交通路线查询为例，假设用户在描述他所希望的交通工具时使用的是"车"，这个概念在交通领域具有多个外延性的指称，可以泛指"机动车"或"非机动车"，也可以具体一些，指称"自行车"、"三轮车"、"助动车"、"摩托"，或者是"公交车"、"出

租车”，也可以指私人用有的轿车。具体将“车”的概念指称到哪一级，要视用户所需要的服务内容和计算机所能够提供的服务程度而定。尽管计算机生成的向用户进行查询的对话也要受到这个概念所指称的内容的约束，但由于本文讨论的主要是分析已有的自然语言语句，并不考虑自然语言生成的问题，因此我们只是把询问语句设置成关于该领域最普通的提问方式。

3.语义求解

3.1.规则和基本定义

范畴：

在范畴语法中，每个句法范畴对应一个相应的语义类型，整个范畴的集合是由基本范畴的集合和定义在其上的递归生成规则所构造而成的。

定义：句法范畴的集合 Cat 是由下列规则递归生成的最小集合：

a. 基本范畴集合={N,S}⊆Cat

b. 若 A,B∈Cat,则 B/A,A\B∈Cat

规则：

非结合的 Lambek 演算 （Non-associated Lambek Calculus , NL）的自然推导形式如下所示：

(1) a. $A \Rightarrow A$ id　　b. $\dfrac{\Gamma \Rightarrow A, \Delta[A] \Rightarrow B}{\Delta[\Gamma] \Rightarrow B} Cut$

(2) a. $\dfrac{\Gamma \Rightarrow A, \Delta \Rightarrow A \backslash B}{(\Gamma, \Delta) \Rightarrow A} \backslash E$　　b. $\dfrac{(A, \Gamma) \Rightarrow B}{\Gamma \Rightarrow A \backslash B} \backslash I$

(3) a. $\dfrac{\Delta \Rightarrow A \backslash B, \Gamma \Rightarrow A,}{(\Delta, \Gamma) \Rightarrow A} \backslash E$　　b. $\dfrac{(\Gamma, A) \Rightarrow B}{\Gamma \Rightarrow B / A} \backslash I$

(4) a. $\dfrac{\Gamma \Rightarrow A \cdot B, \Delta[(A, B)] \Rightarrow D}{\Delta[\Gamma] \Rightarrow D} \cdot E$　　b. $\dfrac{\Gamma \Rightarrow A, \Delta \Rightarrow B}{(\Gamma, \Delta) \Rightarrow A \cdot B} \cdot I$

范畴词典：

范畴词典定义了系统中所用词汇的语义逻辑式、句法范畴。本文例句中使用的词汇范畴为：

基本范畴：

外滩、广场、……	——	**外滩*、广场*、……**	:	**N**
公共汽车、公交车……	——	**bus***	:	**N**
轿车、出租车、的士……	——	**car*、taxi*、……**	:	**N**
自行车、脚踏车……	——	**bike***	:	**N**

非基本范畴：

从、打、在	——	$\lambda x\, V_{from}^{*}(x)$	:	**VP/N**

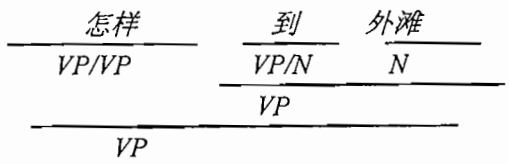
注：这里将介词视为动词是为了处理的方便。

到、去	——	$\lambda x\ V_{to}^{\bullet}(x)$	: VP/N
乘、坐、……	——	$\lambda x\ V_{by}^{\bullet}(x)$	: VP/N
怎样、怎样、如何、……	——	λVp（？Vp）	: VP/VP

汉语中带有“怎么、怎样”等疑问词的句子的句义依赖于句子所处的情景，如前面的分析所示，我们用？表示疑问的语义指向。考虑到该领域中口语表述的习惯方式，为了处理的方便，2 价动词“乘、坐、到”等都被视为 1 价动词，即忽略它们的施事成分。

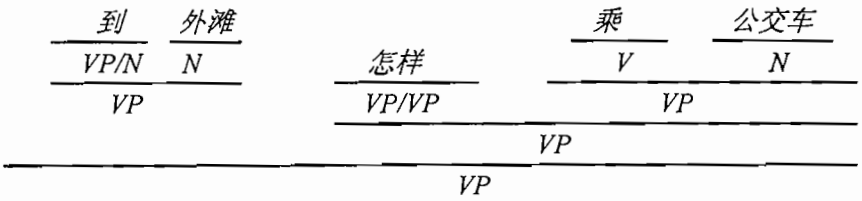
3.2.含有疑问词的疑问句处理

例句 1：*怎样 到 外滩?*



①怎样	——	λVp ？Vp
②到	——	$V_{to}^{\bullet}$
③外滩	——	外滩$^{\bullet}$
④到外滩	——	$V_{to}^{\bullet}$(外滩$^{\bullet}$)
⑤怎样到外滩	——	λVp （？Vp） $V_{to}^{\bullet}$(外滩$^{\bullet}$)
	——	？$V_{to}^{\bullet}$(外滩$^{\bullet}$)

例句 2：*到 外滩 怎样 乘 公交车?*



①到	——	$V_{to}^{\bullet}$
②外滩	——	外滩$^{\bullet}$
③到外滩	——	$V_{to}^{\bullet}$(外滩$^{\bullet}$)
④怎样	——	λVp (?Vp)
⑤乘	——	$V_{by}^{\bullet}$
⑥公交车	——	bus$^{\bullet}$
⑦乘公交车	——	$V_{by}^{\bullet}$(bus$^{\bullet}$)
⑧怎样乘公交车	——	λVp (?Vp) $V_{by}^{\bullet}$(bus$^{\bullet}$)
	——	? $V_{by}^{\bullet}$(bus$^{\bullet}$)
⑨到外滩怎样乘公交车	——	$V_{to}^{\bullet}$(外滩$^{\bullet}$)∧? $V_{by}^{\bullet}$(bus$^{\bullet}$)

3.3.补语动词的处理

汉语疑问句的疑问词可以出现在句首，也可以出现在句中或句尾，这有别于英语中的wh-结构。但疑问词出现的位置对整个句子的句义是有影响的，如在交通查询系统中，疑问词出现在句尾时，整个句子的意义就超出了该系统所能理解的范围（例句 2）。

例句 3：*到 外滩 怎样?*

例句 4：*到 外滩 怎样 去?*

我们在此所需要考虑的是句首的疑问词移位至句中的情况（例句 4），这时疑问此后必须加补语动词才能保持句义不变，因此我们对范畴词典进行修订，参照文献[1]中关于polymorphism 问题的处理，所依据的 NL 中的演算系统中也应该增加相应的规则：

NL(增加的规则)：

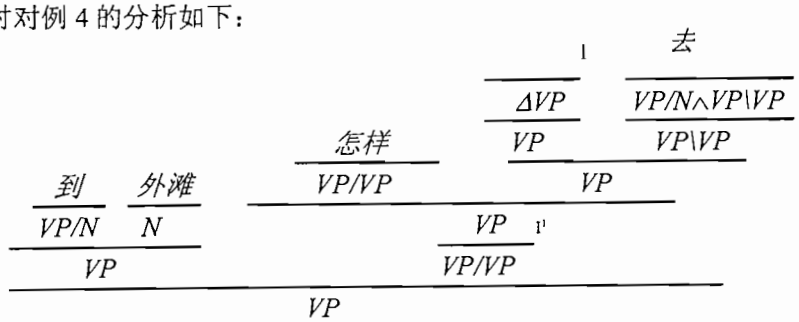
$$\text{a)}\ \frac{\Gamma \Rightarrow \phi : \mathrm{A} \quad \Gamma \Rightarrow \psi : \mathrm{B}}{\Gamma \Rightarrow (\phi,\psi) : A \wedge B} \wedge R$$

$$\text{b)}\ \frac{\Gamma[x:A] \Rightarrow \chi[x] : \mathrm{C}}{\Gamma[z:A \wedge B] \Rightarrow \chi[\pi_1 z] : C} \wedge L_a \qquad \frac{\Gamma[y:B] \Rightarrow \chi[y] : \mathrm{C}}{\Gamma[z:A \wedge B] \Rightarrow \chi[\pi_2 z] : C} \wedge L_b$$

范畴词典(增加的词项)：

去、走、…… —— **λVp Vp** ：**VP\VP**

这时对例 4 的分析如下：



①到	——	V_{to}^{*}
②外滩	——	外滩*
③到外滩	——	V_{to}^{*} (外滩*)
④怎样	——	λVp(?Vp)
⑤去	——	λVp Vp
⑥v	——	ΔVP
⑦v 去	——	λVp Vp ΔVP
	——	ΔVP
⑧怎样 v 去	——	λVp (?Vp) ΔVP
	——	? ΔVP
⑨怎样去	——	λVp (?Vp)
⑩到外滩怎样去	——	λVp (?Vp) （V_{to}^{*} (外滩*)）
	——	?V_{to}^{*} (外滩*)

3.4.词汇的语义义项处理

我们在第一节中提到一词有多个外延指称的情况，当求解其词义时，有些是需要从显式的语句中求出，有些则没有显式地说明，因为该词的词义已经由动词或形容词确定，这时如果再要求用户对词义进行说明就显得对话不够自然。如："骑车"中的"车"的词义已经明确为"自行车"，"几路车？"中的"车"也已经明确为"公交车"，"哪路车?"中的"哪"的指向是一个代表公交车路线序号的数字或名称，等等。

我们的解决方法是在词典中增加这些动词、形容词对多义词的语义义项描述，如下所示：

范畴词典(增加的词项)：

骑、蹬、……	——	$\lambda x\,(V_{by}(x) \wedge (x = bike^*))$	:	**V**
开、驾、……	——	$\lambda x\,(V_{by}(x) \wedge (x = car^*))$	:	**V**
路、线、……	——	$\lambda x \lambda y\,(No^*(x) \wedge (y = bus^*))$	:	**A**
几、多少、什么、……	——	**?x**	:	**AP/A**

这样我们就可以解决类似例 4、例 5 的问题：

例句 5：*到 外滩 骑 车 怎么 走?*

例句 6：*到 外滩 乘 几 路 车?*

分别得到：

例 5： $V_{to}^*(外滩^*) \wedge ?\ V_{by}^*(bike^*)$

例 6： $V_{to}^*(外滩^*) \wedge V_{by}^*(No^*(?x) \wedge bus^*))$

限于篇幅所限，例 4、例 5 详细的分析过程从略。

4.总结

现代汉语语义形式化的工作目前还处在一个初级的阶段，在本文所提到的领域中，由于短语结构的前后顺序非常灵活，如果按照通常的方法用产生式书写句法规则将非常复杂，在句法分析阶段就会遇到很大的问题，从而影响以后的语义分析。在研究中引入类型逻辑语义学的理论为解决类似问题提供了一种新的手段，同时也注意到，本文讨论的语言现象和采用的例句只是在一个较狭窄的领域中，并且为了处理的方便，对某些词汇功能进行了调整，这样的调整是否具有普遍意义，还需要在实际中进一步验证。

在实际应用中，最主要的工作是建立范畴词典，目前的方法是由人工从收集到的例句中提取句法、语义信息建立词典，其缺点在于描述时没有区分领域专用词和一般常用词，这对于系统的可扩展性有很大影响，因而分别建立这两种不同词汇的范畴词典也是下一步的主要工作内容。

参考文献

[1]GLYN V.MORRILL, 《Type Logical Grammar-Categorial Logic of Signs》, KLUWER ACADEMIC PUBLISHERS

[2]AARNE RANTA,《Type-Theoretical Grammar》, CLARENDON PRESS.OXFORD

[3]ROBIN COOPER , KUNIAKI MUKAI, JOHN PERRY, 《Situation Theory and Its Applications》,CENTER FOR STUDY OF LANGUAGE AND INFORMATION

[4]PATRICK SAINT-DIZIER, STAN SZPAKOWICZ, 《Logic and Logic Grammars for Language Processing》,ELLIS HORWOOD LIMITED,1990

[5]于江生，《范畴语法简介》，北京大学计算语言学研究所，http://icl.pku.edu.cn