多语信息交流平台的中间语言系统及支撑环境设计*)

熊文新[1] 宋 柔[1] 袁 琦[2]

(北京语言大学语言信息处理研究所 北京 100083)[1] (中国电子信息产业发展研究院 北京 100044)[2]

摘 要 探讨了中间语言充当多语信息交流平台基础架构的必要性和可行性,介绍了一个基于中间语言的多语信息处理平台的总体设计和实现策略。讨论了如何在构建中间语言系统过程中引入子语言、受限语言技术,中间语言系统在网络环境下的XML实施,以及系统实施过程中的多引擎处理策略和学习反馈模块等支撑环境建问题,最后给出了一个示例在系统平台中运行的过程。

关键词 中间语言,支撑环境,多语言信息

The Design of a Multi-Lingual Communication Platform Based on Interlingua

XIONG Wen-Xin[1] SONG Rou[1] YUAN Qi[2]

(Center of Language Information Processing, Beijing Language and Culture University, Beijing 100083)[1]

(Research Center of Computer and Microelectronic Industrial Development, MII, Beijing 100044)[2]

Abstract In this paper, the necessity and possibility of Interlingua as the architecture of Multi-lingual information communication platform is discussed, and the overall design and implementation strategies of building such as platform based on Interlingua are introduced. Introduction of the technology of sublanguage, controlled language during the process of Interlingua system, XML representation of Interlingua in networking environment, multi-engines and self-study modules during the processing are also presented. An example will be elaborated to illustrate the workflow of the described system.

Keywords Interlingua, Multi-communication, Supporting system

1 引言

伴随相关理论研究和工程实践以及新兴的网络数据交换标准的推进,基于中间语言的处理系统开始获得重视,尤其是在机器翻译领域。比如美国机器翻译协会(AMTA)就专设中间语言特别兴趣组(SIG),团结了大批高等院校和工业组织的研究机构,并一直进行着有关的理论研究和实践探索。相关成果不断涌现,诸如卡耐基梅隆大学(LTI/CMU)在基于知识的翻译系统KANT的基础上推出新的Kantoo[4]、南加州大学(ISI/USC)等的Pangloss以及新墨西哥大学、马里兰大学的探索。近来尤为突出的是联合国大学高等研究院(IAS/UNU)在早期Atlas和多国语言机器翻译系统MMT基础上推出的通用网络语言(UNL)[6],吸引了工业界和学术界众多的国际合作。

一个值得重视的趋势是在中间语言研制当中,尤其是知识构造中,越来越多的手段,像获取知识、知识表示及利用等多种方式被引入进来,知识建造的精度和准确度越来越高,从而摆脱了长期以来中间语言方法只能用于建造玩具系统的观念。知识库并非中间语言系统所专有,几乎所有人工智能系统都需要知识库建设(只不过知识形式各异、获取手段不同而已),因此,建造中间语言知识系统所积累的方法和手段,也将促进整个语言信息处理的发展。

2 中间语言的设计模式

根据网络时代日益增长的多语信息交流需求,以中间语言系统为基础,结合其他语言处理方法和网络处理手段,我们构建了一个基于中间语言的多语信息处理平台。该平台使用户能够方便地将自己编制的信息转化成中间语言格式,实现网上"一旦构成,各语适用"目的。在考虑平台总体设计时,主要考虑了如下几点。

2.1 中间语言简化处理与工程实现

采用中间语言,可以大大减少工程开发的工作量。假定采用直接转换法,设有n种语言需要进行互译,需$n\times(n-1)$个模块,而中间语言方法,则仅需$2\times n$个模块。中间语言剥离了直接转换法中的源语分析和目标语生成捆绑在一起的模块,只要中间语言体系构造良好、表述准确、处理方便,就能较好发挥多语交流的中心平台作用。

并且,如果中间语言体系构建得好,就其作为自然语言理解终点和生成源点的知识系统而言,对于自然语言处理的其他系统的知识库建设都将有着直接利用价值或间接借鉴作用,这也是我们在该项目中着力于中间语言建设的原因之一。

2.2 中间语言架构与子语言(Sublanguage)建设

中间语言方法一般多采用基于知识(knowledge-based)的处理策略。用作知识处理时,作为其源语言的中间语言必须构拟良好,具有语法语义上的明晰性,并且能较好地满足知识推导的要求,亦即中间语言自身应是一个完备的语言系统。

2.2.1 纯粹完整的中间语言尚不可行

虽然中间语言经过精心构造,理论上可以做到无歧义,然而在表述精确性和精简性之间还需要折衷处理,这是因为中间语言必须能够担负多语信息交流的媒介作用。仅就单一语种的自然语言而言,尚未有一个百分百完全可实用的分析生成系统见诸报道。更由于不同语言在句法、语义、语用等诸层

*)国家自然科学基金项目(编号69902003,60372106)。熊文新 博士生,主要研究方向:机器翻译、语言工程;宋 柔 教授,博士生导师,主要研究领域:中文信息处理、语言工程;袁 琦 教授级高级工程师,主要研究领域:机器翻译、中文信息处理。

面并没有一一对应关系，企图构建一个针对世界知识无所不包的中间语言体系，虽然是人们梦寐以求的理想，然而在现阶段的实现过程中无疑并不现实。

2.2.2 中间语言的构造

我们曾经参与的由联合国大学高等研究院的 UNL 项目[8]中，其中间语言主要是基于概念语义的体系。该系统定义了 41 种语义关系，其中既涉及与谓词框架有关的体词概念类格关系，如典型的施事 agt、受事 obj、目标和终点 gol，同时有涉及事物类概念之间联系的修饰关系 mod，甚至还有表示事件间关系的顺序 seq 等。除了语义关系，还定义了数十个附着于关系或中间词语基础上的各类属性，如加载于动作概念上的时态属性，例如{@present, @past, @future}；有表示体貌的诸如经历@experience、开始@begin 等，还有体现说话人焦点、意图、态度等内容的属性。整个中间语言系统通过无歧义的表示概念意义的中间词汇和概念意义之间的关系和属性来构成。这个体系完全是基于语义关系基础上的。

在总结以往经验的基础上，我们在中间语言多语信息交流平台中设计的中间语言，并不抛弃目前较为成熟的句法分析。在中间语言词汇中，我们以每一个词语的义项 entry 为基础，同时，对中间词语的表示仍然记录其固有的句法属性和词法信息特征，如词类和构词法信息。在中间语言体系中，除了前此的语义关系和属性内容外，句法信息内容仍然作为进行分析和生成的基础，句法语义接口仍是其中的核心利用之一。中间语言能够表示分析结果的层次性，自然语言分析的树型结构也能方便地体现。

2.2.3 划分子语言

根据直觉，领域知识是确定词汇概念和语言系统的重要参数。假定构拟的通用中间语言体系是一个语言全集，通过引入领域属性，则可以确定一个特征函数，并进而划分一个语言的子集。其形式如下：

$$Language\_select(x)=\begin{cases}1, & \text{if } x\in Domain\_attribute(A)\\ 0, & \text{if } x\notin Domain\_attribute(A)\end{cases}$$

如果待确认的元素 a 具有 A 领域的参数区别特征，则语言系统可以调整为领域 A 的一个子语言集 sublanguage。同理可建立语言风格参数 Style_variation 集。通过引入参数设置，在处理原有中间语言集时，就能调其对应子语言子集，进行相关处理，从而获得较好效果。例如，针对英文的“mouse”，在确定文档隶属计算机硬件设备领域，则可以选取其中文含义为“鼠标”，而在生物学领域则大抵可归纳到中文含义“老鼠”。在中间语言概念意义构建过程，我们采用多种句法语义信息，同时尽可能地包含了语体和领域等可以用以消歧的属性信息。由于自然语言的复杂性，目前我们并没有全面开展各参数的子语言集的配置工作，而是局限在科技领域的正式文体。

2.3 中间语言表示体系与网络环境支持

XML 语言凭借方便的可交互操作性和众多的开发解释支持工具成为网络信息最好交互形式之一。在自然言处理界，人们或通过 XML 来学习句法规则[3]，或用其改造原有的机器翻译流程[7]，或进行双语翻译对齐的语料库和短语库的知识表示[2]，并且都取得了预期效果。可以说，XML 的实施利用大大方便了语言工程的后续工作。

结合美国机器翻译协会（AMTA）中间语言可读性工作组（Interlingua Readability workshop）制定的编码实践和关于双语语料库的标注体系[2]，我们规划了一个能够标注句法语义关系和中间结果的 XML 表示，使得采用中间语言方法表示的信息内容能够较方便地反转回自然语言形式，或交由其他自然语言处理系统调用。

举一个以中间语言表述的自然语言句子片断，设想要传达“人们建造巴别塔的准备就绪”这样一个想法。图 1 表示了关于这种想法的 XML 表述。

```
1 〈doc document_id="Document" author_id="Author"〉
2 〈language code="CHN-SIM" /lang〉〈nation nation_id="CHN" /
  nation〉
3 〈style style_id="formal"/style_id〉〈domain domin_id="pop" /
  domin_id〉
4〈sent id="1"〉
5 〈NODE=1, HEAD=2, syn=sbj, sem=agt, lex="people(icl>hu-
  man)"〉人们〈/NODE〉
6 〈NODE=2, HEAD=2, syn=entry, sem=entry, lex="begin(obj>
  event)"〉开始〈/NODE〉
7 〈NODE=3, HEAD=2, syn=obj, sem=obj, lex="build(gol>
  thing)"〉建造〈/NODE〉
8 〈NODE=4, HEAD=2, syn=obj, sem=obj, lex="a(icl>quanti-
  ty)"〉一〈/NODE〉
9 〈NODE=5, HEAD=5, syn=mod, sem=mod, lex="huge(isa>at-
  tribute, atr)thing)"〉高〈/NODE〉
10 〈NODE=6, HEAD=3, syn=man, sem=etr, lex="tower(icl>
  building)"〉塔〈/NODE〉
11 〈/sent〉
   ……
12〈/doc〉
```

图 1 中间语言分析结果的 XML 表示

以上 XML 表述表明，这是由中国作者编辑的文档（见行 1），使用的是简体汉字，对应的语言是汉语普通话（见行 2），文体风格是正式体，属于科普领域（见行 3）。行 4～行 11 是该句的句法语义表示。NODE 表示在当前词语在原句中的线性位置，HEAD 表明与当前词语相关联的核心成分的线性位置；SYN 与 SEM 表明当前词语与核心词语之间构成的句法和语义关系；行 6 的 entry 表明该节点是本句的入口，由中间语言词语知识库登载的词语信息，可知这是一个带动词宾语的动词谓语句（词语信息由于具有唯一性，在处理过程中，可直接通过 XML 语句中的 lex 体现的中间词语到词汇库中获取）。其句法语义关系可以从图 2 看出。

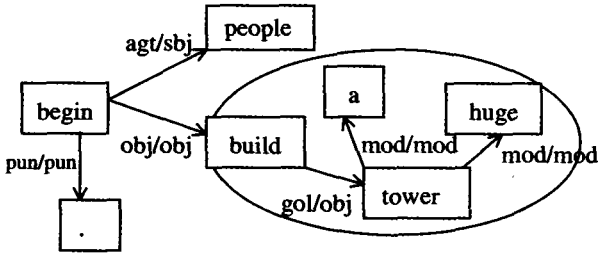

图 2 中间语言分析结果的句法语义描述关系

图 2 是图 1 中间语言表示的句法语义关系图。其中，箭头引出方向是从谓词（predicate）到其论元（argument），从被修饰成分（modified）到修饰成分（modifier）。图中有向边标示的“sem/syn”分别代表连接有向边的两个节点之间的句法语义关系。中间语言以意义内容传递为主要内容，故而语义关系是其核心表示；而自然语言的形式和意义是一个矛盾统一体，相同的语义关系，或由于语用因素的渗入，在表层语言形式上可能存在不同的表现手法。为体现出这种细微差别，反映出源语言的特性，我们引入了广义的句法关系（之所以称为广义，是我们把主题等所谓的语义概念也纳入到主语的范畴中，虽然可能在理论上有争议，但它们体现在汉语表层的线性序列上，在某些方面有着类似的体现，因而在应用系统的实现上起到简化分析的作用，这是我们面向应用在工程实践上的一种折衷处理），供生成目标语言语句时作为参考。而由于

XML 表示可以记录树结构中不同节点，作为指针，可以获取相关上下文信息，解决诸如回指和复指等内容，从而实现在篇章范围内的语言处理。因此，虽然我们把处理范围目前仍局限在句子范围，但跨句处理在人机交互的情况下仍是可能实现的。

图中，我们还可以发现与分析入口的核心谓词发生宾语关系的是一个事件类，因而其关涉对象又可扩展为一个新的事件，比如还可由其入口动词 build(gol＞thing)创建一个新的关系网。在某一个事物类概念内部与其他成分之间还可能存在修饰、限定等不同的联系，形成一个复杂的体词性概念结构。通过可以递推的核心动词和与其关联的论元，及其各论元之间的关系，能够表示复杂的多重嵌套的事件内容。

2.4 多引擎引入与自学习机制

多年机器翻译的历史经验证明，采用规则转换生成的译文，效果并不是很好[11]。而基于实例的翻译，效果却令人满意，因为翻译实例源自真实的自然语言译文。而规则处理对于自然语言各类多变的表达方式，颗粒度太粗，难以刻画，造成机器翻译的味道浓重。适时引入译文自然度较高的实例法，将较好地提高信息交流的可接收度。

我们在系统平台设计中采用多引擎的处理策略。在基于规则处理的中间语言转换前端，加载基于实例的中间语言转换模块。通过接收用户输入端的自然语言语句，转换成中间语言表示。如果能从已有实例知识库中取得完全匹配的正解，则直接输出；否则根据输入内容与实例内容之间的句法、语义关系距离，计算其相似度。如果大于某一给定阈值，则直接生成译文结果；否则继续调用规则处理方式，实施下一步的中间语言转换。可以看出，在处理流程上，当实例引擎不能获得最优解时，系统才蜕化成纯粹的基于规则的中间语言处理方法。

无论是经过实例生成的结果，还是中间语言规则处理的结果，都将回显给用户。显示方式既可以是由中间语言逆转换回来的源语言自然语言语句形式，也可以直接是中间语言表达式自身。后者更适合经过训练的专业员。用户可以针对中间语言的反馈结果进行调整，直至对系统输出的转换内容满意为止，并将最终校正结果存入后端知识库系统中。

采用多引擎的交互式处理，化解了自然语言分析和生成的复杂性，如输入句子句法语义的分析、词义消解、高自然度译文语句的生成。而在构建知识库系统过程中，将确定已有译文结果的对齐等，同时也将促进自然言处理知识的发展。

3 中间语言系统的实现

鉴于自然语言分析技术的前端目前尚不足以保证机器翻译及其他语言工程系统的需求，尤其对待汉语的分词、词性标注及词义排歧等，困难尤甚；在作为底层的基础分析的准确性没有得到彻底解决的情况下，先天不足的分析将直接影响到后续处理的输入，由此造成的累积错误将使语言处理系统准确率难以保证。这正是当前全自动机器翻译面临的困境之一。

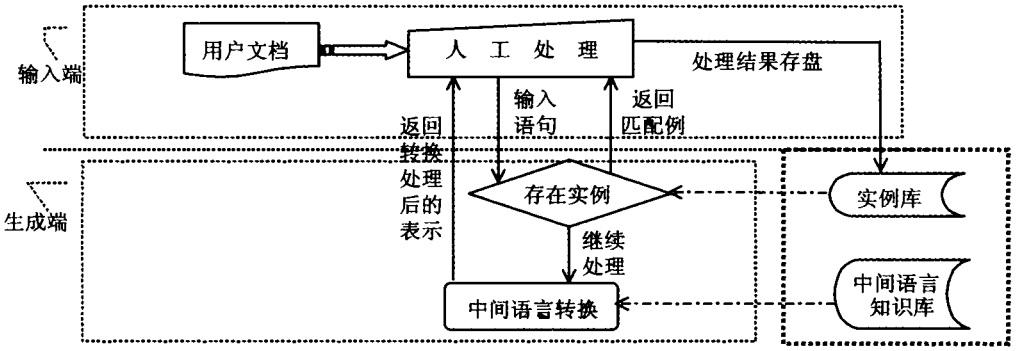

图 3 中间语言多语交流环境工作流程

为避免重蹈全自动机器翻译之覆辙，我们在系统中设计嵌入前端的输入编辑环境，引入受限语言技术，希望通过人机交互的方式来达到最终提升中间语言分析结果的准确性。图 3 显示了该平台环境的工作流程。

3.1 输入端

在输入端，用户既可以像传统机器翻译那样，直接向系统提交用母语书写的句子，系统分析器将其转换为中间语言表示的结果形式，并按用户需求，回显中间语言结果或经中间语言重新转换回母语的句子，以便于用户稽核系统是否真正理解源语信息。用户对结果不满意时，可进行修订加工，直至系统反馈内容满足用户需求为止。

用户同时可以利用系统提供的编辑环境进行处理。系统根据中间语言知识库系统中已经存在的知识内容，帮助用户选择合适的表达方式。比如针对知识库中缺少的词汇内容，系统将报警，并向用户提出选择合适的概念语义类供参考，或提示用户是否有其他表述方式。而当用户输入的是一个具有多个中间语言词语相对应的词，则提醒用户从系统选取的候选词集中选取合适的表示。针对存在多种分析结果的短语结构和句型，系统将根据从真实文本中统计出现的频度信息，按照高频先现的原则，列出最大可能的结构形式，供用户进行选择。

所有这些都是为了解决全自动分析时的黑箱操作，使系统分析结果能够明白地由用户操纵，通过交互操作，最大限度地保障待交流的信息内容与用户的需求相吻合。作为中间语言知识库中的规则规范是建立在受限语言研究的基础上实施的。用户输入端解决的是将自然语言映射到中间语言的一对多的情形。在语言处理技术尚达不到完全自动准确的情况下，由系统处理其能处理的那部分，而将存疑内容交由人工交互完成，是提高语言处理效率和质量的关键。这也是我们引入受限语言处理技术的目的所在。

3.2 后端知识服务器

为充分提高系统对源语言的分析效率和准确性，我们在后端分析处理器中引入翻译记忆(TM)技术。后端系统具有包容多种知识源的功能，既有对源语言的分析和生成处理，又存储有具体的源语-中间语言翻译对实例。

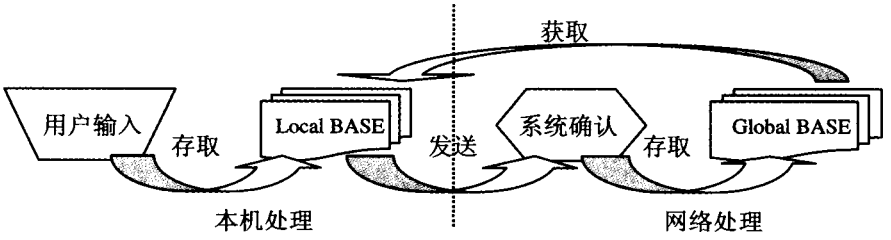

图 4　翻译记忆知识库处理流程

参照图 4，当用户在输入端确认系统接收用户输入反馈的中间语言表示后，系统将自动记录源语言语句与中间语言表达式的对应实例，并以数据库方式存入个体用户级知识库系统的实例库 LocalBASE 中。以后倘若有类似的待翻译语句，将直接从已实现的实例库提取，优先呈现给用户，而不重新经过源语言分析，除非用户明确对反馈回来的实例结果表示拒绝。

实例知识库系统分为本地和远程两类。远程实例库 GlobalBASE 一般是已经被系统管理员确认的系统级知识库。当用户确认处理结果后，系统将对齐的实例对存入以用户自身账户设立的 LocalBASE 中，供用户后续翻译的处理。缺省情况下，实例知识库是作为翻译用户本地使用的；如果用户希望这种知识能够全局共享，则可以将其发布到翻译服务器，经过系统管理员确认后，存入系统知识库。一旦进入系统知识库，则能被授权访问该网络的用户使用，或下载到用户本地机上使用。用户如果找不到本地实例，可以通过 Web service 方式自动访问 GlobalBASE。

3.3　生成端

生成系统是在接受用户输入后，经由分析器转换成中间语言的内部表示之后，根据内部实例库（用来实现前一层次的实例生成）和中间语言知识库（根据中间语言词汇和输入语句上下文实现环境中赋予的动态关系和属性，依据生成规则）生成用户指定的目标语言形式。

虽然生成模块接受的输入是中间语言内部表示，由于不同自然语言之间的句法语义特性，不同语种的输入者对于系统反馈的中间语言中间结果表示会有不同的认知倾向。比如说汉语的人，对中间语言体系中表示动作行为的时体、名物类的单复数属性并不敏感，因此即便经过交互式处理，用户对系统反馈回的中间结果表示或用母语表示的语句表现出的这些属性的遗漏，也不会有所反映。对其他对此信息敏感的语言使用者来说，不可避免地导致了信息传递的缺损。在 UNL 系统中，由于不同国家、不同语言之间生成的中间语言编码有各自的“特色”，Boitet 描述 UNL 项目时，曾经就多语翻译间的中间语言应用提出自己的看法 [1]。

因此，在构建从中间语言到目标语言的转换时，为保证鲁棒性，有必要引入对用户处理的容错机制，以满足实际应用之需。我们曾在参与 UNL 项目时也提出过相应的解决措施 [9]。

由于在汉语句法规则方面，宾语的种类可以包含不同的语义类型，如处置的对象（类似于 obj）和从无到有的结果（类似于 gol）。因此，在向汉语转换的过程中，生成具有这两种语义类型的关系时，自然语言的线性序列同样都是由核心动作指向到其指涉的体词概念。这两种不同的语义关系在表层体现上得到中和（neutralization），句法实现上都是“V＋N”形式。根据自然语言线性排列及其深层语义关系的一对多的特性，我们的生成规则能够处理类似情况，使得规则处理得以化简。

汉语与其他语言不一样之处还有诸如量词的处理。西方语言有的是“数（量）”的范畴，而缺乏量词这一观念。全面引入量词到中间语言体系中，对西方语种的使用者并不习惯，因而他们生成的中间语言关系中，往往容易缺损这一信息。纯粹仅仅唯中间语言格式规范而从之，那么或者与原文编辑者的语感不符；如果不加入该信息，则又将影响到汉语的生成质量。我们采取的策略是将“量”的概念载入中间语言词汇集中，在将汉语词语与中间语言词语挂接（Linking）的过程中，根据其对应关系，在中间语汇反映的体词类概念中，加入其对应汉语词的量词属性特征。在由中间语言向汉语生成的过程中，根据其他语种中带有“数（量）”关系的体词语汇，追加相应的汉语量词，从而使得在不增加中间语言的关系和属性的情况下，能够生成较为通顺的汉语句子。

在针对汉语的生成端，我们同时为开发用户提供了一个集成的生成调试辅助环境，能够实现对汉语词语与中间语言词汇之间挂接时的去重和排序；能够对待转换的中间语言输入实施中间语言词汇的合法性检查，找未良好定义（well-defined）的词语形式和中间语言词库中未发现的词汇表示；能够检查输入的中间语言表达式自身撰写的规范合适性；能够提取生成过程中使用的规则及其频次和使用的上下文环境；能够浏览、检索、删除、修改生成规则库中的规则内容，从而创建一个良好的规则库系统。

目前，我们根据前此开发的汉英、英汉机器翻译系统的英语和汉语生成模块，实施了这两个语种的生成工作。

4　一个示例

以下通过一个实例，说明我们如何利用多语信息交流平台向 Internet 发布信息。

4.1　输入检查和中间结果校核

用户使用系统时，可以根据需要选择操作模式是采用自动转换方式还是交互式处理。如果选择前者，系统读入用户输入之后，将调用转换模块转换成中间语言表达式；后续操作可以进一步分为回显结果是中间语言内部表达，或是输入语言形式的表示。如果选择交互处理，则在用户输入信息的同时，系统即时读入信息内容，并根据中间语言知识库及时反馈当前输入与知识库的不符之处，并根据具体语境提出解决方案，具有即时校对（checking）的作用。

现在假定目标用户是一个说汉语的信息发布者，并且选择的模式是交互式汉语回现处理，他需要发布的信息内容是“人们开始修建一座巨大的塔。”（为方便起见，我们直接用汉语句子表示用户想要表达的信息内容）。

当用户用汉语键入“人们开始建筑一座高塔”。系统首先扫描用户输入，提示“建筑”具有两个义项，“建造行为 build”（修建、建设、建造、建立、建成）和“建造结果 building”（建筑物）。并根据前此句法分析结果，此“建筑”是“开始”的宾语，而“开始”的论元要求是一个事件类概念，从而推导出该词可

（下转第 266 页）

适应性软件构架采用可复用构件与其它技术，在分布式、异构环境中简化软件进化的复杂性[6]。我们采用了构件技术来设计可接入层的各种学习算法，以简化算法设计和算法进化的复杂性，而且使算法构件对于不同领域的 AIS 应用具有可复用性，以提高 AIS 开发应用的效率和质量。此外，我们的构架也潜在支持分布式免疫计算，这是进一步的工作。

自适应(self-adaptive)系统能根据环境或条件的变化自行改变其结构或行为，无需人工参与[7]。AIS 应该是一种自适应的系统，但目前还难以全面实现，根本原因在于其复杂性。本文从软件构架设计的角度探讨 AIS 的适应性解决方案，虽然其中大多适应性工作需要人工参与，但这种基于构架的途径对于将来的自适应免疫计算具有指导作用。

结论 适应性是 AIS 的一个重要特性，除了学习算法所针对的异物适应性之外，在 AIS 的软件开发应用方面还存在数据源进化和多算法进化等适应性问题。为此我们扩展并改进了已有的 AIS 构架，提出一个新的 AIS 软件构架，以满足这些适应性要求。该构架以基因计算为中心，设计了元基因来适应数据源的变化，并设计了可接入构件和算法验证机制来解决多算法进化的适应性问题。在该构架支持下，数据源的进化或改变不影响学习算法的设计，同时使学习算法能适用于多种数据源。该构架简化了 AIS 软件开发的复杂性，可提高开发应用的效率，对于实现下一步的分布式免疫计算和将来的自适应免疫计算具有指导意义。

参 考 文 献

1 Hofmeyr S, Forrest S. Architecture for an Artificial Immune System. Evolutionary Computation. Morgan-Kaufmann, San Francisco, CA, 2000, 7(1): 1289～1296

2 Ji Z, Dasgupta D. Artificial Immune System (AIS) Research in the Last Five Years. Published in the proceedings of the Congress on Evolutionary Computation Conference (CEC) Canberra, Australia, 2003

3 Cayzer S, Smith J, Marshall A R J, Kovacs T. What have Gene Libraries done for AIS? Digital Media Systems Laboratory, HP Laboratories Bristol, HPL-2005-116, 2005. www.hpl.hp.com/techreports/2005/HPL-2005-116.pdf

4 Kim J, Bentley P J. Towards an Artificial Immune System for Network Intrusion Detection: An Investigation of Clonal Selection with a Negative Selection Operator, the Congress on Evolutionary Computation (CEC-2001), Seoul, Korea, 2001. 1244～1252

5 Kim J, Bentley P J. A Model of Gene Library Evolution in the Dynamic Clonal Selection Algorithm. In: Proceedings of the First International Conference on Artificial Immune Systems (ICARIS2002) Canterbury, 2002. 175～182

6 Aniort P. A Distributed Adaptable Software Architecture Derived From a Component Model. ACM, Computer Standards & Interfaces, Special issue: Adaptable Software Architectures. ISSN: 0920-5489. 2003, 25(3): 275～282

7 Oreizy P, Gorlick M, Taylor R N, et al. An Architecture-Based Approach to Self-Adaptive Software, IEEE Intelligent Systems, 1999, 14(3): 54～62

8 Fayyad U M, Irani K B. Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning, Proceeding of The Thirteenth International Joint Conference on Atificial Inelligence, 1993. 1022～1027

(上接第 259 页)

能的最优解为表示“建造行为”的谓词，即“修建”，并建议用户选择合适的汉语词，供用户确认。

4.2 实例检索

一旦用户同意优选结果，系统将自动选用该结果，表示建造行为的谓词类 build，并在分析出该句的句法语义表示后，在实例库中以该词对应的中间语言概念词语及其语义句法关系为索引，查询相应的实例。设在实例库中，以 build 为索引的实例包括有“创立理论”、“增进情谊”、“建造房屋”等中间结果(虽然在中间语言词汇集中，中间语言词语与汉语的义项对应，但为增强系统鲁棒性，我们有时以光杆词语来索引，以适应不同用户选择不同中间语言词语造成的理解偏差，毕竟对义项的把握并非每个人都感觉一致)。在这些实例中，谓词和体词之间都存在 gol 关系(从无到有地产生某种结果这样一种行为，而体词表示的结果本身)。

(1) gol(build(gol ＞abstract thing), theory(icl＞abstract thing))

(2) gol(build(gol ＞relation), will(icl＞relation))

(3) gol(build(gol ＞building), house(icl＞building))

build 的语义类暂时未列出来。在以上检索到的实例中，theory，will 的语义类别是“抽象物”和“关系”，而(3)中的 house 语义类属于建筑物。与分析所得的“塔”的语义属性同类，因此优先选择(3)确定用户输入词“修建”的中间语言词汇是 build(gol＞building)。而在实例库中，(3)体现的汉语对应表示是“建造房屋”。

4.3 知识积累

根据第 2 步的实例匹配处理，系统将生成“建造塔”，并以汉语形式反馈给用户。如果用户对此予以确认，则系统将“建造塔”存入用户实例库 LocalBASE 中，并将此成果提交发送到给多语平台的后端知识服务器，由有权限的系统管理员确定是否追加到系统平台的知识库 GlobalBASE，作为全局共享的知识。

如果用户不满足系统反馈的结果，可以通过多次交互直至产生所需语句。最终的中间语言处理表示结果如图 2 所示。配合实例参考和规则生成，则形成系统反馈的结果“人们开始建造一座高塔”。而最终形成的各类关系和属性同时将作为翻译成果存入实例知识库中，供后期翻译服务时继续采用。

参 考 文 献

1 Boitet C. Advantages of the UNL language and format for web-oriented cross-lingual applications. Seminar on linguistic meaning representation and their applications over the World Wide Web, 2000

2 常宝宝，詹卫东，柏晓静，等. 服务于汉英机器翻译的双语对齐语料库和短语库建设. 见：第二届中日自然语言处理技术国际研讨会论文. 北京大学，2002

3 Déjean H. Learning Syntactic Structures with XML. Proceedings of CoLL-2000 and LLL-2000, Lisbon, Portugal, 2000

4 Schütz J. One web, One Language: The universal networking language. 2003

5 Mitamura T, Nyberg E. the KANTOO Machine Translation Environment. In: Proceedings of the 4th Conference of the Association for Machine Translation in the Americas on Envisioning Machine Translation in the Information Future, Cuernavaca, Mexico, Oct. 2000

6 Sabarís M F, Alonso J L R, Dafonte C, et al. Multilingual Authoring through an Artificial Language. EAMT Summit VIII, Santiago, Spain. Sep. 2001

7 Senellart J, Boitet C, Romary L. SYSTRAN New Generation: The XML Translation Workflow. In: Proceedings of MT Summit IX, New Orleans, USA. Sep. 2003

8 Uchida H, Zhu M. The Universal Networking Language beyond Machine Translation. International Symposium on Language in Cyberspace, Seoul, Korea. Sept. 2001

9 熊文新. 基于中间[illegible]语生成规则处理. 见：1998 中文信息处理国际会议论文[illegible]学出版社，1998. 515～523

10 熊文新. 中间语言[illegible]程中的增强处理. 计算机工程与应用，2005(9)：171～173