《中文信息学报》稿件排版格式

文章编号：1003-0077（2011）00-0000-00

中国手语信息处理述评 *

姚登峰[1,2,3]，江铭虎[1,2]，阿布都克力木.阿布力孜[1,2]，李晗静[3]，哈里旦木.阿布都克里木[4]，夏娣娜[5]

（1.清华大学人文学院计算语言学实验室，北京 100084；

2.清华大学心理学与认知科学研究中心，北京 100084；

3.北京市信息服务工程重点实验室(北京联合大学)，北京 100101；

4.清华大学计算机科学与技术系智能技术与系统国家重点实验室，北京 100084；

5.工业和信息化部电子工业标准化研究院，北京 100007）

摘要：为了能够有效地对中国手语进行信息处理，需要针对中国手语的特性提出相应的信息处理方案。本文根据国内外的研究进展情况，从基于规则和基于语料库的角度，讨论了中国手语信息处理过程中遇到的有关问题，并提出可借鉴的中国手语信息处理技术，同时从中国手语自身的词法、句法出发，参考国外手语语言学的最新研究成果，讨论了中国手语信息处理中有关信息表征、理解、生成等问题。最后指出未来手语的信息处理将会更多地建立在跨学科、多模式的基础之上 ，该项研究将有力的促进信息无障碍技术的发展。

关键词：中国手语；信息处理；书写系统

中图分类号：TP391　　**文献标识码：**A

The Discussion of Information Processing of Chinese Sign Language

Dengfeng Yao[1,2,3] ,Minghu Jiang[1,2] ,Abudoukelimu.Abulizi[1,2], Hanjing Li[3], Halidanmu.Abudukelimu[4]，Dina Xia[5]

(1.Lab of Computational Linguistics, School of Humanities, Tsinghua University, Beijing, 100084, China;

2. Center for Psychology and Cognitive Science, Tsinghua University, Beijing, 100084, China;

3. Beijing Key Lab of Information Service Engineering, Beijing Union University, Beijing, 100101, China；

4. State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology, Department of Computer Sci. and Tech., Tsinghua University, Beijing, China；

5. China Electronics Standardization Institute,Beijing,100007, China)

Abstract: In order to use computer to process Chinese sign language, an information treatment plan is needed to present. This paper discusses the problems related to Chinese sign language information processing and proposes the processing technology according to the domestic and foreign research progress. Due to the lexical and syntactic characteristics of Chinese sign language and the latest research results in foreign Sign Linguistics, this paper puts forward the resolving scheme to information processing of Chinese sign language. First, to solve defects existing in the transfer scheme, the second is to focus on the phenomenon of sign language called

* **收稿日期：**　　**定稿日期：**

基金项目：国家自然科学基金资助项目 (61171114; 61433015; 91420202)；国家社会科学基金资助项目（14ZDB154; 13&ZD187）； 教育部人文社会科学研究规划基金资助项目（14YJC740104）；北京高校青年英才计划项目（YETP1753）

classifier predicates. Finally we point out that the future study of sign linguistics will be based more on the interdisciplinary and multi-mode and promote the development of the technology of information accessibility.

Key words: Chinese Sign Language；Information Processing；Writing System

1 引言

1996 年 12 月《吉隆坡宣言》指出手语不仅是聋人之间必不可少的交流工具，还是绝大多数聋人的第一语言。根据第六次全国人口普查及第二次全国残疾人抽样调查，推算出 2010 年末我国听力残疾人数为 2054 万人[1]，是我国人口最多的“少数民族”。中国手语（分为自然手语和文法手语，若无特殊说明，以下均指自然手语）是中国听力言语残疾人（聋人）交际和思维的主要工具。在聋人的知识习得、事物认知、信息获取、生存生活和参与社会等方面起着相依相伴的重要作用。国内外手语语言学研究表明手语有自己的语法规则、词汇结构，其视觉特性是任何有声语言中没有的语言现象 [2][3]。手语还存在相当于有声语言或文字符号的语音结构层，由手形、动作和表情等组成，并由这一结构层和其它非手势手语（No-Manual Sign 简写为 NMS）等多通道来表情传意。有声语言通常通过附加语素或添加词项来延长句子以表达更多的信息，但手语往往利用其多通道表达更丰富的信息。比如手势者修改手语行为、做出夸张的面部表情，或利用手势者周围的空间均可改变其手语含义。

中国手语研究起步相对较晚，但在国际语言学快速发展的背景下，近几年也如雨后春笋般迅速发展起来。中国手语已存在于世，尽管对其是否为独立的语言尚未完全达成共识，但这并不影响学术界对如此庞大族群语言——中国手语的研究和探索。如何将这门特殊的视觉语言进行有效的信息处理是个挑战，也是摆在我国科技工作者面前的一项重要任务，尤其是在国家大力推行少数民族语言保护政策和推行信息无障碍的社会背景下，这项工作显得更有意义。为此，中科院、清华大学和复旦大学等科研机构开展了中国手语的信息处理研究，并取得了一定的进展。其中中科院计算所在上世纪九十年代就开发了中国手语合成系统和识别系统，目前还与微软亚洲研究院、北京联合大学合作基于微软体感装置 Kinect 开发中国手语识别系统。清华大学则进行了中国手语的计算机信息处理研究，开展了语用标注、手语文本分词系统等研发工作。

中国手语信息处理是一项系统的工程。本文就中国手语信息处理的有关问题提出几点思考，并按照本文图 1 所示的手语信息处理几个基本步骤组织内容，其中手势识别属于计算机视觉内容，仅在第 5 节进行必要的介绍；第 2 节详细介绍中国手语的信息表示及手势切分问题；第 3 节介绍了为实现手语的信息处理而建设的手语语料库；第 4 节描述了为实现计算机处理手语而特别设计的手语语料标注；第 5 节在简述语料库建设的基础上，介绍了手语信息处理；第 6 节介绍了手语机器翻译碰到的问题以及初步解决方案；第 7 节介绍了手语生成器的情况；最后第 8 节是全文的总结与展望。

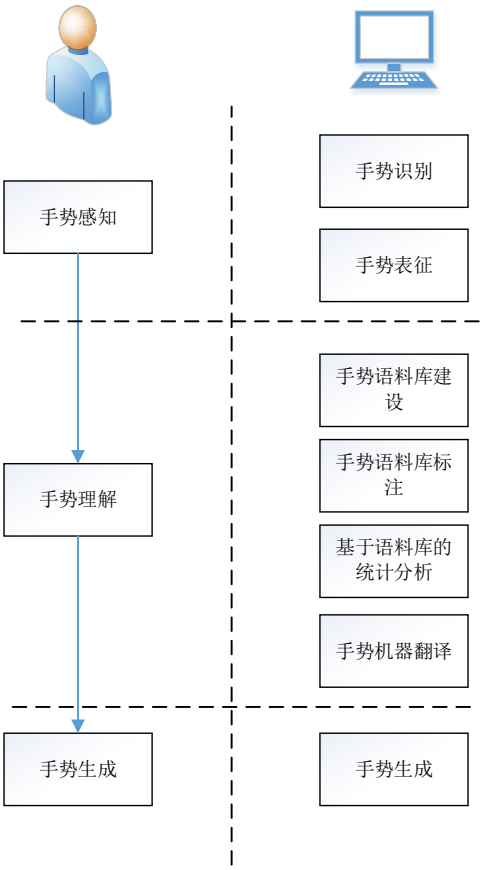

图 1. 手语信息处理流程图

2 中国手语的信息表示及手势切分问题

任何语言能够被计算机处理的前提是该语言拥有书写系统，并能够机读化。尽管国外学者认为手语存在书写系统，比如 SignWriting 系统[4]、ASL-phabet [5]、Stokoe 符号集 [6]、HamNoSys [7]。然而这些书写系统的受众太少。从 SignWriting 群发(sw-l@majordomo.valenciacc.edu)的报告上看，世界上大概有 14 所学校正在使用。Diane Brentari 解释说受众少是因为人口、政治和技术等因素[8]。其实问题的本质在于借鉴有声语言的单信道来处理手语的多信道存在很大的困难。有声语言拥有书写系统，它通过声音作为载体输出，这种音频是基于时间轴的数据流，语音信道是随着时间的推移而改变的一组值[9]。有声语言的自然语言处理系统是基于文本的，只需要记录语音对应的书面文字这个唯一的信道，因此只要求用户具有良好的识字能力。而手语本质是多信道：手的位置、形状、手的方向、眼睛凝视、头倾斜、肩部倾斜、身体姿势和面部表情，在手语中所有这些信道信息都代表着语言含义。中国手语的多信道性质，使得将手语编码成线性单信道字符串尤为困难。

从目前所报道的文献来看，中国手语尚未有一个能被聋人群体接受的标准书写系统，因此并没有任何文字保存。限于目前的技术条件，尚不能直接对手语视频进行语言处理，只能将中国手语转写成近似的汉语书面语言，再进行计算机信息处理。需要指出的是这里的转写不同于语音记录。文献[10]指出语音记录(notation)和转写(transcription)很相近，但亦有不同。语音记录倾向于指用图形符号、字母、文字等书写体系来记录言语中词语的发音，比如用国际音标 IPA 为有声语言标音，或者手语中用汉堡转写系统 HamNoSys(Hamburg Notation System for signed languages)记录手语词的打法。转写则是对更长的面对面的交流或口头表达

等的图形符号或文字记录等，它需要用到专门的标音体系（语音转写或音位转写），或文字体系（实录转写）。例 1、例 2、例 3 显示了汉语、中国手语语音记录、中国手语转写的例子。

例 1:

汉语：给你介绍一下，我在崇文区残联(工作)，他在北京大学教书。

例 2:

中国手语语音记录：介/我你，我/CW 区/残联，他/北京大学/教 。

例 3:

中国手语转写：(微笑)["介"(自身→对方)[话题],指(自身)/CW(崇文)—"区"/残—联，指(第三方)/北—京—大学/教]

从例 2 可以看出，语音记录的文本或句子为一维语言，更适合进行计算机处理。但中国手语直译后的汉语文本或句子"瘦"得只剩"骨架"，没有"血肉"，其内容要损失大概 50%[11]。甚至比不上例 3 的转写句子。由此可看出将中国手语编码成线性单信道字符串的复杂性和难度。这些书写系统在对实际手语动作抽象时会省略一些细节，并且在开发书写系统时哪些细节可以省略，哪些细节则不可忽略，这是一个极具挑战性且容易出错的课题。正因为手语本身特有的复杂性和空间性，使得手语信息处理具有挑战性，为其改进或发明书写系统或者转写方案，这是进行中国手语计算机信息处理不可跳过的第一步。

当然也可以使用国外开发的书写系统对中国手语转写。如国外较流行的有 HamNoSys[7]、Sign Writing[4]等转写系统。这些转写方案已被国外的若干手语语料库使用。文献[12]以"bears"这个词为例比较了 Stokoe 系统、HamNoSys、SignWriting 三个转写系统，比较结果发现 HamNoSys 的线性结构有利于计算机的读取与识别。文献[13]为了将手语手势嵌入手机来帮助残疾人，对 Stokoe、HamNoSys、Sign writing 三个符号转写系统进行了比较，结果发现 Sign writing 比 Stokoe 符号系统、HamNoSys 符号转写系统更易理解，更适合嵌入手机。文献[14]比较了 Sign writing 和 HamNoSys 符号转写系统，发现 Sign writing 的图形组织不够正规，HamNoSys 符号转写系统更易机器读取。文献[15]综述了用于机器翻译南非手语的手语符号表示法，比较了 Stokoe、HamNoSys、Sign writing 三个符号转写系统，Stokoe、HamNoSys 被认为在技术格式上不切实际，Sign writing 则更易阅读。根据以上比较研究可以得出，Stokoe、HamNoSys 不易理解，但易于计算机阅读；Sign writing 易理解，但不易计算机阅读，且在 Stokoe、HamNoSys 两者中，HamNoSys 被广泛用于手语机器翻译和手语三维模型生成。

由于种种原因，Stokoe、HamNoSys、Sign writing 等系统尚未传入我国。目前国内主要采用文献[16]中的中国手语转写方案，例 3 就显示了该方案的转写例子，可以看出该方案涉及词性、构词法、方向性、句型以及非手控等方面的信息，易于人们阅读理解,但该方案主要用于语言学研究，未考虑计算机信息处理所需的基本技术，比如切分、标注、句法、语法分析等，这些技术的缺失将限制生成流畅手语动作的能力。最大的缺憾是该方案虽有方向信息，但缺少手形、手掌方向、运动方向等手控方面的信息，且因中文的歧义性，为手语转写体系的机器阅读与理解造成很大困难。

若采用汉语作为中国手语的转写语言,与国外手语信息处理的特殊之处在于需要解决转写文本自动分词和消歧的问题。当然目前的汉语分词技术较成熟，使之应用于手语信息处理已经不是难题，它为下一步手语的信息处理创造了条件。需要注意的是中国手语转写文本分词除了可以借鉴国内外分词技术及算法研究的优势，还需要从自身的词法、句法等出发，提出与之相应的手语分词方案 ，特别是要处理好汉语最小语言单位"字"和手语最小语言单位"手势"的关系 。通常一个汉语复合词有可能由两个手势构成， 比如"妻子"，在汉语分词里是一个词的单位，但在手语里却是合成词，因为手语对"妻子"的表示是 "结婚"+"女人"，

或者 “女人”+“结婚”，这样本来在汉语里是一个语素的“妻子”，在手语里却是由两个语素构成的合成词。这种情况在中国手语里大量存在，经常是汉语里一个名词为一个语素，在手语里却变成了两个语素甚至三、四个语素。与以上情况相反，汉语的两个语素，在中国手语里是一个语素的也大量存在。最典型的是动宾一体，所谓动宾一体其实是比照汉语语法来说的。比如汉语中，踢足球是两个词素，分别为动词“踢”和宾语“足球”，但手语仅一个手势就可以表达“踢足球”。这是手语作为视觉语言的特性所决定的，相对汉语的语法特点，手语语法相对简化，而不能按汉语语法规则来分析手语。

3 手语语料库的建设

对手语进行识别、理解、生成等信息处理问题，国内外无外乎采用基于规则或基于语料库统计的方法来进行处理。文献[17]考察了汉语和中国手语之间的同异，建立了汉语中国手语机器翻译的一系列规则，在此基础之上采用规则解释方法实现了一个汉语至可视化语言中国手语的翻译系统。由于真实语言的输入集是无限的，这种通过有限规则，特别是少量规则的建模方法，很难满足手语语言处理的全部需要。以翻译手语的典型语言现象——分类词谓语为例，最简单的方法是在传统的英语词典里存储语义特征，例如：+人+车辆+动物+平面表面。针对每个单词或词组在英语词典里存储一组 3D 坐标，将英语词典中特定动词或介词与其他特征，如运动路径、固定位置、相对位置、形状、轮廓等相关联，可以帮助识别谓词要表达什么样的信息， 从而进一步缩小谓词可能的分类词手形集合，产生谓词的 3D 运动[18]。由于 3D 场景部署有许多可能性，这种方法在组合上是不切实际的，例如考虑汽车可以行走所有不同形状和坡度的道路。其他用得最多的是采用基于启发式规则的方法来计算运动路径，比如基于英文源文本的一些有限的特征集或者语义元素集合来设计运动路径。这种方法需要将基本的特征集组合以便产生一个单一的分类词谓语运动的动画组件库，将这些组件与相应的英语特征或语义元素相关联，这样就可以选择适当的动画组件并在转换时可以组合产生 3D 运动。这些基于规则的方法有个前提条件，需要手势者事先决定用哪些空间信息来交流，并决定如何表示其排序，这样才能描述如何建立一个独立的分类词谓语。并且只能生成单一的分类词谓语。对于生成多个相关分类词谓语还很困难，更重要的是这些基于规则的方法都缺乏规划整个场景元素的能力[19]。

因此目前在手语信息处理领域，基于语料库的统计方法成为主流。从以上可看到，尽管手语没有书写系统，这并不妨碍各国开展本国手语语料库的建设，国外已开始建设手语视频语料库 [20][21][22][23][24]。由于手语是没有书写系统的语言，国外普遍将手语视频作为手语语料来进行处理，再用本国语言给手语进行转写。如德国手语语料库用德语转写，美国手语语料库用英语转写。虽然本国语言不是手语的专门书写系统，可能会遗漏很多语言学细节，但聊胜于无，这些语料库从零开始，为手语信息处理创造了条件。从各国目前建设的语料库用途来看，主要用于语言学研究，比如研究手势变异、语义、形态、音韵、语法等，同时也有将语料库用于手语词典、手语教学、手语翻译以及特定领域手语应用等方面。具体详见表 1。

表 1 各国手语语料库研究情况

语料库分类	语料库用途	代表性例子
通用手语语料库	支持词汇变异、语义、形态、音韵、语法等语料库全面内容的研究	1. 德国手语语料库（德国汉堡大学德国手语与聋人交流会） 2. 爱尔兰手语语料库（英国经济社会研究所于 2008 年至 2011 年） 3. 西班牙通用手语语料库 4. 基于文件的澳大利亚手语语料库
专用手语语料库	制定手语标准	1. 英国伦敦大学建立的英国手语语料库（BSL），结合美国手语，对不同地域的手语进行分析比较，研究手语的地域差异
	用于手语教学	1. 瑞典手语语料库 2. Diane Lillo‐Martin 提出建立的双模式双语双国儿童语言语料库(bi‐bi‐bi corpus)) 3. 荷兰手语研究者 Onno Crasborn 等建立的维基手语语料库（sign language corpora wiki） 4. 阿拉伯手语语料库
	编纂字典	1. 美国手语视频语料库(American Sign Language Video Corpus） 2. 汉堡科学院建立的德国手语语料库
	语言学研究	1. 斯洛文尼亚手语语料库及语法项目 2. 澳大利亚手语档案工程
	特定领域的手语语料库研究	1. 基于航空信息系统（Air Travel Information System(ATIS)）的航空信息系统手语语料库， 2. 针对天气预报领域的德国手语语料库

相对于国外手语语料库的研究，国内的研究较少且单一，一般集中在专用手语语料库的研究，多用来支持手语的一般性的词法、句法、语义现象的描写、解释和特定领域下针对特定目的手语研究。如中国科学院计算所、北京联合大学与微软亚洲研究院合作的基于 kinect 的手语识别和翻译系统项目中涉及的 226 句常用语及 2400 个有关不同场所、场合的常用语；中国黄晓晓建立的基于情景的手语语料库[25]，它包含个人在家庭学校等场合的日常交流。

4 手语语料库标注

为了使手语语料库适用于手语信息处理，标注是必不可少的工作，为此一般语料库建设学者都制定了一致的转写和标注方案。其中手势的识别释义（ID-glosses, unique identifier of sign types）是语料库建设中最为基础和重要的标注内容。它是用一个含对应意义的标注工作语言的词（比如汉语、英语），去表达手势。识别释义包括该词的词典形式和所有形态和音位变体。统一了标注者的识别释义，方可避免同一手语词被不同标注者贴上不同的标签，进而促进机器和用户准确而无遗漏地搜索到此手势的所有例（token）[10]。

为了使标注后的文本适用于信息处理，文献[2]提出了装饰字符串（decorated string）的标注概念，如图 2 所示，这是使用“装饰字符串”标注书写一个句子。在这个句子中手势者用双手打出三个手势：JOHN，NOT，ARRIVE。该图中，否定-摇头的横线条表示以否定的方式来摇头。图中“眼睛凝视”横线条下面，手势者需要凝视他或她身体旁边的位置，用这个位置代表 John。一般在美国手语中，手势者可以使用眼睛凝视来表示曲折动词的呼应对象。

图中的注释（指单词）是语言学家用来记录手势者手部活动的。其中黑色横线条并不是代表手语的信息，而是非手动特征（NMS），黑色横线条看上去像“装饰”字符串。这个标注系统用“空符号”（Ø）作为一个语言学单位的占位符，它表示手势者的手部不做任何动作。在该例中，Ø 表示手势 ARRIVE 开始之前眼睛凝视了一会儿。

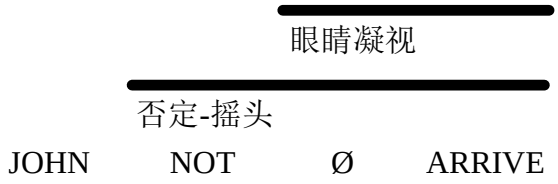

图 2. “装饰字符串”标注

这种标注的好处就是方便计算机进行信息处理，由此生成的语法树如图 3 所示。

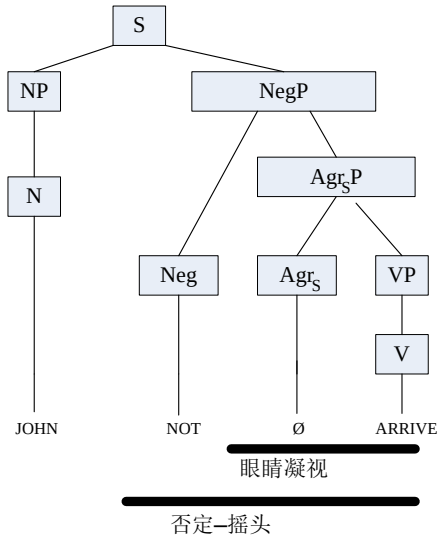

图 3. 语法树

这个树解释了手动手势的语法结构，但它没有说明 NMS 横线条如何跟它相关联。由于树是用来表示文本字符串嵌套结构的图形化方式，因此可以考虑树如何表示为图 4 的括弧结构（一维）。

[[[JOHN]] [[NOT] [[Φ] [[[ARRIVE]]]]]]

图 4. 括弧结构

这样我们可以看到 NMS 横线条就超出括弧结构的表示范围。因此括弧结构无法清楚地表示 NMS 横线条。虽然如此，但装饰字符串的提出在手语语言学上是一大进步。

文献[26]总结了以往树结构的理论工作，比如树结构可分成多维度[27]，以及能够表示视觉语言的语法[28][29]。采用 NaÏVE3D 树作为基础，他们提出了 P/C 思想，将装饰字符串概念做了扩展，如图 5 所示，亦可看做三个信道。

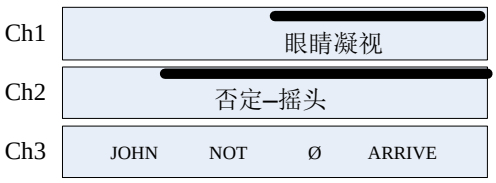

图 5. 装饰字符串扩展

进行计算机处理时，仍使用 3D 树进行语法处理，如图 6 所示。

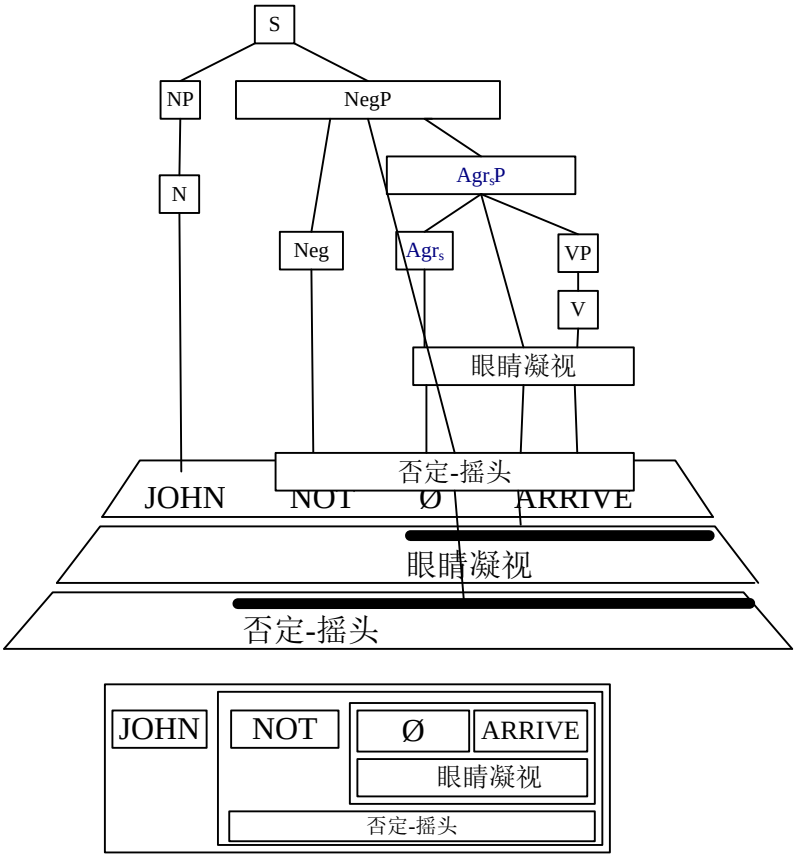

图 6. 3D 语法树

从上方看，3D 树看起来如图 5 的二维括弧结构。水平轴代表时间，垂直轴代表多个信道。整个句子包含在一个单一矩形里，该矩形对应于树中的 S 节点。从左到右它跨越了整个句子，从上到下它指定了所有信道的句子输出。对于 John 框右侧，是一个包含了其余句子的大矩形；它是 NegP 节点。当覆盖多个信道的节点被分为几个子节点时，每个子节点可以覆盖父节点所覆盖信道的子集。例如 NegP 节点将其 Agr_sP 子节点分配给两个顶信道，其否定摇头子节点分配给底信道。这种 P/C 思想旨在解决手语多信道的并列与非并列关系，具体表现为，每个矩形每次只在一个方向上分割，此外以非重叠的方式覆盖所有时间（从左到右）内父节点所有信道（从上到下）的矩形子节点。由于括弧图中的矩形是类似于 3D 树结构中的节点，将“矩形”术语替换成“节点”。当一个节点分支从左到右，称之为组合节点，我们说它已经被分成组件。对于组合成父节点的子现象，组件从左到右的顺序应被解释为指定了时间序列。子节点以非重叠的方式覆盖了其父节点的整个时间范围。组合节点就像传统语法树的节点（就像图 5），其中节点分成连续的子节点。

但是这种思想的缺点就在于需要事先假定在 P / C 树内部结构中一个节点可以分割或组合（但不是同时两者兼具）。此外分割节点的子节点被假定为他们之间没有时间并列关系。这就限制了其扩展范围。相比有声语言的类似字符串编码，P/C 思想能够更好的为手语语言学信号进行编码。这无疑可作为对中国手语信息处理的借鉴。

5 基于语料库的手语信息处理

从 1982 年 Shantz 和 Poizner 合成的美国手语计算机程序开始，各国在手语信息处理领域上取得了突出的成果，主要用于手语识别和手语生成，文献[30]给出较全面综述，使用方法有神经网络、HMM、向量机、机器学习等。需要指出的是，目前的手语识别研究已经从静态手势识别过渡到动态手势识别，从使用可穿戴设备提取特征过渡到基于计算机视觉提取特征。采用自然的不佩带任何装置或物品的手语输入方式,获得准确快速的识别结果,是目前该领域的研究核心与发展方向。如 Vogler 和 Metaxas 利用手语的基本单元而不是手语词汇进行连续手语识别,对 22 个词构成的句子实验结果表明,这种方法的识别率和传统方法的识别率相近[31][32][33]。文献[34]在不关联空间域和时间域特征的情况下,完全依赖于密集的局部特征,采用特征包和多类支持向量机识别手势。文献[35]采用 PCA 方法提取手势图像前 M 个特征值的特征向量，用最小欧式距离实现手势分类,由于 PCA 对尺度、旋转、光照变化等不具备鲁棒性,所以该方法需收集各种情况下的手势训练样本。

国内的手语识别和生成研究较早，文献[36]表明我国已建立了一个能够识别大词汇量的中国手语识别系统，该系统对 1064 个中国手语孤立词的识别率达到 90%。通过嵌入式训练，对由 220 个词构成的 80 个句子的手语识别率达到 95.2%，同时一个中国手语自动翻译系统也由该研究者设计建成，对 5177 个中国手语孤立词进行离线识别，识别率为 94.8%。文献[37]通过鲁棒回归分析和变阶参数模型对小规模的动态手势进行识别，将手势图像运动参数应用于手语表观建模，并提出了一种手势运动估计方法，然后将这两种特征作为表观特征创建手势模板，通过最大最小优化算法进行基于模板的手势分类识别,该方法在手势图像运动信息的基础上对 12 种手势进行识别，准确率超过了 90%。由此可看出我国在基于数据手套的手语识别研究方面已处于世界领先地位，但在基于视觉的手语信息处理领域，尤其是动态手语识别方面与其他发达国家还有一定差距。

然而以上模型都没有提出手语理解算法来解决手语语言处理问题。造成这种情况的原因除了识别率不太理想，还在于视频语料采集繁琐，人工标注困难，以致用于手语信息处理的手语语料库普遍未达到一定的规模。图 7 显示了各国手语语料库的规模对比，由此可以看出规模都在 50 小时左右，而流畅手语是每秒 2-3 个手势，因此生语料库的规模大概在 36-54 万个手势。国外学者指出一般手语视频语料的 RTF 因子为 100，也就是 1 个小时的语料至少需要 100 个小时做标注[38]。照此推算，按一天标注 8 小时算，50 小时左右的视频语料需要 21 个月左右才能完成标注，如此庞大的标注工作量使得手语熟语料获取困难。此外还有一个原因在于手语语料库没有根据手语特点建立相应的模型，比如一些语言学家做了一个手势的多种打法动作捕捉数据语料库[39]，他们记录了手势输出的手形、手的位置、方向、运动和非手动元素的时移参数。但这些模型并没有说明许多手语语言学现象如何表示。比如表示分类词谓语出现的手形数量并不多，但这些模型却记录手形的信息特别多，而对表现分类词谓语特征的手部方向的信息记录太少，以至使指定复杂运动路径更为困难，而这些复杂路径是表示分类词谓语所必需的。

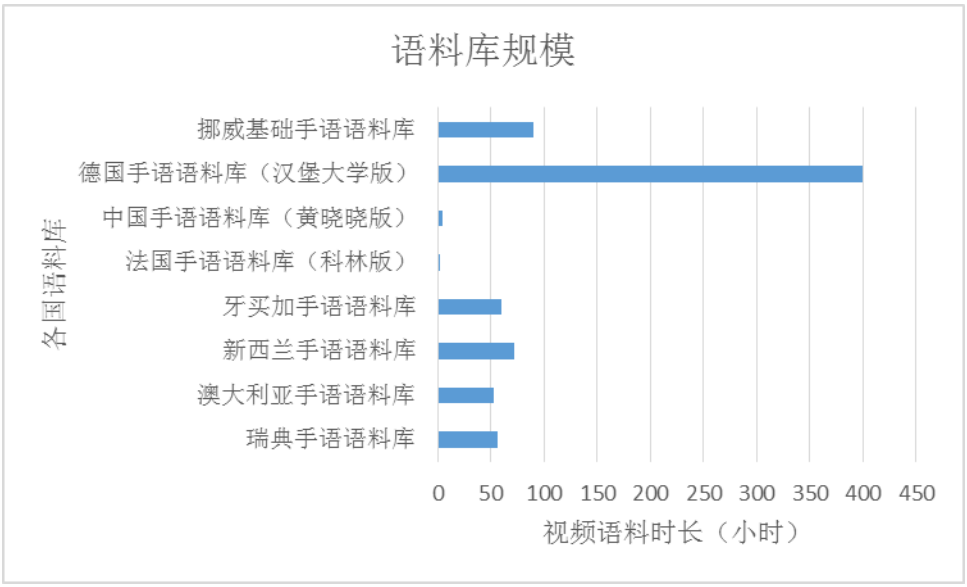

图 7.　各国手语语料库的规模对比

有声语言能够成功使用统计模型，是因为网络时代信息的数字化和网络化为统计模型带来了取之不尽、用之不竭的数据资源。手语语料因为视频采集繁琐和标注困难，缺乏相应的应用规范和模型，使得手语的生语料和熟语料数据依然匮乏，手语应用统计模型仍然面临严重的数据稀疏问题，此外单纯的概率模型也不能全部解决手语语言处理的自动化问题。因此目前力图用传统的统计模型和机器学习方法来研究手语机器翻译还很困难，至少在没有可靠的方法来为语料库建立一个手势者的 3D 模型，或大规模视频语料自动标注技术尚未出现之前是不切实际的。

6 手语的机器翻译

手语的机器翻译并不是简单的将汉语一个词对应一个手势翻译出来。与主流有声语言不同，手语具有视觉空间的立体性特征。这种特殊性对于传统的计算语言学方法是一个极大的挑战。文献[40]认为进行手语机器翻译时需要模拟真人手语译员事先在心里形成 3D 空间影像，然后将空间的对象位置映射到物理手势空间，以表达手语的概念。具体来讲，像“猫在床下跑”的例子，首先选择一个基于“猫”实体特征（小型动物、四条腿对象、跑动过程中等特征）的手形闭集（即分类词手形，此手形与其手部在手势者前面空间进行的运动一起组成分类词谓语），以及手势者希望讨论的实体空间特征，其表面（猫在平坦的地面上）、床下空间的大小、形状、位置（猫在床下任意位置）、运动（跑动、非静止）等。然后手势者针对需要表达床下空间的轮廓、手势者周围空间的位置（在伸展两只手的范围内选择哪个空间位置来代表施动者猫和被动者床）、3D 空间的运动(在床下有限空间内表示跑动)、物理/抽象的维度（床需要表示多大、跑动幅度需要多大？）或某些其他需要被传递的对象属性，比如床是不是席梦思、猫是不是每天都在床下跑、猫是否还有同伴陪它跑等因素，而相应地制定手部立体运动。此外还要根据汉语“猫在床下跑”的上下文环境提取语用特征，比如想表达猫捉老鼠很勤快，还要配合眉毛和眼睛的动作（眉开眼笑）、脸部表情（表达夸张的情绪），必要时还要头部动作(头部稍微向前倾)、身体姿态（抬起肩膀、身体上部左右摇晃等）及其他方式来表达“猫在床下跑”的含义。

从以上过程可以看出手语机器翻译**困难在于手语表达的每一块空间信息都必须被编码为一个语素，通常需要许多语素传达各种各样的空间信息来表示手语，特别是在用于组合**

空间信息来描述场景中对象之间的空间关系或比较的情况下。文献[41]做了一个统计分析，比如分析手语句子“一个人走向另一个人”的语素时，总共有 28 个语素，包括：两个面对面的实体、都在同一水平面上、都在垂直方向、自由运动、都有一个特定的距离、在直线路径运动等等。作者认为生成各种手语，此多语素模型需要一个巨大的、甚至可能是无限数量的词素集合。

此外手语机器翻译的另一个困难在于复杂的空间相互作用和3D场景限制很难编码成一组组成规则[42]。例如“汽车行驶在颠簸的道路上，经过一只猫。”该句有两个分类词谓语。为了生成这些谓词，手势者必须知道如何部署场景，包括猫、道路、汽车的位置。要为汽车选择运动路径，有开始/结束位置，手部必须流畅地表示路径轮廓，例如颠簸、丘陵起伏、曲折的。靠近猫的道路、地面平面、曲线道路也必须表达出来。此外生活常识包括一些世界知识也必须要了解：（1）猫一般坐在平面地上；（2）车一般沿着地面道路上行驶。可以看出要想成功完成这两个分类词谓语机器翻译涉及大量的语义理解、空间知识和推理。由以上分析可以看出分类词谓语机器翻译的复杂性。曾有学者评论手语的**分类词谓语是超语言的空间手语、非空间多语素结构或构成空间参数化表达式**[41]。

而传统计算语言学方法的缺陷就在于不能模拟手语三维场景中的对象空间布局，为了解决这些问题，国外研发的手语机器翻译系统在此方面做了有益的尝试。文献[43]提出（并建立了原型）的英语到美国手语（ASL）的翻译系统 ZARDOZ 系统，使用一组手工编码架构作为一种中间语言翻译组件。作者选择将分类词词根表示成高度未指定的词汇条目，词条的动作将取决于生成语法，因此，他们可以像对待任何其他单一的词条一样对待分类词词根。他还对常见的分类谓词的手形和运动类型进行了分类。因此他研发的 ZARDOZ 系统初步解决了分类词谓语的问题。在该系统中，分类词谓语表达的特定主题可以由独特的中间语言框架来表示，该中间语言框架由翻译体系结构的分析/理解部件进行选择和填充。他还讨论了空间和常识推理方法如何用来填充生成流畅的分类词谓语手形和运动所需要的动画具体细节。

不过限于目前的 AI 推理的发展水平和空间表示技术，开发这样的系统显然并不现实，因为它需要相关领域知识，并且是个很耗时的工作。不过它对于中国手语机器翻译可以提供借鉴，因为手语中分类词谓语不是独立词汇，无需遵循手语词汇音系学中对称和统领的条件[44]，它由拥有各自语法功能和意义的词素共同组成，如手形、移动、手势者的身体表情等。由于手语分类词谓语有极其复杂的内部结构，而且没有汉语词汇与之一一对应。若要研发中国手语机器翻译系统，这种分类词谓语现象则是必须研究解决的问题之一。事实上，虽然 ZARDOZ 的方法没有实用化，但很多系统都参考了这些方法[45]。这个例子提示我们，在对手语进行机器翻译时，要对手语特有语法现象有足够的认识和了解，起码要熟悉手语音韵学、语义学、句法学和形态学等独有的规律和特征。

7 手语生成器

与其他传统机器翻译系统不同，手语机器翻译还需要一个手语生成系统，以负责生成手语动画。大量的研究表明，聋人虽大多学过本国书面语言，但因听力障碍，他们先天口语习得存在着困难，由此导致大多数聋人高中毕业生的阅读水平相比健听人滞后三到四年[46]。在此情况下，一个好的中国手语机器翻译系统最好附带手语生成系统，这样才能真正达到信息、服务并具的无障碍。目前手语动画生成技术才开始出现在适用于聋人用户的软件和网站上。

虚拟人体建模和动画研究已比较成熟，现有技术已足够开发能够清楚表达和快速响应手语动画的人物模型[47]。当然仅有动画人物模型是不够的，还需要一个中国手语生成器。即给定一个汉语文本或抽象的语义输入，计算语言学部件需要告诉动画人物该怎么做（假定语

言学和动画组件之间的接口已设定了正确的指令集）。这样就需要一个动画脚本，专门负责告诉动画人物如何做。因为中国手语是一个没有标准书写系统的语言，制定动画脚本规范也就没有统一的格式。以美国手语(ASL)为例，有很多机器翻译系统专门为 ASL 动画脚本规定了格式，每个系统开发的脚本语言也不一样。TEAM 系统使用了嵌入式参数的注释表示 ASL 句子[24]，这样就影响了非手势手语和音韵平滑的质量。TEAM 系统使用一个非常小的示范词典，并且手势的动画动作作为参数化运动路径的模板，该模板与 Jack Toolkit 和 Jack Visualizer 兼容[48]。而 ViSiCAST 系统中的词典存储了关于手语的语音信息，不仅包括语音的 SGML 规范，还包括特殊的次范畴、句法和形态特征等。这种存储用到了手语手势标记语言[49]，此标记系统本来是用于手语书写/转录系统，它着重于如何指定手形、手掌方向和运动细节，以表达手语语义。该系统处理的 ASL 动画脚本使用了手势标记语言以及 HamNoSys 手语书写系统的 XML 版本。比较先前两个系统使用的运动控制语言，SGML 适用于较重要的运动类型。因此，定义 ASL 的过程使用标记应该更直观。若要演示动画，ViSiCAST 系统设计师使用能接受 SGML 输入的动画角色，并产生动画。

虽然脚本技术不一，但机器翻译系统需要动态修改脚本，以便输出手语动画。这些系统必须针对手语句子生成一个手势和面部表情序列，然后他们要将得到的序列合成一个实际动画。目前国外主要用数据驱动的方式来生成手语动画，并且在已有的手语视频语料库基础上，使用统计模型和机器学习方法来研究手语动画的生成。

8 总结

以上研究表明，很多学者在手语机器翻译和计算手语学方面做了大量工作。比如新的手语表示模型、扩展的手语注释文法、 Movement-Hold 语音模型、话语表示如何管理手语对话空间定位的实体、分类词谓语的空间复杂性和三维表示，这些研究工作为手语信息处理创造了条件。

目前我国手语信息处理仍处于起步阶段，近几年才开始利用语料库资源进行手语机器翻译的研究。中国手语的计算机处理虽然起步时间不长，但可站在较高的起点，可借鉴国内外的研究成果，少走弯路，同时结合中国手语自身的词法、句法等特点，走出自己的新路。例如可以尝试应用大脑的认知理论、手语语言理解、脑成像等技术来研究手语的信息处理， 特别是最近出现的深度学习理论，有望解决手势的表征模型问题。

总之中国手语信息处理研究有着广阔的前景，虽然存在很多具有挑战性的难点和问题，我们可期待汉语语言学、手语语言学、神经科学、计算机科学和心理认知学等学科的学者进入该领域探讨跨学科的研究，以便取得更大的进展，以期为无障碍交流环境提供软件和硬件的支持，也为扩展计算语言学起到抛砖引玉的作用。我们相信未来的手语语言学研究发展趋势必将形成文、理、工、医交叉、多学科整合模式，获得不同视角、多领域的跨学科研究成果。

参考文献

[1]. 中国残疾人联合会. 2010 年末全国残疾人总数及各类、不同残疾等级人数 [OL]. 2011. http://www.cdpf.org.cn/sytj/content/2012-06/26/content_30399867.htm

[2]. Kegl J, MacLaughlin D, Bahan B, et al. The syntax of American Sign Language: Functional categories and hierarchical structure[M]. Cambridge, MA: MIT Press, 2000.

[3]. Vally C, Lucas C. Linguistics of American sign language[J]. 3rd edition, Washington, DC: Gallaudet University Press, 2002.

[4]. Sign Writing. Sign Writing[OL]. 2011. http://www.signwriting.org/

[5]. Supalla S, Cripps J H, McKee C. Revealing Sound in the Signed Medium Through an Alphabetic System[C]//Poster presented at the First SignTyp Conference, Storrs, CT. 2008.
[6]. Stokoe， William C.， Dorothy C. Casterline & Carl G. Croneberg. 1965. A dictionary of American Sign Language on linguistic principles. Silver Spring， MD: Linstok.
[7]. Prillwitz S, Hamburg Zentrum für Deutsche Gebärdensprache und Kommunikation Gehörloser. HamNoSys: version 2.0; Hamburg Notation System for Sign Languages; an introductory guide[M]. Signum-Verlag, 1989.
[8]. Johnston T A. W (h) ither the deaf community? Population, genetics, and the future of Australian sign language[J]. American annals of the deaf, 2004, 148(5): 358-375.
[9]. Huenerfauth M. American sign language generation: multimodal NLG with multiple linguistic channels[C]//Proceedings of the ACL Student Research Workshop. Association for Computational Linguistics, 2005: 37-42.
[10]. Johnston T. From archive to corpus: transcription and annotation in the creation of signed language corpora[J]. International Journal of Corpus Linguistics, 2010, 15(1): 106-131.
[11]. 吴铃. 手语语法和汉语语法的比较研究——寻找聋人失落的书面语[J]. 中国特殊教育, 2006, 8: 010.
[12]. Morteza Zahedi, Hermann Ney, Gerhard Rigoll, Robust Appearance-based Sign Language Recognition[D]. Rheinisch-Westfälischen Technischen Hochschule Aachen 21.09.2007
[13]. Adhinarayanan VenkataSubramaniam1, Karthikeswaran Duraisamy, Dinakar Subramaniam and Marikkani Chelladurai Embedding Sign Representation in Mobile Phones to Assist Disabled[J]. Computer Technology and Application 2 (2011) 42-47
[14]. Michael Filhol, Annelies Braffort,Sign description : how geometry and graphing serve linguistic issues[J] LIMSICNRS, Orsay.G762
[15]. Hutchinson J. Literature Review: Analysis of Sign Language Notations for Parsing in Machine Translation of SASL[J]. 2012.
[16]. 倪兰. 中国手语动词方向性研究[D]. 复旦大学, 2007.
[17]. 徐琳, 高文. 面向机器翻译的中国手语的理解与合成[J]. 计算机学报, 2000, (1):60-65.
[18]. Supalla T R. Structure and Acquisition of Verbs of Motion and Location in American Sign Language[D]. Ph.D. Dissertation, University of California, San Diego, 1982.
[19]. Liddell S K. Grammar, gesture, and meaning in American Sign Language[M]. Cambridge University Press, 2003.
[20]. Bauer B, Hienz H. Relevant features for video-based continuous sign language recognition[C]//Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on. IEEE, 2000: 440-445.
[21]. Huenerfauth M. American sign language generation: multimodal NLG with multiple linguistic channels[C]//Proceedings of the ACL Student Research Workshop. Association for Computational Linguistics, 2005: 37-42.
[22]. Marshall I, Safar E. Grammar development for sign language avatar-based synthesis[C]//Proceedings HCII. 2005: 1-10.
[23]. Stein D, Bungeroth J, Ney H. Morpho-syntax based statistical methods for sign language translation[C]//11th Annual conference of the European Association for Machine Translation, Oslo, Norway. 2006: 223-231.
[24]. Zhao L, Kipper K, Schuler W, et al. A machine translation system from English to American Sign Language[M]//Envisioning machine translation in the information future. Springer Berlin Heidelberg, 2000: 54-67.
[25]. 黄晓晓. 基于情景语料库的自然手语构词研究[D]. 南京师范大学, 2012.
[26]. Huenerfauth M. Representing coordination and non-coordination in American Sign Language animations[J]. Behaviour & Information Technology, 2006, 25(4): 285-295.

[27]. Bird S, Liberman M. A formal framework for linguistic annotation[J]. Speech communication, 2001, 33(1): 23-60.

[28]. Martell C H. An extensible, kinematically-based gesture annotation scheme[C]. 3rd International Conference on Language Resources and Evaluation. 2005.

[29]. Tucci M, Vitiello G, Costagliola G. Parsing nonlinear languages[J]. Software Engineering, IEEE Transactions on, 1994, 20(9): 720-739.

[30]. Ong S C W, Ranganath S. Automatic sign language analysis: A survey and the future beyond lexical meaning[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2005, 27(6): 873-891.

[31]. Vogler C, Metaxas D. Adapting hidden Markov models for ASL recognition by using three-dimensional computer vision methods[C]//Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation., 1997 IEEE International Conference on. IEEE, 1997, 1: 156-161.

[32]. Vogler C, Metaxas D. ASL recognition based on a coupling between HMMs and 3D motion analysis[C]//Computer Vision, 1998. Sixth International Conference on. IEEE, 1998: 363-369.

[33]. Vogler C, Metaxas D. Toward scalability in ASL recognition: Breaking down signs into phonemes[M]//Gesture-based communication in human-computer interaction. Springer Berlin Heidelberg, 1999: 211-224.

[34]. Niebles J C, Wang H, Fei-Fei L. Unsupervised learning of human action categories using spatial-temporal words[J]. International journal of computer vision, 2008, 79(3): 299-318.

[35]. Dardas N H, Petriu E M. Hand gesture detection and recognition using principal component analysis[C]//Computational Intelligence for Measurement Systems and Applications (CIMSA), 2011 IEEE International Conference on. IEEE, 2011: 1-6.

[36]. Gao W, Ma J, Wu J, et al. Sign language recognition based on HMM/ANN/DP[J]. International journal of pattern recognition and artificial intelligence, 2000, 14(05): 587-602.

[37]. Zhu Y, Xu G, Huang Y. Apearance-based dynamic hand gesture recognition from image sequences with complex background [J]. Journal of Software, 2001, 11(1): 54-61.

[38]. Dreuw P, Neidle C, Athitsos V, et al. Benchmark Databases for Video-Based Automatic Sign Language Recognition[C]//LREC. 2008.

[39]. Arena V, Finlay A, Woll B. Seeing sign: The relationship of visual feedback to sign language sentence structure[C]//Poster presented at CUNY Conference on Human Sentence Processing, La Jolla, CA. 2007.

[40]. Huenerfauth M. Spatial representation of classifier predicates for machine translation into american sign language[C]//Workshop on Representation and Processing of Sign Language, 4th Internationnal Conference on Language Ressources and Evaluation (LREC 2004). 2004: 24-31.

[41]. Liddell S K. Sources of meaning in ASL classifier predicates[J]. Perspectives on classifier constructions in sign languages, 2003, 199: 220.

[42]. Bangham J A, Cox S J, Elliott R, et al. Virtual signing: Capture, animation, storage and transmission-an overview of the visicast project[J]. IEEE Seminar on “Speech and language processing for disabled and elderly people.” . 2000.

[43]. Veale T, Conway A, Collins B. The challenges of cross-modal translation: English-to-Sign-Language translation in the Zardoz system[J]. Machine Translation, 1998, 13(1): 81-106.

[44]. Battison R. Lexical Borrowing in American Sign Language[J]，Linstok Press, Silver Spring, MD . 1978.

[45]. Huenerfauth M. Generating American Sign Language classifier predicates for English-to-ASL machine translation[D]. University of Pennsylvania, 2006.

[46]. Holt J A. Stanford Achievement Test—8th edition: reading comprehension subgroup results[J]. American Annals of the Deaf, 1993, 138(2): 172-175.

[47]. Wideman C J, Sims E M. Signing avatars[C]//Technology And Persons With Disabilities Conference. 1998.

[48]. N. Badler, R. Bindiganavale, J. Bourne, M. Palmer, J. Shi, and W. Schuler. 1998. A parameterized action representation for virtual human agents. In Workshop on Embodied Conversational Characters, Lake Tahoe, CA. http://www.cis.upenn.edu/~rama/publications.html

[49]. Kennaway R. Synthetic animation of deaf signing gestures[M]//Gesture and Sign Language in Human-Computer Interaction. Springer Berlin Heidelberg, 2002: 146-157.

作者简介：姚登峰（1979—），通讯作者，男，博士研究生，讲师，主要研究领域为手语认知与计算。Email:yaodengfeng@gmail.com；江铭虎（1962—），男，教授，主要研究领域为语言认知与计算。 Email: jiang.mh@tsinghua.edu.cn；阿布都克力木.阿布力孜（1983—），男，博士研究生，主要研究领域为语言认知、认知神经科学。Email: keram1106@163.com。李晗静（1974—），女，教授，主要研究领域为自然语言处理。Email: tjthanjing@buu.edu.cn；哈里旦木.阿布都克里木（1978—），女，博士研究生，主要研究领域为自然语言处理、人工智能。Email: abdklmhldm@gmail.com。

注：1、第 1~3 位作者请随论文提供一张一寸登记照片。

2、文中图表请统一采用黑白图，文中对图的说明应与图表一一对应，表达清晰。

3、每篇论文可以注明一个通讯作者，如果需要标识，请在提交个人简介的时候标注清楚。