手语计算 30 年:回顾与展望

姚登峰[1),2)] 江铭虎[2)] 鲍 泓[1)] 李晗静[1)] 阿布都克力木·阿布力孜[2)]

[1)](北京市信息服务工程重点实验室(北京联合大学) 北京 100101)

[2)](清华大学人文学院计算语言学实验室、心理学与认知科学研究中心 北京 100084)

摘 要 手语的自然语言处理是计算机学科中的一项重要任务.目前随着信息技术的飞速发展,以文本和语音为主要载体的传统语言计算的工作重点已从编码、输入方法和字音的研究逐渐转移到语法层面,并进入深度计算的阶段.然而手语信息处理却严重滞后,处于空白起步阶段.究其原因,主要是缺乏用于机器学习的具有一定规模的手语语料库资源,同时传统的语言计算技术也存在不足,这些都阻碍了手语机器翻译、手语问答系统、手语信息检索等信息处理的应用研究.该文首先阐述了手语计算与传统语言计算的本质差异在于空间建模,这种差异导致了前者核心任务是单信道与多信道转换,后者根本任务是消歧.从词法、句法、语义、语用、应用等层面对手语计算进行了回顾,重点介绍了手语机器翻译和分类词谓语计算,指出分类词谓语是手语计算的关键以及取得突破的切入点.从展望的角度,认为互联网时代体感设备的出现、认知神经科学的兴起、深度学习的进展等新技术为手语计算带来了新的机遇.将手语计算与传统语言计算进行比较,分析了手语计算的趋势和未来的研究方向,手语的认知计算是从手势的物理特征到语义表征的映射转换过程,其计算趋势是填补音韵特征、语义单元这样的中间步骤,避免直接从底层特征得到语义概念,关注在手语行为与语言特征的关系上进行机器学习,建立融合空间特征的统计学习模型.未来研究方向包括资源建设、文景转换、隐喻理解,其中文景转换有助于实现空间信息抽取,即物体的空间方向、位置等信息,结合知识库消除自然语言的模糊性,进而实现三维场景构建.指出手语计算正从萌芽期过渡到发展期,若取得重大突破,手语计算将扩展语言计算体系,推动人工智能的发展.

关键词 手语计算;分类词谓语;机器翻译;空间建模;多信道;空间隐喻

中图法分类号 TP18 **DOI 号** 10.11897/SP.J.1016.2019.00111

Thirty Years Beyond Sign Language Computing: Retrospect and Prospect

YAO Deng-Feng[1),2)] JIANG Ming-Hu[2)] BAO Hong[1)] LI Han-Jing[1)] ABUDOUKELIMU Abulizi[2)]

[1)](*Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing* 100101)

[2)](*Laboratory of Computational Linguistics, School of Humanities, Center for Psychology and Cognitive Science, Tsinghua University, Beijing* 100084)

Abstract The natural language processing of sign language is an important task in the field of artificial intelligence and information processing. Currently, with the development of information technology, the focus on the information processing of spoken language and written language, is gradually shifting from the word coding and input method to the grammatical level, and then to depth computing. However, sign language information processing is seriously lagging behind and remains at the starting stage. The main reason for this situation is that no ready-made sign

收稿日期:2016-08-16;在线出版日期:2017-12-01.本课题得到国家自然科学基金重点项目(61433015)、国家社会科学基金重大项目(14ZDB154)、教育部人文社会科学研究青年基金(14YJC740104)、国家语委重点项目(ZDI135-31)、北京市属高校高水平教师队伍建设创新团队建设提升计划(IDHT20170511)、北京市教委科技计划项目(KM201711417006)、清华大学自主科研项目两岸清华大学专项(20161080056)及北京联合大学人才强校优选计划资助.**姚登峰**,男,1979 年生,博士,副教授,中国计算机学会(CCF)会员,主要研究方向为语言认知与计算、信息无障碍.E-mail:yaodengfeng@gmail.com.**江铭虎**(通信作者),男,1962 年生,博士,教授,博士生导师,主要研究领域为自然语言处理、认知神经科学.E-mail:jiang.mh@tsinghua.edu.cn.**鲍 泓**,男,1958 年生,博士,教授,博士生导师,主要研究领域为图像处理、机器学习.**李晗静**,女,1974 年生,博士,教授,主要研究领域为自然语言处理.**阿布都克力木·阿布力孜**,男,1983 年生,博士,讲师,主要研究方向为语言认知与计算.

language corpus resources can be used for machine learning and deep learning. Sign language machine translation, sign language question-answering system, information retrieval and information processing cannot be applied because of the lack of research foundation. The essential difference between sign language computing and traditional language computing is spatial modeling and it leads to that the core task of sign language computing is to convert single-channel representation to multi-channel representation, while the fundamental task of the traditional language computing is the disambiguation of single-channel representation. From the lexical, syntactic, semantic, pragmatic, and applied levels, sign language computing is reviewed, and the sign language machine translation and classifier predicates in computing are emphatically introduced. Classifier predicates are the key of sign language computing, and it is the breakthrough point of sign language computing. New technologies, such as the emergence of somatosensory devices, the rise of cognitive neuroscience and the progress of deep learning, have brought new opportunities to sign language computing in the Internet age. From the perspective of outlook, sign language computing is compared with spoken language computing. The trend of sign language computing and the breakthrough points are analyzed. The cognitive computing of sign language has been regarded as a mapping conversion process from the physical characteristics of gestures to semantic representation. The trend of sign language computing is to fill these intermediate steps, such as phonological features and semantic units. It avoids the semantic concepts obtained directly from the underlying physical features, focuses on the machine learning on the relationship between sign language behavior and language features, and establishes the statistical learning model of fusion spatial features. These breakthrough points include resource construction, text-to-scene and metaphor understanding. Among them, the text-to-scene in the sign language is helpful to realize the spatial information extraction including spatial orientation, object position, and the ambiguity of natural language can be eliminated by combining with the knowledge base, so as the three-dimensional scene construction can be achieved and creates a breakthrough in understanding the spatial relationship and generating the virtual scene. It is pointed out that the sign language computing is from the embryonic period to the development period. Driven by the interdisciplinary, sign language computing may make a substantial breakthrough. The astonishing progress of traditional language computing has promoted artificial intelligence and human-computer interaction to develop further. If a series of problems about sign language computing can be solved in fields of theory, technology and engineering, it will greatly speed up the development of artificial intelligence and natural language processing.

Keywords sign language computing; classifier predicates; machine translation; spatial modeling; multi-channel; spatial metaphor

1 引 言

中国是一个人口大国,听力障碍人数也居世界之首,据统计 2010 年末我国听障人群比人口最多的少数民族——壮族还多 440 万人,占全国总人口的 16.79‰[①].因此中国拥有世界上最丰富的手语(Sign Language)资源,不但在手语的手势、句法、语义和语用等层面拥有丰富的语言资源,而且中国手语方言也是世界上最多的,其手语的多样性也为世界所罕见.面对如此庞大、丰富的稀有资源,我们却面临着尴尬的局面:手语资源的利用率极其低下,没有有效的手段去挖掘.这种情况类似于手语这个知识宝矿被深埋在海底难见天日,而我们缺乏有效的探测和挖掘手段,从而无法实现随心所欲的大海捞

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. http://www.cdpf.org.cn/sytj/content/2012-06/26/content_30399867.htm, 2016.4.26.

针，只好无奈地望洋兴叹. 与此形成鲜明对比的是，随着互联网时代海量数据的爆炸式增长，日益增长的以声、像、图、文为载体的 web、软件等为传统语言计算提供了大容量、多样性和高增速的语料资源，从而人类可以根据需要进行深入分析，挖掘出这些语料中所蕴涵的知识，这方面最显著的成果是 MIT (Massachusetts Institute of Technology)通过使用 37 种语言的大数据语料验证了依存距离最小化的存在，这一成果发表在美国 PNAS(*Proceedings of the National Academy of Sciences of the United States of America*)期刊上[1]. 作为一种典型的人类语言，手语计算还仅停留在语料收集和人工标注阶段，根本无法上升至构建于其上的知识组织、归类分析及深层挖掘.

为了有效地利用互联网时代的海量信息，语言计算已成为信息科学的重要支柱. 语言计算可帮助我们获取语言中的信息，然而认知心理学告诉我们，人类运用自然语言进行交流获得的效果中，讲话内容仅占 7%，强度和语调占 38%，而面部表情和肢体动作却占了 55%[2]，其中，强度和语调涉及情感的加工计算，面部表情和肢体动作涉及手语的加工计算. 李德毅院士借此强调长达半个世纪的自然语言处理学科却仅关注于讲话内容的理解，对以面部表情和肢体动作为代表的手语计算却关注甚少①. 手语同传统语言(spoken language & written language)一样，都是人类在对客观世界感知体验和认知加工的基础上形成的产物，但大脑处理手语的机制不同于传统语音，如传统语音以时间序列为基础，而手语除了体现时间序列外，还体现在以空间运用和动作感知为基础. 语言学家指出通过手语可以把语言特征与人类本身分离而单独进行研究，促进我们对人类语言本质的认识，探究人类语言的结构、儿童语言的习得和人脑语言的认知机制等等[3]. 若能在手语计算上取得重大突破，则可解决手语信息处理的难题，使得扩展语言计算理论成为可能，具有广阔的应用前景和理论价值. 由此可见，其研究不仅对计算机科学、语言学，而且对神经科学、人工智能和认知科学的发展起到推动作用，对自动问答、信息抽取等 NLP(Natural Language Processing)应用，都能提供重要的资源和技术支持.

目前造成手语计算困境的根本原因在于手语自身特点所导致的先天性困难以及当前语言计算技术的局限. 困境就是机遇，目前的互联网分析、永不停止的语言学习等基本需求倒逼手语计算尽快走出实验室而置于互联网之中. 互联网时代给手语计算带来机遇和突破，使扩展语言计算理论成为可能. 本文主要从自然语言处理的角度，回顾手语计算在过去 30 年的发展历程及主要成就，重新对手语计算的现状进行阐述和讨论，并对未来的趋势和挑战进行了分析. 手势(语)识别很大程度上属于图像处理领域，因此本文不作为重点进行介绍. 第 2 节详细分析手语计算与语言计算的关系；第 3 节对手语计算进行回顾；第 4 节分析目前手语计算的机遇；第 5 节提出手语计算未来的研究方向；最后是全文的总结与展望.

2　手语计算与语言计算的关系

语言计算这一术语最早由孙茂松先生[4]于 2005 年提出，随后开始有更多的文献采纳这一术语，但均未给出清晰的定义. 俞士汶先生[5]认为语言计算与计算语言学没有实质性区别，认为语言计算包括词法分析、句法分析和语义计算，并不包括语音处理，即语音识别和语音合成. 本文借鉴语言计算的说法，将以手语为研究对象的语言计算称为手语计算，并与包括手语动作识别与合成在内的手语图像处理相区分. 正如语音识别与合成也需要传统语音的计算理论一样，手语合成、手语识别并不是单纯的手语动作合成、手语动作识别，手语合成是计算机根据输入文本语义(自然语言处理领域)，合成出手语动作的连续图片或者动画(计算机图形学领域)，即研究如何计算动画参数使虚拟人表达的动作与输入文本在语义上保持一致(手语合成还有一类方向即研究增强虚拟人模型的视觉形象真实感，此方向与手语计算无关，不在本文讨论之列)；手语识别包括手势动作识别(计算机图形学领域)和手势含义识别(自然语言处理领域). 由此可以看出，手语识别或合成需要用到手语计算的知识，是手势动作识别或合成与手语含义识别理解的综合，是人类操作计算机进行识别理解或合成. 因此我们将手语动作识别与合成从手语识别与合成分离出来，限定手语动作识别与合成均没有语言的成分，属于计算机图形学的范畴，将“手语计算”归属于自然语言处理的范畴，是对手势含义的识别和理解.

传统语言计算和手语计算具有差异，不仅是因

① 李德毅. 大数据时代的认知计算，https://www.csdn.net/article/2013-11-13/2817475-MDCC-Big-Data-Cognitive-Com-puting，2017.11.16.

为手语缺乏和书写系统相关的信息处理基础，更不是简单地将传统语言与手语一一对应的翻译.问题的本质在于现有传统语言的计算理论是建立在单信道的基础上的，而手语计算是基于多信道的，将传统语言的单信道计算理论应用于手语的多信道计算技术其实不是一件简单的事情.传统语音的输出一般以语音为载体，是随时间推移而变化的一组数值[6].传统语言的书写系统也是如此，它只需要记录语音对应的书面符号，其书面符号和语音都是基于时间轴的数据流，同样都是单一的信道.这种语音或书面字符串构成了传统语言的自然语言处理系统的基石.而手语的本质是多信道载体，不仅难以将手语编码成线性单信道字符串，即使最终能编码成单信道字符串，势必会在各级加工过程中遗失很多载有语言信息的细节，因为手语语言学家认为手语的手部形状、手部位置、手掌方向、头部动作、眼睛凝视方向、面部表情、肩部动作和躯干姿势等这些信道都包含语言学意义上必不可少的信息，这些信道信息互为依存，相互联系，缺一不可.

正是以上本质的不同之处，造成了以下传统语言计算与手语计算的差异：

(1)传统语言计算的根本任务在于“消歧”，贯穿到词法、句法、语义等各个层面.在手语计算里，消歧也是任务之一，但不是核心任务.手语本质是多信道的，如果一个手部信道具有很大的不确定性，则面部表情、肢体动作等其他信道所携带的信息能够减少这种不确定性，甚至可以完全消除这种不确定性.目前的认知神经科学研究已证实了这一点，指出听障者对手语理解的过程与健听人有着显著的差异，因为听障者只需较少的语音信息即可辨别单个手势，并且辨别时间比口语单词更短.这种语音信息更多地受限于手语的音位结构、早期同步可用性(early and simultaneous availability)，这两者可能会共同促成手势的快速识别.有文献表明手势作为视觉信号，本身决定了它可提供大量的早期同步音韵信息，通常手势动作在大约 145 ms 后，其发音部位和手掌方向可被识别[7]，大约 30 ms 后其手部位置、形状和手掌方向等可被识别.这种早期同步可用性显著地缩小了心理词典的候选手势队列.其次手势音系和语素结构可能不同于口语.如口语里的花园路径现象(指语言处理过程中一种特殊的局部歧义现象)在手语里并不常见.此外口语里有 30 个以上单词共享[kan]、[mæn]和[skr]等音标，而手语里很难发现有多个手势共享一个初始音韵参数(即相同的手部配置和目标位置).这个音位结构同样也限制了候选手势初始队列的大小.以上心理学发现表明听障者能够利用一些视觉线索预测手语的词法结构.

通常消歧就是要消除语言中的不确定性，它与语言的信息量相关，在这方面，传统语言的信息熵研究文献较多.汉语被公认为是最简洁的语言，其信息熵较高[8]，因此汉语的消歧相比其他语种的传统语言成本更高、效率更低，需要更多地用到语境和世界知识，即语用知识.关于手语的信息熵尚未见到报道，但根据我们自建的手语语料库进行的统计，一个手势的最大长度是 8 个词汇，约 16 个汉字.比如汉语“打篮球”有两个词，但在中国手语里是一个手势.由此推测，手语的信息熵应比传统语言要高.实际上手语语料库中手语的语法比口语简单，很少见到长难句，并且手语每个信道的熵值还不同，其中面部表情、肢体动作等这些非手动特征信道的内容可视为语用知识，这些信道内容导致读者依赖语境就能获得超过传统语言单信道传递所需信息量的信息，因此在同等信息熵的情况下，手语的信息冗余度应比传统语言要高，从而起到了缓解熵值和消除部分不确定性的作用.由于计算手语的信息熵需要较大规模语料库的支持，因此有待于具体实验的验证.

(2)与传统语言计算相比，手语计算的核心任务是将单信道表征和多信道表征相互转换.目前传统语言的计算理论大多集中于计算单信道的码字平均长度，对多信道关注甚少.Shannon 第一定理指出码字的平均长度只能大于或等于信源的熵.传统语言计算主要关注于怎样构建一个具体的码字，使得单一信道在信息传输速率不大于信道容量的前提下实现可靠的通信.而手语不同，在为手语计算建立最优信道编码系统时，需要求出多个信道信息容量之和的最优解，从而使得只要信息传输速率小于信道容量，编码系统就可以使信息传输的错误概率任意小，即手语信道编码需要实现一维到多维的演变.手语的熵值越大，其输入输出的信息量也就越大，对多信道的考验就越高.目前一些手势输入输出设备尚未普及，最重要的原因就是多信道输入输出的问题没有得到很好地解决，导致一些手语输入输出设备的工作效率与传统语言相比非常低.因此我们亟待解决手语的输入输出问题，发展多信道编码的理论.将语言计算的研究重点逐步过渡到多信道信息编码之中，带动传统语言与多信道编码理论并轨，形成最优的信道编码系统，从而提高通讯的效率.

(3)手语的多信道性质造就了视觉空间的立体

性特征，这种特点对于传统语言计算是一个极大的挑战. 传统语言计算的先天性缺陷就在于传统语言本身是单信道的，无法模拟手语三维场景中实体运动对象的空间布局. 如果需要用计算机将传统语言翻译成手语，Huenerfauth[9]认为计算机需要模拟手势者心理的三维空间，然后把语言里涉及的实体对象映射到心理空间，最后再用手势映射到物理空间，以表达源文本的含义. 以句子“轿车在房子旁边”为例，需要选择表征实体对象“轿车”特征的分类词手形(即手形闭集)，这些特征包括四轮、小型交通工具、停止状态等，还需要考虑所要表达空间的特征，如房子的大小、形状、轿车在房子哪个位置等. 最后在手势者伸展两只手的范围内选择哪个位置来代表施事者轿车和受事者房子，选择后还有一些因素需要考虑，如抽象维度和其他对象的属性等，进而相应地实施手部运动完成表征. 此外还要根据语境来配合眉毛和眼睛的动作(愁眉)、脸部表情(苦脸)，甚至根据表达的需要辅以夸张性动作，如头部和躯干动作. 由此可见，从单信道向多信道转换涉及到复杂的场景加工和空间隐喻，此外像基本常识和世界知识也是必不可少的. 目前传统语言里的文景转换主要考虑到空间实体的部署，尚未涉及到多信道的转换加工. 即手语计算考虑的空间概念比传统语言更为精细，在实现空间概念转换这一基本任务时，需要补充更多的手语描述中缺失的信息.

(4)空间关系是手语的最基本关系. 长期以来早期的手语计算存在着一个误区，即把文法手语(按照汉语顺序打出来的手势序列)作为研究对象，这种手语语法强调以口语线性序列的方式表达，并没有空间性. 自然手语的语言管道是肢体与视觉，用到了空间性特点[10-12]. 许多心理学研究表明文法手语没有使用空间性而自然手语用到空间性，可能是听障者对这两种手语有不同理解表现的原因，即文法手语的理解就比自然手语困难[13]. 学者们已注意到手势者身体前部的手语空间代表不同的意义，如 Sutton-Spence 等人[11]提出将手语空间分为拓扑空间与句法空间. 其中拓扑空间(指将实际空间里的对象位置映射到手语空间里的对应位置)一般用来说明人或物体等对象的位置和运动方向，因此手语可以便捷精确地用拓扑空间表达实际空间关系，从而建立手语空间关系与实际空间关系的对应关系. 比如描述驾车的场景时，可打出“轿车”手势在手势者前面空间运动，当表达弯弯曲曲时，不断地前后转向；当表达崎岖不平时，上下运动来表示. 这时手语很自然地完成了对真实空间的描述，由此可见手语与空间的关系密不可分. 因此空间计算是手语计算跳不过去的课题，从而空间建模、空间隐喻、空间语义等概念贯穿了手语计算的词法、句法、语义和语用等各个阶段.

(5)空间关系对手语计算的影响. 有学者以听障者为研究对象，比较了这两类空间的认知加工差异，发现听障者在看过拓扑空间句子后，完成判断题的反应速度要比看过句法空间的更快，这说明听障者对于这两种空间有着不同的认知加工过程[14]. 脑损伤案例也支持了这一观点[15]，从而从认知神经科学的角度说明，手语的空间特性存在不同的层次，这种空间关系的特点对手语计算有很大的影响.

以词法阶段为例，空间建模主要是对非真实性空间的运用，在代词运用、后文提到的呼应动词、比较手势等应用比较多. 这些词法会根据主语和宾语出现的位置而改变手势动作方向，以此来呼应主语和宾语的关系，并未涉及真实空间的描述. 其中在代词运用上，手势者一般用靠近躯干的位置来指代人、场所或物体，并利用空间距离来表示指代对象之间的关系，从而利用空间实现了代词功能，这与传统语言有较大差异. 首先传统语言使用单信道容量有限，而手语使用多信道，理论上可无限次划分，因此手语里的代词所指数量可达到无限次；其次传统语言代词一般指一类对象，而手语里的代词所指更具体，即某个实际对象[16]. 因此手语里运用空间指代事物及其方位很方便形象，但其指代对象较多时，容易造成混淆，从而给词法计算带来了不确定性.

再以句法为例，很多句子借用真实性空间特性来呈现，比如方位词句子、下文提到的分类词谓语等，分类词谓语计算的分类词系统不同于其他手语现象的地方就在于它需要将语言和空间特性相结合，具体是将分类词手形辅以运动，构成包括一个或多个对象的复杂方位和谓语，从而表达分类词谓语的空间概念，而其他手语现象就没有这个功能. 由于采用特殊的视觉表达方式，分类词谓语在用两个手势来代表主语和宾语，并通过运动来表达两者空间关系时，手语可以不必依赖传统语言中单信道表征的空间介词来表示方位，从而借助空间场景的类比表征建构分类词的多信道——分类词手形. 这种单信道表征和多信道表征相互转换涉及复杂的场景部署加工运算.

(6)空间计算与空间隐喻密切相关. 空间计算涉及到单信道表征和多信道表征相互转换，而空间隐喻是指将空间方位投射到非空间概念上的隐喻.

手语以空间为概念框架和表达中介，更加依赖高效连贯的空间隐喻.例如听障者无法用相应的词汇来表达成功或失败的含义，但可以用向上或向下的手势进行表达，这是一个空间隐喻的过程，实际上也是将单信道向多信道表征转换的过程.传统语言也存在着类似垂直方向上的空间隐喻，如东亚语言(包括汉语、日语等)中广泛存在用“上”表示过去，用“下”表示未来的表达，因此手语中的空间隐喻与传统语言中的空间隐喻存在一定程度的对应.传统语言的隐喻理解主要关注如何制定模型与算法来获取相关知识，以实现隐喻的识别和解释，这些思路与方案能否用在手语的隐喻理解还有待研究，因为语言学界普遍认为手语隐喻具备了手语的像似性(iconicity)和隐喻性(metaphor)双映射的特点，这不同于传统语言隐喻的单映射[12].未来可通过认知神经科学来探究手语隐喻的大脑加工过程，进而推导出更科学的手语隐喻加工模型，从而实现识别和理解空间隐喻.

(7)各国手语计算存在着差异，但这不是手语计算的核心任务.因为任何一个国家的手语与本国的传统语言存在着关联，只要不是处理多信道的内容，手语的单信道计算完全可以借鉴传统语言的计算理论.即使手语和传统语言语法结构不同，背后的计算模型和数学理论大致是相通的.同样，不同国家之间的手语翻译也可以借鉴传统语言的计算理论，而且难度只会比传统语言要小.因为传统语言语音感知和手语存在着差异，即手语的发音器官(如手部、头部和身体躯干等)是完全可见的，由此表明手语的象似性比传统语言更为明显.从空间关系来讲，手语对物体间真实运动或静态方式的表达具有高度视觉象似性.因此手语计算与传统语言计算之间存在着差异和关联，并且不同国家之间手语计算也存在着差异和关联.目前手语机器翻译多指传统语言与手语之间的翻译，尚未见到不同手语之间翻译的研究报道.从研究价值来看，传统语言与手语之间机器翻译具有较强的理论意义和实践价值，而不同手语之间机器翻译的研究并不那么迫切，因此未来一段时间内传统语言与手语之间的机器翻译将是主要研究任务之一.

3 手语计算的发展回顾

从 1983 年 AT&T 最先取得数据手套专利开始，手语信息处理已有 30 多年的发展历史.纵观这些年的研究文献，与动辄每年百篇到千篇论文的手语语言学和传统机器翻译领域相比，手语计算有影响的文献并不多.以 ACLWEB 为例，截至 2016 年 4 月 20 日，ACL(Association for Computational Linguistics) Anthology 已收录了 36 000 篇论文，而与手语有关的论文才 76 篇.与此形成鲜明对比的是，手语动作识别的文献丰硕，以 Web of Science 为例，以 Sign Language Recognition 为关键词的论文就有 3454 篇.因此手语计算仍处于起步阶段，还有很多工作要做.手语计算的基本问题包括手语编码、空间建模、句法分析、语料库建设、语用计算、机器翻译等环节.

3.1 手语计算与手语图像处理的关系

尽管手势动作识别属于图像处理领域的范畴，并且大部分文献确实使用了图形学特征来识别手势，然而也有一部分文献为了提高手语动作的识别率，或多或少地使用了手语语言学的一些知识，最典型的就是 Stokoe 理论[17]，该理论认为手势同时由三部分(参数)组成，包括手势位置、手形以及运动轨迹.20 世纪 70 年代初期，语言学家给这 3 个参数增加了一个要素：方向.之后语言学者大都采用 4 个参数.1999 年 Vogler 等人[18]第一次基于 Stokoe 理论提出的双手手形和运动参数，结合 PaHMMs 模型来识别手语.此后 Stokoe 理论在手语识别文献中得到了广泛的应用.如国内姜峰等人[19]基于 Stokoe 模型提出了手语力效要素的定义和描述方法，继而给出了非特定人手语数据的规整策略，并用于手语识别.Lichtenauer 等人[20]为未登录手势提出一种自动构建分类器的方法，具体做法是收集很多手势者的手势特征，并与新手势的特征进行比较，从而为目标新手势构建分类器模型.这种方法依赖于很大的基础特征训练集(75 人打出的 120 个手势)，并允许使用一次性学习来训练一个新手势分类器.Bowden 等人[21]还提出使用一个训练样本就能正确分类新手势的手语动作识别系统，并给出了一个单一样本的训练案例.他们的方法使用了两级分类器，其中初级分类器使用硬编码检测手形、布局、运动和位置.在应用动态分类手势前，二级分类器采用 ICA (Independent Component Analysis)对 34 维的初级特征向量进行有效地噪声消除.在少量训练实例和缺乏语法知识的情况下获得了很好的效果.Kadir 等人[22]扩展了前述工作，并基于 Boosting(级联弱分类器)、躯干为中心的描述(将运动规格化为二维空间)来检测头部和手部的运动轨迹，然后使用两级分类器，其中初级分类器生成语言学的特征向量，二级分类器在马尔可夫链上使用 Viterbi 方法获取最高识别概率.Cooper 与 Bowden[23]延续了这项工

作，在初级分类器的基础上还检测了训练样本和分类样本的手势特征，接着二级分类器使用第一阶马尔可夫链来获取概率最高的结果. 这些文献表明基于 Stokoe 语言学特征，其手语识别率等同于基于图形学特征的识别率. 这显然有违引进语言学特征的初衷，其根本原因在于 Stokoe 模型认为包括手形、运动、位置等在内的音系参数是同时出现的. 由于手势的轨迹是时序性的，在传统语言中声音也是线性时序性的，因此手语也应与传统语言一样，是时序性表征(sequential presentation). 音系结构通常通过形态变化来表现，如 Wendy Sandler 认为美国手语(American Sign Language，ASL)只有两种基本音段，即运动和位置这两个音韵参数，分别对应传统语言里的元音和辅音[24]. 同样还有 Baus 等人[25]认为，手语拥有与传统语言类似的“辅音-元音-辅音(CVC)”结构，具体就是“位置-运动-位置”结构，但是与语义相关的只有位置参数. 至于方向、手向两个参数，虽然可能会对大脑语音加工有影响，但尚未发现其作用，这些理论并未见到应用于手语图像处理的报道.

目前众多学者基于图像处理理论，从前端的数据预处理，即手势的跟踪与分割，到鲁棒的特征提取与分析，以及手语的统计建模这几个阶段进行了大量且深入的研究[26-28]，但是手语识别仍集中在单音节单个手势的研究，连续手语识别还未取得突破性进展，识别准确率仍不理想. 并且这些手语识别仍处于实验研究阶段，是面向特定手语使用者而设计的，离真正实用化和商品化还有很长一段距离，因此引入手语计算理论，并进行创新已势在必行.

3.2　词　法

不同于传统语言的最小语义单位——词(word)，手语里的最小语义单位为手势(sign). 目前对手语的词法研究集中于手语编码、手势切分、手势空间建模等工作上.

3.2.1　手语编码

Yao 等人[29]讨论了中国手语的信息表示，介绍了现有手语书写系统 Stokoe[30]、HamNoSys[31]、Sign Writing① 等，并指出这三个系统存在着易理解和易机读两者不可兼得的问题. 从手语编码的角度来看，这些系统不仅考虑了手势本身的含义，还考虑了四个音韵参数，其中德国汉堡大学设计的 HamNoSys 系统，目标就是作为手语的音标系统来使用，这类似于 IPA(International Phonetic Alphabet)国际音标，它充分考虑了手语的音韵特征，如图 1 所示，因而被广泛用于手语的机器翻译和手语的三维模型生成.

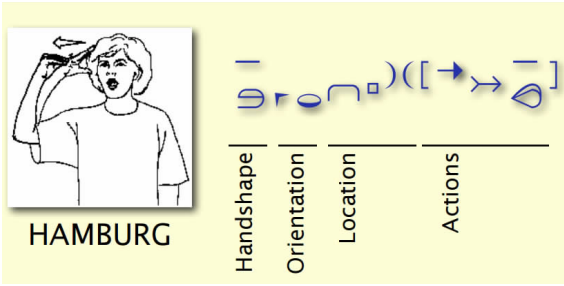


图 1　手势“HAMBURG”的 HamNoSys 编码(图片来源于文献[31])，该编码融合了手形、方向、位置、运动等四个音韵参数，其中手形符号表示 C 手形(以便与 A、B 等字母手形和 1、2 等数字手形相区分)，方向符号表示远离躯体，向左上方方向(以便与其它方向相区分)，位置符号表示位置在前额上(以便与胸部位置、嘴部位置等其他位置相区分)，运动符号表示直线运动，同时手形变成数字 0 手形(以便与其它曲线运动、圆弧运动等相区分). 这些合成起来表示“HAMBURG”的含义

从文献来看，很多国家的手语语料库都使用了本国传统语言作为转写语言. 因为目前传统语言的计算技术已比较成熟，若本国手语能使用本国传统语言进行编码，则可充分利用和借鉴本国传统语言计算的成熟技术. 其次，目前手语转写主要是人工转写，依靠转写人员(包括听障者和健听人)对手语语料进行转写，若采用本国传统语言进行编码，有利于减少培训转写人员的时间和精力成本. 最后本国传统语言是本国听障者的第二语言，听障者一般是双语者，因此将本国传统语言的书写系统作为本国手语的编码，听障人群也能接受. 显然，采用本国传统语言进行手语编码也会遗失很多携带信息的细节，但这是目前成本最低、见效最快的手语编码方案.

因此手语编码目前有两种方案，一种是创建新的编码方案，综合考虑语音和语义等因素，使之更适合计算机处理，如 HamNoSys 系统；另一种是使用现有的本国传统语言编码方案，并针对手语进行改良，使计算机处理手语更为方便. 由于前者学习成本高和受众少，后者普及率更高. 这类似于汉语编码，汉语编码有很多种，如基于汉字图形的编码、基于汉字组件的编码、基于笔划和偏旁部首的编码等，还有汉字的多维编码，即综合考虑字音、字形和字义等问题. 最终最受用户欢迎的汉字编码是基于拼音的连音输入法，即用户仅输入连续拼音字符串，程序自动进行检查和分词，并结合上下文信息自动调整并给出对应的汉字，目前普及率更高的搜狗输入法即为典型

① Sign Writing. http://www.signwriting.org/, 2016.4.23

的代表.这种输入方案强调以人为本,尽量减轻用户的记忆负担,降低用户入门的门槛.因此预计很长时间内,使用本国传统语言书写系统为本国手语进行编码,并配套一些改良措施是手语计算的发展趋势.

3.2.2 手语切分

手语切分根据实际用途分为两种,分别是手语图像处理的手势切分和转写文本的手势切分.

手语图像处理的手势切分:传统语言的语音识别需要定义音系学模型以便进行音节切分,需要人为定义声母段、过渡段、韵母段、闭塞段和停顿段等多个细节,并进行较清晰的划分.在进行手语图像处理时,也需要定义手语的音系学模型,以便进行手势切分,由此可见手语计算和手语图像处理存在着密不可分的关系.1986 年 Liddell 等人[32]提出了经典的运动-保持模型(Movement-Hold Model),该模型认为手势是包含音节的,具体可分为运动和保持两个音位音段(phonological segments),它们按序列生成.这些音段都包含了完整的手部配置和语音特征(手形、运动、位置和方向),这些手形、位置、方向和非手动特征的信息通过每个单元一系列发音特征表现出来.根据这个理论,1999 年 Vogler 等人[33]使用 HMM(Hidden Markov Model)对 22 个词进行了手势切分,证实了运动-保持模型的可行性.此后,他们以该模型为基础,提出了 ASL 识别系统的框架,证明了该模型的有效性,即能够处理手语多信道的表征问题.此后一些手语识别研究均采用了此模型[34].

也有研究从其它角度将手势的运动单元划分成子单元从而实现手势切分,如有些文献指出手语的音节应围绕音核进行组织,衡量音核的单位就是视觉效果最强,不同的是手语音核前后不分响度的高低,只要是视觉强度较差的要素即可[35-36].1990 年 Wilbur 建议像速度和加速度这些物理因素可作为反映手语响度的基础[37].Kong 等人和 Han 等人[38-39]据此将手势的运动单元自动分割成子单元,即运动模式开始时手部速度的加速度和模式结束时手部速度的减速度,利用路径和加速度的不连续性,来表明音段的开始和结束,然后使用动态时间规整(Dynamic Time Warping,DTW)距离测量或特征主元分析(Principal Components Analysis,PCA)汇聚成一个可能的样本路径.但这些论述在手语语言学界未达成共识,仍存在争议.因为很多学者认为 Stokoe 模型实际上是语素(语素是最小的语义单位)[40-43],但不是最小的语音单位(音素是最小的语音单位).这种差异在于手语的表达具有同时性和序列性,进行手势切分时需要很好地解决这个问题.

连续手语识别的手势切分关键在于识别手语的音变现象,因为手语与有声语言口语一样,连续手势序列并不是单个音节的简单组合,手语句子里每个手势的组成部分以不同顺序组织,而且相互影响.受协同发音、韵律等因素的影响,手语也存在音变现象,从而导致连续的手势序列与单独的手势音节有很大的不同.以中国手语为例,目前已发现中国手语有四个音变现象,分别是运动增音(movement epenthesis)、保持缺失(hold deletion)、音位转换(metathesis)和同化(assimilation).其中保持缺失的例子可见例 1 和图 2.

例 1:

北京手语:光暗

汉语:光芒暗下来

图 2 汉语句子"光芒暗下来"的手语打法

完整过程如表 1 所示.

表 1 手语句子"光芒暗下来"的音变过程

	基本手势:光芒暗下来						
	H	M	H		H	M	H
移动增音	H	M	H	M	H	M	H
保持缺失	H	M		M		M	H

其中,M 和 H 分别表示运动(Movement)和保持(Hold)两个音位音段.这是中国手语常见的一个音变过程,但对于这些音变现象建模的研究很少.1999 年Vogler 等人[33]就 ASL 的运动增音进行了建模,使用 HMM 对 22 个词进行了手势切分,验证了音变现象建模的必要性.2007 年 Fang 等人[44]研发中国手语大词汇量连续手语识别系统时,也以运动增音为基础提出了过渡运动模型(TMMs),应用 K-means 聚类算法和 DTW 得到了 91.9%的识别率.这些研究仅用到了运动增音一个音变现象,其

他音变现象未覆盖到. 因此只有精通手语语言学和熟悉手语内部独有的规律和特征，才能深入开展手语计算研究，否则无法将手语计算从手语图像处理中分离出来. 另外，手语动作识别离不开手语计算，仅有计算机图形学的知识是不够的，必须有手语语言学的知识支撑. 如 Wang 等人[45]指出连续手语识别无法套用传统语言中语音识别的二元(bigram)或三元(trigram)模型等上下文相关知识，因为手语中并未定义音素，从而无法为语音建模，这种论述招致手语语言学家的批评. Vogler等人[46]指出手语动作识别由于手语本身的特点而无法套用传统语言计算常用的与上下文相关的 HMM，因此需在手语音系学理论的基础上，建立手势的过渡运动模型.

转写文本的手势切分：要对手语进行信息处理，首先将手语识别转写成文本，大量文献表明[47-52]，目前各国手语语料库的视频语料一般是用本国传统语言的书写系统进行转写和标注的，因此理论上沿用现有的传统语言切分模型来实现手势切分是可行的①. 由于传统语言的一个词甚至多个词可组成一个手势，类似于组成词组. 但两者的意义不同，传统语言里有很多词组其意义并不是其构成部分词汇意义的简单组合，比如“人民”和“大会堂”这两个词的意义并不能简单地合并成词组“人民大会堂”的意义. 而手语的手势可认为是组成词汇意义的组合，类似于固定搭配，如踢足球. 此外，传统语言的一个词对应两个以上手势的情况大量存在，因此为了切分的需要，在转写时，也需提醒转写人员不能将这些特殊手势转写成一个汉语的词，而必须是一个手势一个词. 需要说明的是，传统语言需要分词是因为像中文之类的文本通常由连续汉字序列组成，词与词之间缺少天然的分隔符，所以中文信息处理比英文等西方语言多一步工序，即确定词的边界. 但是像中文文本的手势切分不仅是确定单个手势的边界，还需要确定由多个手势组合而成的复合手势，这种情况也适用于英语之类的西方语言转写文本，因为包括汉语和英语在内的传统语言转写文本里一个词可能对应多个手势，多个词可能对应一个手势. Yao 等人[53]总结归纳了重叠手势、书空手势、复合手势特征的切分知识，应用条件随机场(Conditional random fields，CRFs)进行了手势切分，取得了 F 值 77.4% 的切分率，证实了转写文本的手势切分可行性.

3.2.3　词性空间建模

目前手语词性与传统语言相比，特殊之处在于手语词性除了需要表征词汇本身含义的手势以外，还需要一系列的辅助动作来表达词性信息，即手势者在谈话时会在其躯体周边的空间部署占位符来表征话语里的个体或对象，并通过一系列动作来表达词性的语法信息[54]，以便实现单信道向多信道的转换. 以最常见的动词为例，由于动词涉及到主语和宾语需要进行空间部署来表征，这种动词空间属性导致动词被语言学家划分为三类：简单动词、呼应动词和空间动词[11]. 由于简单动词主要是通过眼睛注视等非手动特征来辅助手部动作完成的，其特点是不通过手势的移动空间来显示语法信息，没有手语语言学中的人称、数或处所词缀等屈折形态标记[55](屈折词是由屈折变化所构成的词，是为了限定某词的语法功能而添加词缀或改变词形). 在转写该手势时可认为等同于传统语言的一般动词. 难点在于其后两类动词，呼应动词使用句法空间，而空间动词使用拓扑空间来表明其语法关系. 其中呼应动词需要指示位置决定运动路径的方向，允许包含人称和数等屈折标记. 如例 2 和图 3 所示. 它们都是通过在语法空间移动来实现的，这是手语与口语的不同之处. 传统语言一般使用虚词等语法手段来表达显性方向，但只有手语的呼应动词可以用视觉上的方向作为词形方向，来表示语义的概念.

例 2：

武汉手语：教(面向第三人称)电脑(中国手语里经常省略“我”)

汉语：我教他电脑

(a) 教　　(b) 电脑

图 3　汉语句子“我教他用电脑”的手语打法

① 实际上采集、转写、标注手语视频非常繁琐且任务困难，众多学者指出在众多语料标注中，唯有手语视频标注的 RTF (Real-Time Factor)为 100，意指 1 h 的手语视频语料需要 100 h 做标注. 因此标注人员也不可能花大量的时间来标注完整的语言学细节，包括句子类型、主手/辅手类型等. 最常见的标注是标注人员根据手语视频直译的文本. 限于时间限制，标注人员也不太可能为这些直译文本添加手势的边界以及复合手势标记. 因此不管是手势识别成文本，还是人工翻译手势成文本，对手势的切分是绕不过去的问题.

对于呼应动词的理解和生成，由于呼应动词在计算机识别或者转写时可识别出人称、数等屈折标记，在理解时没有太多的困难.目前工作重点集中在呼应动词的生成上，2004 年 Toro[56] 使用了 6 个呼应动词，利用已有的文景转换软件生成了 42 幅动画，第一次考虑了这个问题，此后 Toro 在其博士论文中设计了动画算法以便生成 ASL 的一些呼应动词手势，包括相关主语和宾语位置、动词运动路径的建模，得出的经验是需要考虑动词的引用形式、语言学特征、几何信息才能顺利生成呼应动词的手势动画，把呼应动词的人称、数等屈折标记表达出来[57].但是他的工作仍需要被试去寻找视频里的手部位置，并写下角度和坐标，然后另一个被试去寻找运动的模式，未用到机器学习方法.2009 年 Segouat 和 Braffort[58] 在研究法国手语时，定性分析了视频里手势者的运动路径，并训练了两只手之间的运动模型，来达到生成呼应动词手势的目标.但他们的研究是基于手势视频里的二维图像来获取运动数据，易出错且效率低.2010 年 Huenerfauth 等人[59] 采用了动作捕捉传感器，事先收集了手势运动数据，以此为基础对呼应动词进行了建模，通过建立手部运动的数学模型，对样本数据进行训练，确定最佳拟合训练数据的 3 阶多项式系数——最小二乘系数.然后在呼应动词的手势开始时，给定数学模型预测右手的坐标位置，给出主语和宾语在手势者周围的圆弧位置，如图 4 所示.评估结果表明生成的动画效果与真人动画效果相当.2011 年 Duarte 等人[60] 也使用动作捕捉传感器做了同样的工作，但其重点在于重组了动作数据的元素并合成了手势动画.2012 年 Lu 等人[61] 在前者的基础上，提出了基于向量的学习模型，认为呼应动作最重要的是手部在空间的运动，而非手势的起点和终点，因此建立的运动向量模型仅用到手部的三个坐标值，并未采用能够表征起点和终点的原 6 个值，而是单独设计了高斯核算法来预测手部位置，因此他们的多个手势者的实验数据结果表明在呼应动词手势生成上基于向量的学习模型优于以往的学习算法.这些学者在解决呼应动词的问题上做了有益的尝试.

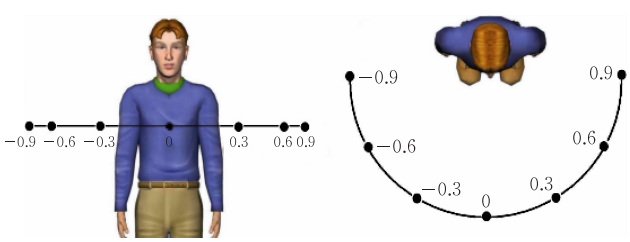


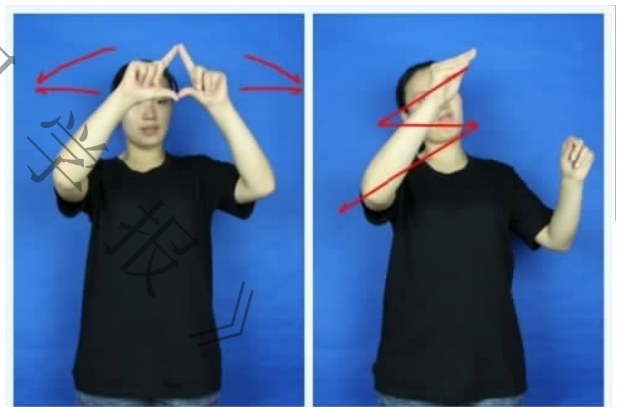
图 4 手势者呼应动词手势的数学模型

最后一类动词即空间动词，可以通过屈折变化表征方式和位置而非表征人称或数量.如例 3 和图 5 所示，可见要先打出主语的完整手势（例如叶子），其次是分类词手形词素（例如植物叶子），这种手形代表一类对象（即陆地上的花草和水中的水草等所有植物叶子类，参见图 5(b)).可见，该空间动词给出了以下信息：路径、轨迹和动词所描述的动作运动速度以及有关动作的位置.空间动词有两部分：运动语素和分类词手形语素[40].空间动词只是表征了位置和语义分类词上的共现，因此它属于分类词谓语范畴的一个子类.由于分类词谓语在手语计算中的重要性，我们将单独用一节篇幅介绍分类词谓语计算.

例 3：

北京手语：叶子 CP(Classifier Predicates)：叶子$_{形状}$_落下来.

汉语：叶子落下来了.



(a) 叶子 (b) 电脑CP：叶子$_{形状}$_落下来

图 5 汉语句子“叶子落下来了”的手语打法

3.3 句 法

手语句子与传统语言的不同之处在于手势者使用了非手动特征来表达句子的含义，具体地讲就是使用面部表情、头部动作、肢体动作等多信道来表达对应口语句子单信道的含义；其次是根据视觉优先原则和主题化等因素，表达的手语句子语序稍微自由，不像传统语言句子成分顺序基本是固定的.如何用句法分析树来表示多信道的内容，众多学者提出了各自的解决方案，如装饰字符串[62]、多维度树结构[63]、NaÏVE3D 树[64] 等，其中 NaÏVE3D 树的例子见图 6.此外他们还提出了通过语法生成这些树的结构[65-66]，Yao 等人[29] 已给出了综述，在此不再赘

述. 总之这些均属于语法驱动的分析方法，要生成它只有手工编写规则和从训练数据中推导规则两个途径，这些均需要大量的人力或标注语料（树库）的支持. 遗憾的是至今没有适用于手语的树库和数据驱动的句法分析方法.

手语里稍微自由的语序主要受语篇因素（如主题化）的影响，因此我们常见到听障者使用不同于汉语语序的各种各样的语序和手势来表达同样的意思. 除了语序与传统语言不同，非手动特征有时也会充当句子成分，如使用头部倾斜、眼睛注视也可能会略去某个传统语言句子的成分，如名词短语主语或直接宾语. 因为在手语会话中讨论实体往往与手势者周围空间的位置相关联，头部倾斜或眼睛凝视往往对准这个位置，这样通过非手动特征，就已经表示了这个实体. 手语句法计算需要重点解决这个问题.

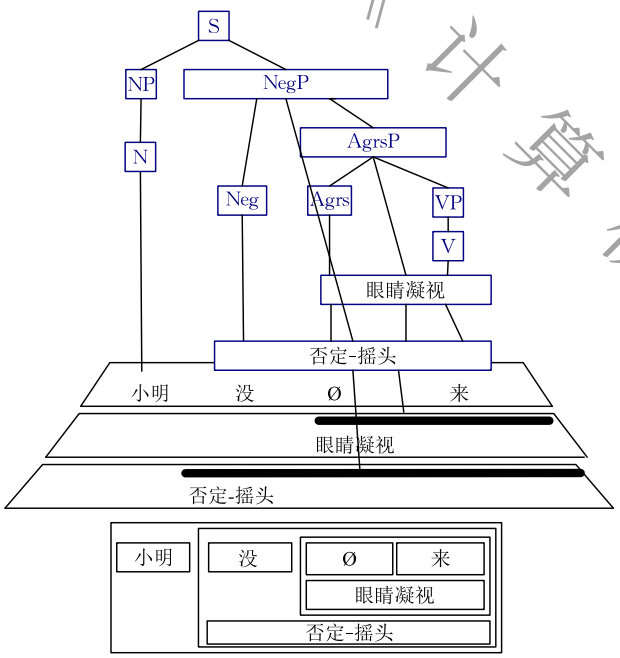


图 6　3D 语法树

此外，评价手语机器翻译的质量，应包含评定这样一句缺少成分的话是否传达了同样的信息. 从句法分析的角度讲，手语是满足投射性条件的语言，传统语言的一些理论可用于手语的分析，但手语相比传统语言多了空间属性，这不仅是个时间序列，因此手语的句法分析要比传统语言困难得多，至少目前还没有手语依存分析工作使用的语料和语料划分标准，现有的传统语言树库能否适用于手语还是个新的课题.

3.3.1　手语语料库

Yao 等人[29]对手语语料库进行了综述，目前大部分手语语料库用于语言学分析，很少用于机器学习. 为了建立一个能用于机器学习的手语语料库，与传统语言语料库一样，需要大规模视频语料自动标注技术. 但手语语料库与传统语言的不同之处还在于需要一个可靠便捷的方法为手语语料库建立一个手势者三维模型.

目前手语语料库主要使用本国传统语言的书写系统来标记手语，显然在此基础上是无法实现手语的空间计算，必须有一整套模型来记录手势的所有空间信息. 以中国手语为例，使用中国手语视频或动画形式将信息呈现给听障人群阅读，相比汉语文本，这种方式更能实现信息无障碍，毕竟中国手语才是听障者的母语，而汉语只是第二语言. 在手语生成系统或者翻译系统在生成手语句子时，目前的技术还不便于使用手语视频，虽然手语视频里手势之间平滑过渡可以通过合成得到，这也是手语识别和对话系统未来的研究课题. 但是要方便地与面部表情等非手动特征组合，便捷地频繁编辑或修改视频，目前只有动画或动画脚本可以做到，这就有必要为手语语料库建立手势者三维模型. 只有建立了手势者三维模型，语料库才有可能存储手部的空间运动等手动特征和面部表情等非手动特征、甚至手势者的手势速度等手语空间信息. 2002 年 Cox 等人[67]使用动作捕捉技术建立了手势动画的词条库，但他们的工作只是为每个手势记录单一引用形式，并没有为整个句子或话语建立标注语料库. 2004 年有学者使用计算机视觉技术来识别手语信息为手势运动建模，但受限于识别精度，这样的语料库三维模型并不可靠[68]. 2009 年 Segouat 等人[58]使用转描机技术（一种用来逐帧的追踪真实运动的动画技术）成功建立了法国手语语料库的三维模型，可以半自动地记录手势视频里的手部位置. 2010 年 Lu 等人[69]使用动作捕捉技术将手部、躯体、头部、眼睛的组合进行追踪，并由此创建了已标注的 ASL 语料库. 但这些需要手势者佩戴数据手套、可穿戴躯体传感器，配合眼动仪进行数据收集，过程仍然繁琐. 因此为手语语料库建立一个可靠便捷的手势者三维模型将是目前的重要任务之一.

有了手势者的三维模型还不够，还需标注必要的信息以便供机器学习系统进行训练，因此需要大规模手语视频语料自动标注技术的实现和配合. 如某手势者在谈论一位叫“小明”的同学，第一次提及小明后，该手势者会指向空间某位置，这样的空间参

照物表示小明,随后再次提及小明时,手势者只需简单地指向该位置即可.有时,一个手语句子并没有提到主语和宾语,而是靠手势者的眼睛凝视或者头部倾斜指向某位置的方式传达主语或宾语的身份信息[62].如果不标注这样的词法和语法信息,则手语动画脚本、动画、机器翻译等系统都无法自动处理手语的空间信息.遗憾的是目前尚未见到有关手语语料自动标注技术的报道,现有手语语料库仍停留在人工标注阶段.

3.3.2 分类词谓语计算

分类词谓语是手语里最常见的语言现象,一般由运动语素和分类词手形语素两部分组成,分类词手形是指表征分类词的手形,如例 3 里叶子形状的手形可以代表陆地上的花草和水中水草等所有植物叶子一类,这样就起到了分类词的功能[70].听障手势者平时在进行手语交流时,几乎每分钟就出现一次分类词谓语,某些类型甚至出现 17 次之多[71].因为生成手语动画,若要描述场景、发音工具、动作、大小以及其它视觉/空间或现场/过程的属性信息,分类词谓语是最理想的表达方式.但分类词谓语不能用传统语言学理论来解释.从 1960 年开始语言学家们一致认为,手语的语言学现象基本可以用传统语言学来解释,但手语中的分类词谓语却是一种独特的语言现象[72].由于分类词谓语是手语计算中最复杂的手语现象(通常与空间语义有关),它突破了什么是语言表达的传统定义,这些都是传统语言计算理论无法解决的,首要问题就是如何表征分类词谓语的手形和运动类型,映射到语义表征时如何转换?另外如何用到空间背景和世界知识,这显然是个难度很大的课题.为了实现分类词谓语的计算,需要将分类词谓语涉及的对象实体映射到心理空间和物理空间,即编码时需要将分类词谓语表达的每一空间信息都量化为一个语素,通常需要多个语素才能表达完整的分类词谓语的含义.实际上分类词谓语所要表达的空间信息远比想象的还要多,特别是描述分类词谓语涉及的对象实体之间空间关系的情况时,空间信息只会更复杂.Liddell 对简单的分类词谓语"一个人走向另一个人"做了统计分析,结果发现需要 28 个语素才能完整地表达空间信息,这些空间信息包括:这两个人面对面、两人之间有一个特定距离、直线路径走动、都在同一水平面上、两人站立在垂直方向上等等.据此他评价分类词谓语是非空间多语素结构模型,因为分类词谓语为了能够完整地表达各种各样的空间信息,所需要的语素数量可能是庞大的,甚至是无限的[73].

由于分类词谓语的表达是动态的,还需要把实体对象的相互作用和三维场景的部署限制编码成一系列的规则[74].仍以"一个人走向另一个人"为例,为了部署场景,除了需要决定实体对象的位置信息,还需要为实体对象"一个人"选择开始和结束的位置,是直线运动还是曲线运动,运动路线是颠簸的还是连绵起伏的,这样手势者才能流畅地表达运动路径.此外在表达该分类词谓语时,道路、地平面也会表达出来,为了防止出现人在地平面下运动的常识性错误出现,一些必要的生活常识和世界知识是必不可少的,如人一般站在地平面上等常识.由此可见分类词谓语的计算离不开大量的语义理解、空间知识和逻辑推理.

因此以上两个难点造就了分类词谓语计算的复杂性,使分类词谓语的计算显然超出了目前机器的计算能力,以致手语语言学学者将其评价为超语言的空间手语、构成空间参数化表达式等论述[74].由于分类词谓语是手语里最特殊的语言现象,同时其计算又最为复杂,我们有理由相信分类词谓语的计算或将成为手语计算皇冠上的一颗璀璨的明珠.国外的手语机器翻译除了 ZARDOZ 系统,其它系统几乎均未能解决分类词谓语的计算问题.美国学者 Huenerfauth Matt 指出,分类词谓语计算是手语计算的最终目标,只有实现了分类词谓语的计算才是真正的手语计算[9,75].

从计算本质来看,分类词谓语的计算仍属于单信道向多信道转换映射的过程,与其它手语现象不同的是,分类词谓语的计算还融合了大量的空间隐喻和场景加工计算.国内外尝试使用理性主义和经验主义方法来解决分类词谓语计算的问题,但是经验主义方法需要一定规模的语料库支撑,传统语言能够成功地使用经验主义的方法,是因为爆炸式增长的 Web、软件等资源为语言计算提供了大容量、多样性和高增速的大规模语料数据.而手语语料因为视频采集繁琐和标注困难,导致手语计算面临着严重的数据稀疏问题.此外即使解决了手语语料匮乏的难题,机器学习也不可能解决所有的手语理解和生成问题.因此相当长的时间内,众多学者仅限于使用理性主义方法来尝试解决分类词谓语计算的问题.最早的理性主义尝试就是在传统英语词典的基础上,增加了一套像猫、床、地面等分类词谓语的语义特征,把词典中特定动词或介词与其他空间特征相关

联，从而在进行分类词谓语计算时根据其空间信息缩小可能的分类词手形集合，最后生成手部空间运动. 也有其它研究使用了启发式规则[76]，如徐琳等人[77]采用规则解释方法开发了一个中国手语机器翻译系统，由于自定义的一系列规则有限，该系统限制汉语句子为简单陈述句和简单疑问句. 当然考虑到分类词谓语计算的复杂性和多变性，这些规则是无法满足分类词谓语计算的所有需求，并且这些规则随意性太强，因为它需要制定规则的人员自行决定如何部署坐标，如何定义运动路径等. 这些方法最大的缺点就是缺乏部署三维空间元素的能力，从而导致最后只能处理一个分类词谓语的计算，无法实现多个分类词谓语的计算. 随后有学者提出了分类词谓语计算的空间规划模型[78]，他们认为传统音系学模型无法表征分类词谓语，生成的时移坐标流无法精确描述分类词谓语，分类词谓语的计算涉及大量的甚至无限的时移参数. 因此空间规划模型的目标是减少计算所需的时移参数数量，设计一系列算法来简化眼睛、头部、手部位置等参数，并与手势语义相关联以便计算三维位置，借用参数化行为表征(Parameterized Action Representations，PARs)模型来规划场景元素[79]. 这种方法也需要事先规定 PARs 模板，是否适用于所有的分类词谓语还有待验证. 因此在未来的一段时间内，分类词谓语的计算还将以理性主义方法为主，至少我们需要更关注分类词谓语的大脑加工，通过建立小系统来模拟智能行为. 对分类词谓语计算的认识积累到一定程度时，弄清楚分类词谓语的认知机理，我们才能提出分类词谓语计算的整体解决方案. 因为成功的分类词谓语的机器翻译必须实际运用一些空间常识来理解所要传达的空间场景，需要进行复杂的空间常识推理，只有这样才能理解源语言——汉语，并通过空间隐喻将三维分类词谓语的表达式翻译出来. 此外大脑的概念网络涉及左右脑的若干神经结构，这些概念网络与左脑外侧颞叶的词汇网络相联系，包含人物、动物或工具的专门信息. 而这些专门信息是生成分类词谓语手形要用到的，这项研究表明听障手势者大脑特别适宜于处理手语——空间自然语言，这就提示我们分类词谓语的机器翻译需要借鉴脑研究的成果，建立起相应的认知加工模型，以便进行知识表示和空间推理，而不能使用传统语言计算理论来生成相应的分类词谓语，从而克服传统语言计算的缺陷.

3.4　语　义

手语的歧义情况比传统语言更复杂，以隐喻“开花-春天”为例，在汉语里用两个词就可以完成从具体域到抽象域的映射，但在中国手语里，仅用一个手势就可以同时表征源域和目标域，即一个手势可以同时表达“开花-春天”两个词的概念，具体是哪种概念，得结合语境来看. 目前手语的语义消歧研究较少，主要集中在单个手势消歧和模拟听障被试大脑思维的语义计算模型. 如一些文献使用了非手动特征来消歧，1993 年 Butterworth[80] 提出英国手语可使用嘴部动作来消歧. 2008 年 Von 等人[81-82] 总结了非手动特征的作用，指出德国手语可以用头部运动来明确参照物，以区分 NOT 和 TO 两个手势，英国手语则需要唇部动作来区分 NOW 和 Today，以达到手势识别中的消歧目的. Van 等人[83] 则提出了多种消歧策略，包括引入 AAC(Augmentative and Alternative Communication)接口、使用闭集领域词汇和 POS 标注(Part-Of-Speech tagging)来减少树的数量等，以实现英语到南非手语的机器翻译. Yao 等人[84] 则从心理语言学的角度提出了听障者大脑理解空间隐喻时的消歧过程与健听人不同，指出了空间隐喻的语义计算有其明确的认知主题和概念结构. 总之，手语与传统语言的消歧情况不同，以拼音为例，汉语拼音转化为汉字时，存在着很多同音字，导致消歧困难，需要利用语境和世界知识. 而手语也存在着同音现象，但手语可通过非手动特征来帮助消歧，并不需要过多地利用语境和世界知识. 此外受限于手语句法分析理论的落后，语义角色标注仍处于空白.

3.5　语　用

与传统语言一样，手语也需要分析篇章或话语结构，即逻辑语义结构、指代结构、话题结构等. 不同的是这些指代、话题结构都与空间语义有关. 听障手势者在表达话语时，一般会在其自身面前的空间里表示一个实体，用其位置、运动或重新定位这个虚构的对象来表示位置、运动、形状或者所讨论的一些对应的现实世界实体的其他属性. 手语语用分析的难度在于手势空间上非拓扑性的使用可以为手语代名词引用或呼应存储位置，这些位置可以建模为无形世界的特殊对象. 这些代名词引用位置(或“标记”)的布局、管理和操作是一个很复杂的问题[73]. 目前有些文献只做了有限的尝试，2009 年 Lefebvre-Albaret 等人[85] 提出根据手势话语视频来重建身体姿态动作，使用了法国手语的音韵特征来消歧，并用非线性过滤器和 Kalman 平滑算法得到身体姿态动作. 2010 年 Huenerfauth 等人[86] 提出可标注空间参

考点(SRPS)(如果英语文本里有就建立),这样当依次引用表达式和 SRP(Spatial Reference Point),任何动词在空间上屈折变化表示一个 SRP 时,其语篇实体与每个 SRP 相关.这些 SRP 的建立和引用都将被记录在与其他语言学注释对齐的平行时间轴的轨道上.这种理论的前提是语用特征会影响手势者将一个实体指定为 SRP 的可能性,对此作者用统计机器学习技术来构建 SRP 模型,较好地解决了屈折动词主语/宾语的三维位置定位问题.

3.6 应 用

目前手语计算的应用仍然局限在手语机器翻译系统、手语(动画)生成系统以及图像处理领域内的手语识别与合成系统等,其信息检索系统、问答系统仍未见到有关报道.其中手语机器翻译系统与其他传统语言不同的是,手语(动画)生成是手语机器翻译的必要组成部分,故手语生成单独开设一节讨论.

3.6.1 机器翻译

1949 年 Weaver 在其论文里正式提出了机器翻译的思想.5 年后美国乔治敦大学与 IBM 公司合作首次试验了传统语言的机器翻译[87].49 年后,即 1998 年美国 Veale 等人[88]才提出 ZARDOZ 系统(一种英语到 ASL 的机器翻译系统).此后手语机器翻译不断借鉴传统语言的研究成果,其语言加工层次持续加深,经过 20 多年的发展,已实现了句法层次的转换,但本质上没有重大突破.虽然传统语言的机器翻译实现了商品化,谷歌、百度等商用机器翻译得到了广泛应用,出现了翻译质量自动评估方法 BLEU(BiLingual Evaluation Understudy)等机器翻译评测标准,而手语机器翻译还停留在实验室样品展示阶段.

机器翻译系统的架构设计可分为三种:直译、转换和中间语言[89].目前绝大多数系统仍处于转换结构,如中国科学院开发的中国手语翻译系统[76]、TEAM 系统[90]、ASL Workbench[91]、ViSiCAST[92-94]等系统,这些系统主要使用了传统语言计算的句法转移方法而实现,甚至使用了现成的 CMULinker 句法分析器来解析源文本.

采用直译架构的有 TESSA[73]等系统,由于该架构无法将英语翻译成自然手语,仅仅是视觉化的英文字符串;且这类翻译属于直译,在词级进行分析和转换,并未针对手语语法特点进行分析和改进,因此受到众多语言学家的批评,如美国学者 Huenerfauth 等人[74]指出有些机器翻译是把英语翻译成为手势英语而不是自然手语,忽视了英语和 ASL 之间的语言差别,声称实现了手语机器翻译是不严谨的.

使用中间语言架构的系统较少,典型的系统有 ZARDOZ[88].该系统解决了许多类型的语义分歧,系统能够采用某种形式的空间或常识推理.由于输入的英文文本进行语法和语义分析得到的信息是被用来选择特定事件的模式,这些模式将记录所有类型的事件和行为(他们的参加者),因此,开发这种模式所需要的高时间成本限制了该系统只能局限于某一领域.这种系统存在着英语和 ASL 共同的表征结构,而且这些模式一般是与语言无关的,它们可以被认为是一种中间语言.ZARDOZ 是唯一被报道初步解决了分类词谓语机器翻译问题的系统.由于这种中间语言架构已涉及到语义层次的机器翻译,并且开发这个系统的工作量巨大,因此该系统并未完成.

由此可见,手语机器翻译系统未普及的原因在于手语计算的落后,如有些机器翻译涉及依存树到串的翻译模型,以便使用一定的规则,但手语计算如何使用依存树还是个新课题.手语机器翻译的句法层级的分析还停留在初级阶段,仍需要借用传统语言的句级机器翻译模型,对于如何处理调序、时态、语态等方面存在的问题还未深入研究.传统语言的统计机器翻译已成为主流,而手语受限于语料库的规模,手语机器翻译系统基本集中在基于规则的方法、或基于规则和统计混合的专用领域方法.

我们认为对于手语机器翻译的研究,无论是如何制定针对手语特点的翻译模型,还是更加有效地利用现有的模型,手语机器翻译都可以补充和完善目前的机器翻译理论,如手势的内部信息比传统语言要丰富,借助于手语的相似性特点,不需要考虑外部语境,听障者往往能猜出其含义,使用短语规则应能更好地获取局部句法知识,此外听障手势者平时使用的手语句子通常比较短,长难句不多,应用的句法知识应比传统语言更容易、更方便.手语机器翻译和传统语言机器翻译都需要进化到语义层次,建立真正意义上的语义翻译模型,只有这样才有可能从根本上解决手语机器翻译的问题.

3.6.2 手语生成

目前手语生成软件分为两种:脚本软件[95-96]和生成软件[97-98].其中脚本软件被认为是创建手语动画的文本处理工具,该脚本系统提供了语言学和人体运动选项,可以指定手势如何出现,人体躯干如何从手势结束平滑到下个手势开始以及用户未指定的各种各样的其它动画问题.倪训博等人[99]为了保证

手语数据生成的有效性，提出了滑动窗结合模板匹配来实现对中国手势手语的关键手形进行自动标记，这是使用了 Stokoe 模型的一个参数来检测和生成手语数据. 生成软件基于一些信息源来规划手语句子，研发人员研究了如何将书写的传统语言句子自动翻译成手语动画，软件使用的输入信息源就是传统语言的书写文本. 从目前来看，手语机器翻译系统使用脚本软件的居多，如前面介绍的 TEAM 系统和 ZARDOZ 系统，也有其它系统采用了生成系统，如 ViSiCAST 系统. 这里以这两类为例举例说明.

TEAM 和 ZARDOZ 之类的脚本系统通常需要自己开发动画虚拟人模型，并自己定义专用的脚本控制语言，这些模型的差异在于设计的粒度，如前者可灵活规定一个附带少量输入参数的动作或运动路径，但是系统的脚本质量还是无法保证非手动特征和音韵平滑的真实性. 而后者针对不同的原子（指最小不可分割的单位）非手动特征做了特殊设计，可以识别所有非手动特征的原子操作，但对于复杂的非手动特征，如重叠的、交互的非手动表达形式，甚至非手动特征的强度随时间而变化，都无法表达出来.

手语生成软件则更为简单，因为它可使用现成标准，如国际标准 SGML（Standard Generalized Markup Language），像 ViSiCAST 就使用了现有 HamNoSys 手语书写系统的 XML（Extensible Markup Language）版本[100] 和 SGML. 也有学者自己定义了标准规范，比如国内 Ye 等人[101] 基于 XML 语法结构为中国手语定义了表达内容格式化描述方法，并应用于中国手语合成系统[102-103]. 这些标准都指定了手形、手掌方向和运动细节. 当然为了生成手语动画，还需要设计与每个词条相关的更多的信息，如 ViSiCAST 使用的标准就包括语音 SGML 规范、特殊次范畴、句法和形态特征等. 因此相比动画脚本系统，生成软件系统生成的动画更为直观.

这些手语生成系统都是对手语翻译的有益探索，但这两者各有利弊，还有很多问题需要改进. 脚本软件目前仍是各自为战，自己开发动画人物模型，没有形成公认的统一标准，此外动画虚拟人物表达粒度不灵活，无法延伸和扩展，使很多语法信息无法表达. 更重要的是这些手语生成软件都未针对空间位置进行扩展和优化，不便于表达手语对话涉及的实体，因此这些系统的输出动画没有反映手势的空间使用和手语词汇的屈折变化[104]，还停留在单信道的计算. 例如 Sign Smith Studio 软件（针对 ASL 定制的商业化脚本软件）的词典只包含了大多数 ASL 动词的未进行屈折变化（屈折变化是为限定某词的语法功能而添加词缀或改变词形所进行的词的屈折变化）的版本①. 如果要生成动词的屈折形式，用户必须使用附带的软件精确地生成手部动作，来生成动词手势，这显然大大延长了生成 ASL 脚本动画的过程.

此外这些系统也没有很好地处理面部表情等非手动特征，主要是受限于句子和其他物理空间限制的词法和语法选择，这种非手动特征决策较为复杂. 对此国内学者做了初步的尝试，如陈益强等人[105] 提出了协同韵律参数控制方法，实现语音、唇动、表情、头部等信道的协调一致，从而实现动画人物的多模式行为合成协同. 此后他们对真人手语表演数据中的手势与头部动作之间的关系进行了深入研究，利用核典型相关分析方法（Kernel Canonical Correlation Analysis，KCCA）建立起手势与头部动作之间的预测关系模型，大大提高了虚拟人行为动作合成的逼真性[106]. 当然由于多信道转换的复杂性，在这方面还有很多工作要做[40].

4　互联网时代手语计算何去何从

4.1　互联网时代给手语计算带来新的机遇

只要有海量语料，就可以借助计算机强大的计算能力，再制定适应的数学模型，挖掘出背后的知识. 而手语计算则无法享受到这点便利，但是互联网时代涌现出的新技术则为手语计算带来了新的机遇.

4.1.1　体感设备的出现

目前建设手语语料库的瓶颈在于手语视频的采集和标注非常繁琐费时，以往为了采集手语信息，采集传感器利用可穿戴式设备，或者利用普通摄像头采集手语视频语料或图像信息. 前者唯一的缺点是设备昂贵复杂，采集过程中需要被试一直穿戴，人机交互性很差. 后者虽具备自然的人机交互性，但这种方法准确率低、速度慢，现有的识别算法均无法获取高精度的识别率[29]. 因此这两种采集过程繁琐，未实现大范围的推广应用. 而体感设备则为建设大规模手语视频语料库提供了可能. 首先体感设备获取的手势信息通常是目标的深度和红外图像信息，它借鉴了人眼原理将传统二维物体转换到三维空间.

① VCom3D, Homepage. http://www.vcom3d.com/, 2016.4.25

最关键的是其设备成本低、使用简单,同时能满足手语语料自动标注的高精度和实时性的需要.以 Leap Motion 为例,Leap Motion 可以 290 帧/s 的速度识别人体手部的 22 个关键坐标①,见图 7.凭借这些信息可以为语料库构建手势者三维模型.同时可以计算出手部的位置、手形、方向和运动四个音韵特征并进行标注,当然如果需要标注形态、句法、篇章等层面的信息,则需要在音韵层面上针对手语的空间特征提出新的句法分析理论和算法.最关键的是体感设备可以采集静态与动态的特征,为探索手语运动的数学模型及其计算理论,建立手语行为与语言特征的统计方法及其两者之间的关系提供了可能.

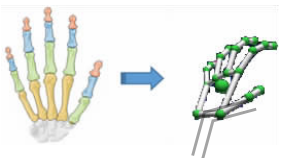


图 7 手部 22 个关节点示意图

4.1.2 认知神经科学的兴起

目前语言计算技术得以广泛应用,主要依赖于深度学习和大数据,将这些人工智能的成功经验与认知神经科学的研究思路结合.目前语言计算仅仅关注于用计算机处理文本,对背后的语言心理过程关注不够,如果能够搞清楚人脑在概念组织、意义推理等能力上的内在认知机制,并在语言计算时结合之,则会对语言计算有更大的帮助.Krahmer[107] 从心理学的角度对语言特征及其说话者做了分析,可以在面向应用的计算模型中发挥积极作用.认知神经科学的研究有助于在人工神经网络技术中设计实现新的学习机理与拓扑结构,将强有力地推动语言计算的进展.例如一种新型的具有长短时记忆能力的递归神经网络(Long Short-Term Memory-Recurrent Neural Network, LSTM-RNN)已经成为当前语言计算的一个标准配置,该模型相比传统的深度神经网络方法可以将错误识别率再降低 20%～30%.它和人脑神经网络在时间上进行信息积累相类似,通过网络拓扑结构的优化和改变可以拥有对序列性数据更好的处理能力.最近几年随着 fMRI(functional Magnetic Resonance Imaging)、ERP(Event-Related Potential)等无创伤脑功能成像技术和脑电技术的相继出现,我们可以近距离观察人类大脑内部的加工过程,从而探索其语言的神经运作机制.这些认知神经科学的手段为手语大脑加工的研究提供了许多科学的证据.例如,国外学者比较了在观看英语句子时,母语为英语的健听人、母语为手语的听障者、母语为手语的健听人这三类被试的大脑反应,结果发现听障者与健听人一样都激活了经典语言区域——左脑外侧裂周区,但听障者大脑还激活了右脑区域,研究者分析可能与手语空间加工因素相关[108].对于分类词谓语的认知研究还发现,除了与空间关系相关的场景加工,分类词谓语还涉及到复杂的空间隐喻加工[9,75],从而为分类词谓语的计算指出了研究方向.因此为了实现认知神经科学与人工智能的有效结合,深入了解大脑的信息处理机制,用计算机人工智能模拟人脑信息处理的方法,为手语计算提供一定的理论支持.由此表明手语计算毕竟是认知科学、语言学和计算机科学等多学科交叉的复杂问题,我们需要从外层(或表层)研究手语理解的理论方法和数学模型,也需要从内层解释人脑理解手语机制的秘密,从人类认知机理和智能的本质上为手语计算寻求依据.认知神经科学为人工智能的发展提供了一条可能的途径,将认知神经科学的相关知识应用于手语计算将会是未来一段时间内非常值得关注的研究方向.

4.1.3 深度学习的进展

深度学习是目前人工智能领域的研究热点,也是互联网时代的重大科学突破.众多文献表明深度学习的本质是通过构建神经网络,更深层次地模拟人脑活动,深层非线性网络结构决定了其具备从少数样本中学习数据集的本质特征的能力.它可以找到数据在时间与空间上的内在联系,进而提高分类的准确性.然而深度学习是最先在语音和图像处理领域取得突破性进展,其后才应用到自然语言处理领域,相比于语音和图像,文本语言是唯一的完全由人脑生成和加工的非自然符号系统,近期的研究报道表明深度神经网络似乎在处理自然语言的优势上并不明显,因此如何更加适应于传统语言的计算还需要更多的研究和探索.在语音和图像的处理过程中,输入信号可以在向量空间内表示,而传统语言通常需要将独立的词汇转换为向量,才能作为神经网络的输入.手语则是融合了图像和人工符号系统的语言,理应比传统语言更适合在向量空间内表示.此外自然语言处理与语音和图像处理的区别还在于传统语言本身就是时间序列,这意味着需要重点解决各种复杂递归结构,尤其在处理句法分析等更加复杂树形结构时,神经网络应能处理这种结构化问题.

① Leap Motion, Homepage. http://www.leapmotion.com/, 2017.4.25

为了解决语言的各种递归结构的问题，且能够处理诸如句法分析对应的复杂树结构，已涌现了 LSTM 之类的特殊神经网络，未来还会有新的深度神经网络来适应语言计算的任务，以实现自然语言的序列输出. 手语的一些空间特征如果利用深度网络做区分性非线性变换，得到的输出作为新的特征向量可显著提高识别率，有助于改善系统的性能. 因为图像经过非线性变换后，已消除了原始特征与低层次描述无关的噪声影响，从而使描述图像特征的准确度和原始特征的相关度大大增加. 因此随着深度学习的研究进展，有可能发现具有潜在复杂结构规则的手语视频等丰富结构数据的本质特征，从而为实现手语计算创造条件. 这些传统语言深度学习的研究将有助于手语计算的进展.

手语计算的本质是多信道的计算，Ngiam 等人[109]将深度学习应用于多模态上学习特征，在 CUAVE 和 AVLetters 数据集上的实验结果表明模态之间可共享表示，使用多个模态可得到更好的特征. 高文等人[110]则从异质模式交互的角度提出了基于多模式接口的交互模型，将手语识别、唇读、人脸特征检测以及特定面部动画相结合，构成了手语转换和口语交流的代理，以便有效实现异种语言模式间交流. 此外，手语主要使用非手动特征表达语用信息，比如使用面部表情来表达情感，相比给定一段文本来判断其情感类别及强度更易于表征，这是一个全新的课题，因此将深度学习应用于手语计算需要进行更多的探索和研究，处理手语的结构化输出需要更为复杂的神经网络，对高效和并行化的训练算法提出了新的要求，这给统计学习意义下的神经网络模型的结构设计、参数选取、训练算法以及时效性等方面都提出了新的挑战. 从生物学的角度看，除了语言能力，包括人类在内的大部分动物几乎都具有良好的视听觉能力. 因此对于模仿人脑结构的人工神经网络，处理语言可能是比加工视听觉信息更为困难，而手语可以实现把语言特征与人类自身相分离而单独进行研究，我们相信深度学习在手语计算方面有很大的探索空间，如何针对手语计算设计有效的深度神经网络模型与学习理论，从融合视觉信息和语言信息的数据中获取真实的规律信息，成功地解决这个难题是实现人工智能不可缺少的关键环节之一.

4.2　从传统语言历史看手语计算

传统语言计算的萌芽期处于 20 世纪 40 年代[111]，经过 70 多年的发展历程，统计模型和数据驱动已成为语言计算的主流方法，各种语言计算的任务都已开始引入概率，并借用了语音处理等任务的评测方法. 机器学习和资源建设成为当前语言计算的主流.

对照传统语言计算的历史，目前的手语计算还处于萌芽期. 很多理论问题未得到根本性的解决，尽管许多理论模型在手语计算中发挥着重要的作用，如音系学分类、动词划分等，但很多基础性课题并未得到彻底、圆满的解决，如空间建模、空间隐喻的表征、分类词谓语的认知机理等. 手语计算尚未建立起一套完整系统的理论框架体系. 虽然一些手语计算理论不断与新的相关技术相结合，比如动画技术、数据手套、可穿戴设备等，用于研究和开发手势者三维模型、手语语料库等，但更多的手语计算仍处于初始阶段，如盲目套用传统语言计算的机器学习方法，或主观地更换新的数据收集设备. 这些盲目的尝试只能是对一些边角问题的修修补补，或者仅能解决特定条件下的一些具体问题，他们并不能从根本上建立全局性的鲁棒性的解决方案. 当然有些研究是受限于目前的技术条件而无法取得突破，如人类对手语的大脑认知加工机理仍未搞清楚，相关的文景转换、隐喻理解等难题在传统语言计算里仍未取得进展，但空间参数是手语理解的最重要因素，建立有效的空间模型与实现空间参数的快速计算是我们需要研究的、大有可为的方向.

4.3　从技术趋势看手语计算

根据 Gartner 2015 年报告分析，见图 8，以自然手语为主要沟通模式的研究在自然语言问答、手势控制、语音识别等技术走出技术萌芽期、泡沫化谷底

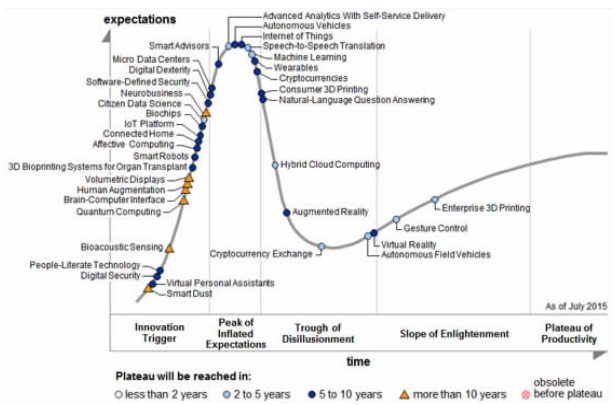


图 8　2015 年新兴技术成熟度曲线(数据来源于 Gartner①)

① Gartner's 2015 Hype Cycle for Emerging Technologies Identifies the Computing Innovations That Organizations Should Monitor. http://www.gartner.com/newsroom/id/3114217, 2016.5.1

期，乃至到达实质生产的高峰期的趋势下有呼之欲出之势. 这些较成熟的技术可以为手语计算提供坚实的理论保障和技术支持.

Yao 等人[112]从认知计算的角度，分析了大脑感知和手语理解的机理，提出了手语认知架构，如图 9 所示. 由此可见，手语的认知计算是从手势的物理特征到语义表征的映射转换过程. 具体来讲，从像素、边等底层特征逐层加工映射成音韵特征，再根据音韵特征加工成低级别的语义单元、再逐步抽象出高级别的语义单元之类的高层特征，最终形成手势语义概念. 由此可见，过去 30 多年的手语识别与计算省略了音韵特征、语义单元这样的中间步骤，直接从底层特征得到语义概念，这样的分析是不太合适的. 而体感设备的出现，填补了这样的空缺，即可以直接从手语的底层物理特征推断出语言学特征——音韵特征. 直接从音韵特征这种语言学特征推断出语义概念，至少要比直接从图形学特征推断出语义概念

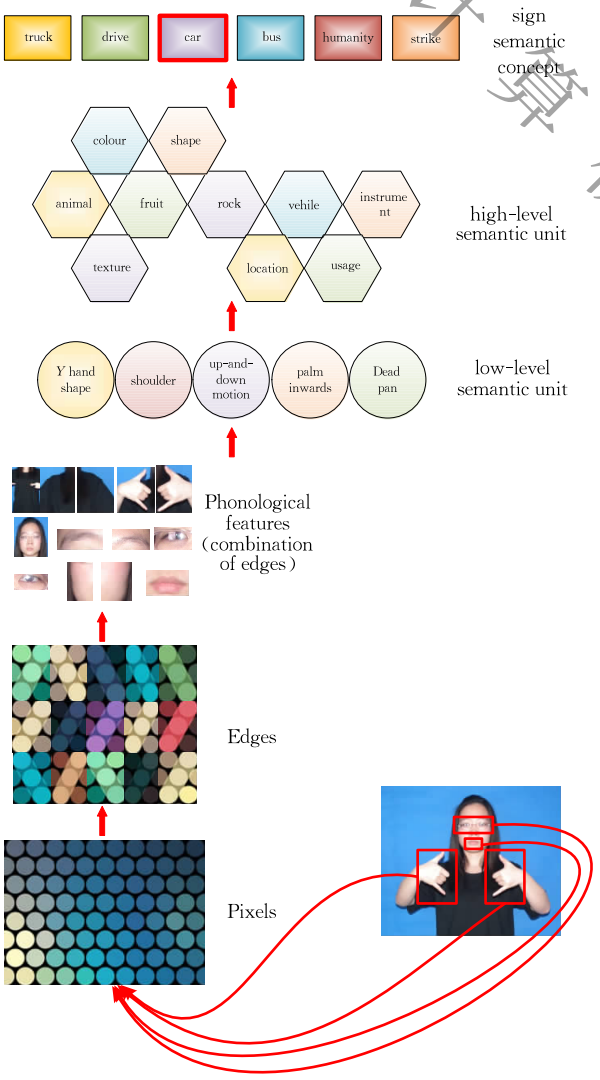


图 9 手语认知计算架构

要前进了一大步. 而从音韵特征逐层加工成低、高层的语义单元，直至语义概念，即在输入数据时，每处理一层，提取的概念特征就抽象一级. 深度神经网络就可以解决这个问题. 当然，再从语义概念按照词法、句法和语用特征等组合成更高级别的语义特征，需借助于认知神经科学的手段. Friederici 等人[113]利用 fMRI 实验发现，人脑存在三个独立的区域涉及到语用知识，即包括世界知识等在内的语用信息与语言形式是各自完全独立的. 人脑如何将语用信息与语义概念等语言形式结合成更高级别的句子或篇章都有待于认知神经科学的揭晓与相关的实验证明.

因此我们有理由认为手语计算正在从萌芽期向发展期过渡. 过渡期内，理性主义和经验主义两种方向并行发展，因为手语计算毕竟是一个庞大的体系，有太多的规则需要归纳和整理，如人工总结一系列规则，构造相应的推理程序，把手语的语法结构映射到语义符号. 像句法理论就需要针对空间特征进行扩展，提出适合于手语的依存语法等. 同时传统语言的经验主义方法则会给手语计算带来启迪，但重点需要在手语行为与语言特征的关系上进行机器学习，建立融合空间特征的统计学习模型.

5 未来的研究方向

传统语言计算的黄金期是在发展期，这一时期出现了一系列重要的理论研究成果，使传统语言计算理论得到了长足的发展并逐渐成熟. 因此互联网时代虽然给手语计算带来了机遇，但手语计算若要发展为与传统语言并驾齐驱的学科，则可能需要在以下几个研究方向上取得突破，方能带动整个手语计算乃至语言计算体系的发展.

5.1 资源建设

手语计算不仅需要语料库资源，还要重视语料库的建设方法以及大数据与挖掘结果的关系. 体感设备的出现，使大规模手语视频标注技术成为可能，并通过这一技术引领手语计算学科的发展. 但体感设备仅仅是使大规模手语语料库成为可能，配套的语言知识资源也要跟上，包括句法资源、语义资源. 因为手语计算最终要过渡到语义层次，比如分类词谓语计算就需要进行深层分析，这些都需要语义资源的支持.

以深度学习为例，手语的音韵特征包含了丰富的内部语义信息. 众多研究表明，手语的音韵参数之一——手部位置是唯一基于语义编码的参数. 因

此手势的内部结构信息可以提供有价值的语义知识，从而填平在学习手势表征时未登录手势与登录手势之间的鸿沟. 以中国手语为例，在编写手语语义词典时，把意识、思维、认识、概念、想念和思想等手势与太阳穴联系起来，以太阳穴为空间位置给出手语语义编码，把这些手势表达思想活动的语义归为一类，在手势嵌入学习时，除了可以从外部语境学习，也可以从手语语义词典进行学习，从手势的音韵特征，如位置、方向、手形和运动等音韵参数的意义来推断手势语义.

5.2　文景转换

手语生成和文景转换一样，都一致性涉及到物体空间关系建模研究，不同的是手语生成仅涉及到空间布局，不需去排列与手势者表达的内容相一致的图片序列，或建模摆放相关的概念实体. 文景转换在自然语言处理领域是一个较新的研究课题，目前相关研究并不充分. 应用到手语计算时，同样也面临着语言的模糊性问题. 传统语言文景转换的难点就在于空间概念的模糊表达影响了场景元素无法平滑的部署. 目前国内外已开发出了文景转换系统的原型，例如中国科学院开发的天鹅系统[114]，美国 AT&T 实验室 WordsEye 系统[115]等，但是这些系统只能实现文本到静态场景的自动转换. 对于手语计算，可能需要自然语言文本到简单动画生成系统或者交互语言动画创建系统这两类系统的支持. 其中前者还处于初级阶段，还没有较完善的系统，后者则集中在让用户通过语言来控制动画角色的动作交互行为，代表性系统有 Alice[116]、AnimNL[117]等，其中 AnimNL 已被应用于 ASL 机器翻译系统. 此外与手语相关的文景转换还涉及到动画脚本的问题，因为这种文景转换需要进行交互，让手势者来控制场景中的布局和动作交互，核心是处理输入的语言，然后绑定到场景中的某个点，手势者可以通过交互进行有目的的修改和设置.

因此这种文景转换充分证明了空间概念是人类认知的基本概念，也是包括手语在内的所有自然语言描述的最基本概念，是理解其他各类关系的基础. 目前传统语言计算对空间计算的局限在于空间知识的管理与分析是一个十分专业的过程，需要专家对认知理论和计算机图形学的一些术语有所掌握，致使很难自动建立这样的空间知识库. 研究和探索基于手语描述的文景转换，有望在空间关系理解和虚拟场景生成取得突破. 一方面，从手语多信道表征的角度考虑空间、空间中的位置、空间中的运动等内容，另一方面，借鉴手语计算的角度来分析传统语言中空间语言的语义等内容. 这两个方向有助于实现空间信息抽取，即物体的空间方向、位置等信息，结合知识库消除自然语言的模糊性，进而实现三维场景构建.

5.3　隐喻处理

隐喻普遍存在于人类语言活动中，是语言计算不可回避的问题. 目前传统语言里的隐喻计算模型的应用比较单一，如国内目前主要的研究集中在名词性隐喻的自动处理，而且多数方法均基于统计理论[118-119]. 但隐喻更是一种认知现象，从手语的空间隐喻例子我们可以看到，对于相同的隐喻，不同的认知主体的理解结果可能并不相同. 因此如果仅仅基于统计方法模型去处理隐喻理解问题是不够的，有必要将认知主体的主观知识和认知角度纳入考虑范畴. 因此我们需要从认知科学的角度考虑隐喻加工，将认知科学方法与隐喻计算模型进行结合，使之处理隐喻问题更加符合认知观点.

传统语言计算对隐喻的处理主要集中在隐喻识别和理解，而对隐喻生成关注甚少. 研究表明听障手势者进行分类词谓语计算时用到了空间隐喻和场景规划，在从汉语语义生成手语的分类词谓语时，从单信道表征向多信道表征转换时自动实现了隐喻生成. 因此若以手语空间隐喻为代表的隐喻生成取得突破，则将会极大地推动语言计算. 空间隐喻是手语的一个高度规律性和极富效率的特征，相对汉语隐喻，听障者只需要更短的时间就可以理解手语隐喻，需要借助 ERP(Event-Related Potential)、fMRI (functional Magnetic Resonance Imaging)等脑神经成像或脑电技术来探究，如汉语隐喻和手语隐喻的理解过程能否从脑电波谱图上区别开来.

我们认为手语隐喻计算研究对于语言计算技术发展有着十分重要的意义，有助于发展一个广泛应用的、不拘泥于隐喻类型的计算模型.

6　总结与展望

综上所述，语料匮乏和理论缺乏仍是制约手语计算的研究、发展和实用化开发的瓶颈问题，实现手语计算可能更多的依赖于多个学科的交叉研究，其涉及的学科方向包括人工智能、计算机图形学、认知神经科学、计算语言学和可视化研究等，其中最突出的是手语生成涉及到的虚拟人物模型需要自然语言处理和计算机图形学的知识. 因此我们需要从其他

相关的学科吸取营养来丰富自己的知识,以适应学科交叉性和边缘性的要求,另辟蹊径解决手语计算的难题.对此,我们对手语计算进行了回顾与展望,总结了 30 年来手语计算的理论成果,分析了手语计算所带来的影响和未来的研究方向,探讨了手语计算的发展趋势.

手语计算作为自然语言处理的新领域,也是互联网时代众多信息的重要载体.《国家中长期科学和技术发展纲要》已将语言计算列为前沿技术,体现了国家的重大科技需求.同时传统语言里已实现实用的信息检索、问答系统等,手语计算还未进行深入研究,如手语版"Siri"尚未出现,甚至"分析互联网"、中英文"知识图谱"等最新进展在手语领域也未开始.手语本身的特点导致了手语计算面临着一系列的困难,不管是理论构建,还是商业应用,这些任务可能需要长期艰苦的努力.但这种情况也提示着我们,手语计算很可能正处于改革的前夜.在人类的历史长河中,手语拥有比传统语言更长的历史,大数据和互联网时代涌现的机遇已经给这种有着悠久历史的古老语言注入了新的生命力,在交叉学科的推动下,手语计算可能会出现实质性突破.传统语言计算的惊人进展已推动人工智能和人机交互等学科大踏步的前进,如果我们能在理论、技术和工程方面,突破手语计算的一系列难题,则可大大加速推进自然语言处理和人工智能的向前发展.

参 考 文 献

[1] Futrell R, Mahowald K, Gibson E. Large-scale evidence of dependency length minimization in 37 languages. Proceedings of the National Academy of Sciences, 2015, 112(33): 10336-10341

[2] Mehrabian A. Linguistics: Silent messages. American Anthropologist, 1973, 75(6): 1926-1927

[3] Aronoff M, Rees-Miller J Eds. The Handbook of Linguistics (Vol. 22). New Jersey, USA: John Wiley & Sons, 2008

[4] Sun Mao-Song. Language computing: A commanding point of strategy of medium and long-term development of information science. Journal of Language Application, 2005, 14(3): 38-40(in Chinese)
(孙茂松. 语言计算:信息科学技术中长期发展的战略制高点. 语言文字应用, 2005, 14(3): 38-40)

[5] Yu Shi-Wen, Zhu Xue-Feng, Geng Li-Bo. Natural language processing technology and language deep computing. Chinese Social Sciences, 2015, 36(3): 127-135(in Chinese)
(俞士汶, 朱学锋, 耿立波. 自然语言处理技术与语言深度计算. 中国社会科学, 2015, 36(3): 127-135)

[6] Huenerfauth M. American Sign language generation: Multimodal NLG with multiple linguistic channels//Proceedings of the Association for Computational Linguistics (ACL) Student Research Workshop. Michigan, USA, 2005: 37-42

[7] Emmorey K, Corina D. Lexical recognition in sign language: Effects of phonetic structure and morphology. Perceptual and Motor Skills, 1990, 71(3f): 1227-1252

[8] Montemurro M A, Zanette D H. Entropic analysis of the role of words in literary texts. Advances in Complex Systems, 2002, 5(1): 7-17

[9] Huenerfauth M. Spatial representation of classifier predicates for machine translation into American Sign Language//Proceedings of the Workshop on Representation and Processing of Sign Language, 4th International Conference on Language Resources and Evaluation. Lisbon, Portugal, 2004: 24-31

[10] Liddell S K. 16 Blended spaces and deixis in sign language discourse//McNeill D ed. Language and Gesture. Cambridge, UK: Cambridge University Press, 2000, 2: 331-357

[11] Sutton-Spence R, Woll B. The Linguistics of British Sign Language: An Introduction. Cambridge, UK: Cambridge University Press, 1999

[12] Clayton V, Lucas C. Linguistics of American Sign Language: An Introduction. Washington, District of Cloumbia, USA: Gallaudet University Press, 2000

[13] Liu H-T. Deaf Students' Story Comprehension Using Manually Coded Chinese, Taiwanese Sign Language and Written Chinese [Ph. D. dissertation]. National Changhua University of Education, Changhua, China, 2004(in Chinese)
(劉秀丹. 啟聰學校學生文法手語、自然手語及書面語故事理解能力之研究[博士学位論文]. 彰化師範大學特殊教育研究所, 彰化, 中国, 2004)

[14] Karen E, Corina D, Bellugi U. Differential processing of topographic and referential functions of space//Emmorey K, Reilly J eds. Language, Gesture, and Space, Lawrence Erlbaum Associates: Hillsdale, USA, 1995: 43-62

[15] Hickok G, Say K A, Bellugi U, Klima E S. The basis of hemispheric asymmetries for language and spatial cognition: Clues from focal brain damage in two deaf native signers. Aphasiology, 1996, 10(6): 577-591

[16] Diane L-M. Where are all the modality effects? //Meier R, Cormier K, Quinto-Pozos D eds. Modality and Structure in Signed and Spoken Languages. Cambridge, UK: Cambridge University Press, 2002: 241-262

[17] Stokoe W C. Sign language structure: An outline of the visual communication systems of the American deaf. Studies in Linguistics: Occasional Papers, 1960, 8: 3-37

[18] Vogler C, Metaxas D. Parallel hidden markov models for American Sign Language recognition//Proceedings of the IEEE International Conference on Computer Vision. Corfu, Greece, 1999, 1: 116-122

[19] Jiang Feng, Gao Wen, Yao Hong-Xun, Chen Xi-Lin. LMA approach in person-independent sign language recognition. Chinese Journal of Computers, 2007, 30(5): 5851-5860(in Chinese)

(姜峰，高文，姚鸿勋，陈熙霖. 手势手语力效分析. 计算机学报，2007，30(5)：5851-5860)

[20] Lichtenauer J，Hendriks E，Reinders M. Learning to recognize a sign from a single example//Proceedings of the 2008 8th IEEE International Conference on Automatic Face & Gesture Recognition (FG'08). Amsterdam，Netherlands，2008：1-6

[21] Bowden R，Windridge D，Kadir T，et al. A linguistic feature vector for the visual interpretation of sign language//Proceedings of the 8th European Conference on Computer Vision. Prague，Czech Republic，2004：390-401

[22] Kadir T，Bowden R，Ong E J，et al. Minimal training，large lexicon，unconstrained sign language recognition//Proceedings of the the British Machine Vision Conference. London，UK，2004：1-10

[23] Cooper H，Bowden R. Large lexicon detection of sign language//Human-Computer Interaction. Berlin，Germany：Springer，2007：88-97

[24] Sandler W. Phonological representation of the sign：Linearity and nonlinearity in American Sign Language. Walter de Gruyter，1989

[25] Baus C，Gutiérrez-Sigut E，Quer J，Carreiras M. Lexical access in Catalan Signed Language(LSC) production. Cognition，2008，108(3)：856-865

[26] Ong S，Ranganath S. Automatic sign language analysis：A survey and the future beyond lexical meaning. IEEE Transactions on PAMI，2005，27(6)：873-891

[27] Wu Y，Huang T. Vision-based gesture recognition：A review，gesture-based communication in human-computer interaction. LNCS，1999，1739：103-115

[28] Mitra S，Acharya T. Gesture recognition：A survey. IEEE Transactions on SMC-C，2007，37(3)：311-324

[29] Yao Deng-Feng，Jiang Ming-Hu，Abudoukelimu A，et al. A survey of Chinese sign language processing. Journal of Chinese Information Processing，2015，29(5)：216-228(in Chinese)
(姚登峰，江铭虎，阿布都克力木·阿布力孜等. 中国手语信息处理述评. 中文信息学报，2015，29(5)：216-228)

[30] Stokoe W C，Casterline D C，Croneberg C G. A Dictionary of American Sign Language on Linguistic Principles. Silver Spring，USA：Linstok，1965

[31] Prillwitz S，Hamburg Zentrumfür Deutsche Gebärdensprache und Kommunikation Gehörloser. HamNoSys：Version 2.0；Hamburg Notation System for Sign Languages；an Introductory Guide. Willkommen，Germany：Signum-Verlag，1989

[32] Liddell S K，Johnson R E. American Sign Language compound formation processes，lexicalization，and phonological remnants. Natural Language & Linguistic Theory，1986，4(4)：445-513

[33] Vogler C，Metaxas D. Toward scalability in ASL recognition：Breaking down signs into phonemes//Proceedings of the Gesture-Based Communication in Human-Computer Interaction. International Gesture Workshop，Gif-sur-Yvette，France，1999：211-224

[34] Awad G，Han J，Sutherland A. Novel boosting framework for subunit-based sign language recognition//Proceedings of the 2009 the 16th IEEE International Conference on Image Processing. Cairo，Egypt，2009：2729-2732

[35] Brentari D. A Prosodic Model of Sign Language Phonology. Cambridge，Massachusetts，USA：MIT Press，1998

[36] Sandler W. The Syllable in Sign Language：Considering the Other Natural Language Modality. The Syllable in Speech Production. New York，USA：Lawrence Erlbaum Associates，2008：379-408

[37] Wilbur R B. Why syllables? What the notion means for ASL research//Fischer S D，Siple P eds. Theoretical Issues in Sign Language Research. Chicago，USA：University of Chicago Press，1990，1：81-108

[38] Kong W W，Ranganath S. Automatic hand trajectory segmentation and phoneme transcription for sign language//Proceedings of the 2008 8th IEEE International Conference on Automatic Face&Gesture Recognition(FG'08). Amsterdam，Netherlands，2008：1-6

[39] Han J，Awad G，Sutherland A. Modelling and segmenting subunits for sign language recognition based on hand motion analysis. Pattern Recognition Letters，2009，30(6)：623-633

[40] Supalla T. The classifier system in American Sign Language//Craig C ed. Noun Classes and Categorization. Amsterdam，Netherlands：John Benjamins，1986：181-214

[41] Newport E L. Task specificity in language learning? Evidence from speech perception and ASL//Wanner E，Gleitman L R eds. Language Acquisition：The State of the Art. Cambridge，UK：Cambridge University Press，1982：450-486

[42] Newport E L，Bellugi U. Linguistic expression of category levels in a visual-gestural language：A flower is a flower is a flower//Rosch E，Lloyd B B eds. Cognition and Categorization. Hillsdale，USA：Lawrence Erlbaum Associates，1978：49-77

[43] McDonald B. Levels of analysis in sign language research//Kyle J G，Woll B eds. Language in Sign：An International Perspective on Sign Language. London，UK：Croom Helm，1983：32-40

[44] Fang G，Gao W，Zhao D. Large-vocabulary continuous sign language recognition based on transition-movement models. IEEE Transactions on Systems，Man and Cybernetics，Part A：Systems and Humans，2007，37(1)：1-9

[45] Wang C，Gao W，Shan S. An approach based on phonemes to large vocabulary Chinese Sign Language recognition//Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition. Washington，USA，2002：411-416

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

[47] Dreuw P, Ney H. Towards automatic sign language annotation for the ELAN tool//Proceedings of the International Conference on Language Resources and Evaluation (LREC Workshop): Representation and Processing of Sign Languages, European Language Resources Association. Marrakech, Morocco, 2008: 1-10

[48] Efthimiou E, Fotinea S E. GSLC: Creation and annotation of a Greek Sign Language corpus for HCI//Proceedings of the 4th International Conference on Universal Access in Human-Computer Interaction. Beijing, China, 2007: 657-666

[49] Zahedi M, Dreuw P, Rybach D, et al. Continuous sign language recognition-approaches from speech recognition and available data resources//Proceedings of the 2nd Workshop on the Representation and Processing of Sign Languages: Lexicographic Matters and Didactic Scenarios. Paris, France, 2006: 21-24

[50] Braffort A, Choisier A, Collet C, et al. Toward an annotation software for video of sign language, including image processing tools and signing space modelling//Proceedings of the fourth International Conference on Language Resources and Evaluation (LREC). Lisbon, Portugal, 2004: 201-204

[51] Crasborn O, Mesch J, Waters D, et al. Sharing sign language data online: Experiences from the ECHO project. International Journal of Corpus Linguistics, 2007, 12(4): 535-562

[52] Gonzalez M, Collet C, Dubot R. Head tracking and hand segmentation during hand over face occlusion in sign language//Proceedings of the 11th European Conference on Computer Vision (ECCV 2010 Workshops). Heraklion, Greece, 2010, 6553: 234-243

[53] Yao D, Jiang M, Huang Y, et al. Study of sign segmentation in the text of Chinese Sign Language. Universal Access in the Information Society, 2017, 16(3): 725-737

[54] Meier R P. Person deixis in American Sign Language. Theoretical Issues in Sign Language Research, 1990, 1: 175-190

[55] Cormier K A. Grammaticization of Indexic Signs: How American Sign Language Expresses Numerosity [Ph. D. dissertation]. University of Texas Austin, Austin, USA, 2002

[56] Toro J. Automated 3D animation system to inflect agreement verbs//Proceedings of the 6th Annual High Desert Linguistics Society Conference. Milton Keynes, UK, 2004

[57] Toro J A. Automatic Verb Agreement in Computer Synthesized Depictions of American Sign Language [Ph. D. dissertation]. DePaul University, Chicago, USA, 2005

[58] Segouat J, Braffort A. Toward the study of sign language coarticulation: Methodology proposal//Proceedings of the 2009 2nd International Conferences on Advances in Computer-Human Interactions. Cancun, Mexico, 2009: 369-374

[59] Huenerfauth M, Lu P. Modeling and synthesizing spatially inflected verbs for American Sign Language animations//Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility. Orlando, USA, 2010: 99-106

[60] Duarte K, Gibet S. Presentation of the SignCom project//Proceedings of the First International Workshop on Sign Language Translation and Avatar Technology. Berlin, Germany, 2011: 10-11

[61] Lu P, Huenerfauth M. Learning a vector-based model of American Sign Language inflecting verbs from motion-capture data//Proceedings of the 3rd Workshop on Speech and Language Processing for Assistive Technologies. Association for Computational Linguistics. Montreal, Canada, 2012: 66-74

[62] Kegl J, MacLaughlin D, Bahan B, et al. The Syntax of American Sign Language: Functional Categories and Hierarchical Structure. Cambridge, USA: MIT Press, 2000

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

[64] Huenerfauth M. Representing coordination and non-coordination in American Sign Language animations. Behaviour & Information Technology, 2006, 25(4): 285-295

[65] Martell C H. FORM: An extensible, kinematically-based gesture annotation scheme//Proceedings of the 7th International Conference on Spoken Language Processing. Denver, USA, 2002: 353-356

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

[67] Cox S, Lincoln M, Tryggvason J, et al. Tessa, a system to aid communication with deaf people//Proceedings of the 5th International ACM Conference on Assistive Technologies. Edinburgh, Scotland, 2002: 205-212

[68] Loeding B L, Sarkar S, Parashar A, et al. Progress in automated computer recognition of sign language//Proceedings of the 9th International Conference on Computers Helping People with Special Needs (ICCHP 2004). Paris, France, 2004, 3118: 1079-1087

[69] Lu P, Huenerfauth M. Collecting a motion-capture corpus of American Sign Language for data-driven generation research//Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies. Association for Computational Linguistics. Los Angeles, USA, 2010: 89-97

[70] Yao Deng-Feng, Jiang Ming-Hu, Chang Jung-Hsing, Abudoukelimu A. Cognitive-semantic analysis of classifier predicates in Chinese Sign Language//Proceedings of the 18th Chinese Lexical Semantics Workshop (CLSW2017). Leshan, China, 2017: 1-10

[71] MacFarlane J, Morford J P. Frequency characteristics of American Sign Language. Sign Language Studies, 2003, 3(2): 213-225

[72] Cogill-Koez D. Signed language classifier predicates: Linguistic structures or schematic visual representation? Sign Language & Linguistics, 2000, 3(2): 153-207

[73] Liddell S K. Sources of meaning in ASL classifier predicates//Emmorey K ed. Proceedings of Workshop on Classifier Constructions—Perspectives on Classifier Constructions in Sign Languages. San Diego, USA, 2003: 199-220

[74] Bangham J A, Cox S J, Elliott R, et al. Virtual signing: Capture, animation, storage and transmission—An overview of the ViSiCAST project//Proceedings of the Speech and Language Processing for Disabled and Elderly People. London, UK, 2000: 6/1-6/7

[75] Huenerfauth M. A survey and critique of American Sign Language natural language generation and machine translation systems. Department of Computer and Information Science, University of Pennsylvania, Philadelphia: Technology Report: MS-CIS-03-32, 2003

[76] Liddell S K. Grammar, Gesture, and Meaning in American Sign Language. Cambridge, UK: Cambridge University Press, 2003

[77] Xu Lin, Gao Wen. Machine translation oriented understanding and synthesis of Chinese sign language. Chinese Journal of Computers, 2000, 23(1): 60-65(in Chinese)
(徐琳，高文. 面向机器翻译的中国手语的理解与合成. 计算机学报，2000, 23(1): 60-65)

[78] Huenerfauth M. Representing American Sign Language classifier predicates using spatially parameterized planning templates//Banich M T, Caccamise D eds. Generalization of Knowledge: Multidisciplinary Perspectives. New York, USA: Psychology Press, 2010

[79] Bindiganavale R, Schuler W, Allbeck J M, et al. Dynamically altering agent behaviors using natural language instructions//Proceedings of the 4th International Conference on Autonomous Agents. Barcelona, Spain, 2000: 293-300

[80] Butterworth B. Aphasia and models of language production and perception//Blanken G, et al. eds. Linguistic Disorders and Pathologies—An International Handbook. Berlin, Germany: Walter de Gruyter, 1993: 238-250

[81] Von Agris U, Zieren J, Canzler U, et al. Recent developments in visual sign language recognition. Universal Access in the Information Society, 2008, 6(4): 323-362

[82] Canzler U, Ersayar T. Manual and facial features combination for video-based sign language recognition//Proceedings of the International Association for Pattern Recognition (IAPR) Workshop on Machine Vision Applications. Nara, Japan, 2002: 318-321

[83] Van Zijl L, Olivrin G. South African Sign Language assistive translation//Proceedings of the IASTED International Conference on Telehealth/Assistive Technologies. Baltimore, USA, 2008: 7-12

[84] Yao Deng-Feng, Jiang Ming-Hu, Abudoukelimu A, et al. Semantic computing of spatial metaphor based on deaf persons' cognition cases. Journal of Chinese Information Processing, 2015, 29(5): 39-49(in Chinese)
(姚登峰，江铭虎，阿布都克力木·阿布力孜等. 基于聋人案例的空间隐喻语义认知计算. 中文信息学报，2015, 29(5): 39-49)

[85] Lefebvre-Albaret F, Dalle P. Body posture estimation in sign language videos//Proceedings of the 8th International Gesture Workshop, Gesture in Embodied Communication and Human-Computer Interaction. Bielefeld, Germany, 2009: 289-300

[86] Huenerfauth M, Lu P. Eliciting spatial reference for a motion-capture corpus of American Sign Language discourse//Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC). Valleta, Malta, 2010: 121-124

[87] Jiang Ming-Hu. Natural Language Processing. Beijing: Higher Education Press, 2006(in Chinese)
(江铭虎. 自然语言处理. 北京：高等教育出版社，2006)

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

[89] Dorr B J, Jordan P W, Benoit J W. A survey of current paradigms in machine translation. Advances in Computers, 1999, 49(1): 1-68

[90] Zhao L, Kipper K, Schuler W, et al. A machine translation system from English to American Sign Language//Proceedings of the 4th Conference of the Association for Machine Translation. Cuernavaca, Mexico. Lecture Notes in Computer Science, 2000, 1934: 54-67

[91] Speers D. Representation of American Sign Language for Machine Translation [Ph. D. dissertation]. Georgetown University, Washington, USA, 2001

[92] Marshall I, Sáfár É. Extraction of semantic representations from syntactic CMU link grammar linkages//Proceedings of the Recent Advances in Natural Language Processing (RANLP). Tzigov Chark, Bulgaria, 2001: 154-159

[93] Sáfár É, Marshall I. The architecture of an English-text-to-Sign-Languages translation system//Proceedings of the Recent Advances in Natural Language Processing (RANLP). Tzigov Chark, Bulgaria, 2001: 223-228

[94] Sáfár É, Marshall I. Sign language translation via DRT and HPSG//Proceedings of the 3rd International Conference on Computational Linguistics and Intelligent Text Processing. Mexico City, Mexico. Lecture Notes in Computer Science, 2002, 5449: 58-68

[95] Elliott R, Glauert J R W, Kennaway J R, et al. Linguistic modelling and language-processing technologies for Avatar-based sign language presentation. Universal Access in the Information Society, 2008, 6(4): 375-391

[96] Traxler C B. The stanford achievement test: National norming and performance standards for deaf and hard-of-hearing students. Journal of Deaf Studies and Deaf Education, 2000, 5(4): 337-348

[97] Fotinea S E, Efthimiou E, Caridakis G, et al. A knowledge-based sign synthesis architecture. Universal Access in the Information Society, 2008, 6(4): 405-418

[98] Huenerfauth M. Generating American Sign Language Classifier Predicates for English-to-ASL Machine Translation [Ph. D. dissertation]. University of Pennsylvania, Philadelphia, USA, 2006

[99] Ni Xun-Bo, Zhao De-Bin, Gao Wen, et al. Data generation and its validity inspection of signer-independent sign language. Journal of Software, 2010, 21(5): 1153-1170(in Chinese)
(倪训博, 赵德斌, 高文等. 非特定人手语数据生成及其有效性检测. 软件学报, 2010, 21(5): 1153-1170)

[100] Kennaway R. Synthetic animation of deaf signing gestures// Proceedings of the International Gesture Workshop on Gestureand Sign Language in Human-Computer Interaction. London, UK. Lecture Notes in Computer Science, 2001, 2298: 146-157

[101] Ye Ke-Jia, Yin Bao-Cai, Wang Li-Chun. CSLML: A markup language for expressive Chinese Sign Language synthesis. Computer Animation and Virtual Worlds, 2009, 20(2-3): 237-245

[102] Song Yi-Bo, Gao Wen, Yin Bao-Cai, et al. Text driven deaf-mute sign language synthesis system. Chinese Journal of Computers, 1999, 22(7): 733-739(in Chinese)
(宋益波, 高文, 尹宝才等. 文本驱动的聋哑人手语合成系统. 计算机学报, 1999, 22(7): 733-739)

[103] Wang Ru, Yin Bao-Cai, Wang Li-Chun, Kong De-Hui. Video semantic description method for Chinese sign language synthesis. Journal of Beijing University of Technology, 2012, 38(5): 730-735(in Chinese)
(王茹, 尹宝才, 王立春, 孔德慧. 面向中国手语合成的视频语义描述方法. 北京工业大学学报, 2012, 38(5): 730-735)

[104] Huenerfauth M, Lu P. Effect of spatial reference and verb inflection on the usability of sign language animations. Universal Access in the Information Society, 2012, 11(2): 169-184

[105] Chen Yi-Qiang, Gao Wen, Liu Jun-Fa, Yang Chang-Shui. Multi-model behavior synchronizing prosody model in sign language synthesis. Chinese Journal of Computers, 2006, 29(5): 822-827(in Chinese)
(陈益强, 高文, 刘军发, 杨长水. 手语合成中的多模式行为协同韵律模型. 计算机学报, 2006, 29(5): 822-827)

[106] He Wen-Jing, Chen Yi-Qiang, Liu Jun-Fa. Sign language gesture driven head movement synthesis. Journal of Frontiers of Computer Science and Technology, 2012, 6(12): 1109-1115(in Chinese)
(何文静, 陈益强, 刘军发. 手势数据驱动的头部运动合成方法. 计算机科学与探索, 2012, 6(12): 1109-1115)

[107] Krahmer E. What computational linguists can learn from psychologists (and vice versa). Computational Linguistics, 2010, 36(2): 285-294

[108] Neville H J, Bavelier D, Corina D, et al. Cerebral organization for language in deaf and hearing subjects: Biological constraints and effects of experience. Proceedings of the National Academy of Sciences, 1998, 95(3): 922-929

[109] Ngiam J, Khosla A, Kim M, et al. Multimodal deep learning// Proceedings of the 28th International Conference on Machine Learning. Washington, USA, 2011

[110] Gao Wen, Chen Xi-Lin, Ma Ji-Yong, Wang Zhao-Qi. Building language communication between deaf people and hearing society through multimodal human-computer interface. Chinese Journal of Computers, 2000, 23(12): 1253-1260(in Chinese)
(高文, 陈熙霖, 马继勇, 王兆其. 基于多模式接口技术的聋人与正常人交流系统. 计算机学报, 2000, 23(12): 1253-1260)

[111] Feng Zhi-Wei. Computational linguistics: Its past and present. Journal of Foreign Languages, 2011, 34(1): 9-17(in Chinese)
(冯志伟. 计算语言学的历史回顾与现状分析. 外国语, 2011, 34(1): 9-17)

[112] Yao Deng-Feng, Jiang Ming-Hu, Abudoukelimu A, Li Han-Jing. Cognitive computing on Chinese Sign Language perception and comprehension//Proceedings of the 14th IEEE International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC). Beijing, China, 2015: 90-97

[113] Friederici A D, Fiebach C J, Schlesewsky M, et al. Processing linguistic complexity and grammaticality in the left frontal cortex. Cerebral Cortex, 2006, 16(12): 1709-1717

[114] Lu R, Zhang S. Automatic Generation of Computer Animation: Using AI for Movie Animation. Germany: Springer-Verlag, 2002

[115] Coyne B, Sproat R. WordsEye: An automatic text-to-scene conversion system//Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. Los Angeles, USA, 2001: 487-496

[116] Kelleher C, Pausch R. Lessons learned from designing a programming system to support middle school girls creating animated stories//Proceedings of the IEEE Symposium on Visual Languages and Human-Centric Computing. Brighton, UK, 2006: 165-172

[117] Allbeck J, Badler N. Representing and parameterizing agent behaviors//Life-Like Characters. Berlin, Germany: Springer, 2004: 19-38

[118] Birke J, Sarkar A. A clustering approach for the nearly unsupervised recognition of nonliteral language//Proceedings of the 11th Conference of the the European Chapter of the Association for Computational Linguistics (EACL-06). Stroudsburg, USA, 2006: 329-336

[119] Nissim M, Markert K. Syntactic features and word similarity for supervised metonymy resolution//Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL-03). Sapporo, Japan, 2003: 56-63

YAO Deng-Feng, born in 1979, Ph. D., associate professor. His research interests include language cognition and computing, information accessibility.

JIANG Ming-Hu, born in 1962, Ph. D., professor, Ph. D. supervisor. His research interests include natural language processing and cognitive neuroscience.

BAO Hong, born in 1958, Ph. D., professor, Ph. D. supervisor. His research interests include image processing and machine learning.

LI Han-Jing, born in 1974, Ph. D., professor. Her research interest is natural language processing.

ABUDOUKELIMU Abulizi, born in 1983, Ph. D., lecturer. His research interests include language cognition and computing.

Background

Research on sign language information processing has continued for 30 years since AT&T first obtained its patent on data gloves in 1983. Sign language computing is an important topic in the field of natural language processing. However, no major breakthrough has been achieved because of the lack of a raw and annotated corpus. In this paper, we presented a brief review of the development of sign language computing research in the past at the lexical, grammar, and pragmatic levels. Milestone works were specifically recalled to provide a comprehensive insight into this technique. Then, we analyzed the opportunity brought about by the petabyte era for sign language computing. On one hand, benefiting from the rapid development of computer hardware technology, the emergence of the somatosensory equipment and the improvement of computing speed make annotating the large-scale sign language corpus, which could not be achieved before, possible now. On the other hand, new theories and methods for sign language computing continue to emerge, thus further promoting the development of sign language computing. All of these developments indicate that sign language computing will go from the embryonic period to the development period. Moreover, we discussed the next pivotal frontiers, which are possible research directions in the petabyte era, including resource construction, text-to-scene, metaphor comprehension, and classifier predicates in computing. If the four issues can be solved in the development period, then sign language computing will likely develop into a discipline that can keep pace with spoken language computing.

This work was supported by the NSFC Key Project "Research on Chinese Cognitive Processing Mechanism and Computing Model" under Grant No. 61433015, Major Project "Research on Neural Mechanisms of Chinese Nonliteral Language" of the National Social Science Foundation of China under Grant No. 14ZDB154, and the Humanities and Social Sciences Project "ERP Research on the Spatial Metaphor Processing and Neural Mechanisms in the Chinese Sign Language" of the Ministry of Education in China (MOE) under Grant No. 14YJC740104. These projects aim to realize the cognitive computing of Chinese and Chinese Sign Language. An important content is to explore the neural mechanisms of brain processing of the Chinese Sign Language, based on which we can simulate the cognitive process to implement the cognitive computing of Chinese Sign Language. Our research team has been working on brain cognition and linguistic computing for years. Currently, Chinese computing is heavily dependent on the support of big corpus data, based on which machine learning is applied to achieve natural language processing. However, in addition to Chinese, English and other mainstream languages, other languages may not have large data corpus, and corpus construction can be tedious and time consuming. Our team focused on the issues without the aid of a large data corpus and how the human brain perceives and processes language. Sign language is a typical human language, but people pay little attention to sign language computing. This review paper can help us to obtain a comprehensive understanding of the retrospect and prospect of sign language computing and thus advance our research toward more intelligent information processing. If we can learn from the cognitive theory of sign language, we can realize the cognitive computing of sign language to avoid the problem of data sparseness of sign language corpus, which makes possible the expansion of the natural language processing theory of sign language.