

面向信息处理的中国手语音系学标注加工规范

赵源¹, 金澎², 敬思远², 姚登峰^{1,2,3*}, 孟子文¹

¹北京市信息服务工程重点实验室(北京联合大学), 北京

²特殊教育语言智能四川省哲学社会科学重点实验室, 四川乐山

³清华大学人文学院计算语言学实验室, 北京

{annzy,tjtdengfeng,ziwengM}@bnu.edu.cn, jandp@pku.edu.cn, syjing628@126.com

摘要

随着对手语进行大规模数据化处理的需求日益增强, 手语的音系学标注及规范化工作愈发迫切。然而, 手语作为一种视觉-空间语言, 不同于有声语言, 其多信道(手形、位置、手掌朝向、运动方式以及面部表情、躯干动作等非手动特征)信息的复杂性与缺乏统一标注规范, 一直制约着手语语料库构建与自动分析技术的发展。针对这一问题, 本研究在手语音系学理论的指导下, 提出了一套面向中国手语音系学标注加工的系统化规范。该规范由原则和细则两部分构成: 原则部分明确标注对象的粒度、标注单位的界定与分层方式; 细则部分则给出多信道特征的具体标注实例与操作指南。该规范的实施为中国手语多信道特征的系统标注提供基础支撑, 将有助于推动手语识别、翻译、生成以及教学平台的深入发展, 加速中国手语信息处理标准化与规范化的进程。

关键词: 中国手语; 音系学标注; 多信道特征; 规范化; 语料库构建

A Processing Specification for the Phonological Annotation of Chinese Sign Language

Yuan Zhao¹, Peng Jin², Siyuan Jing², Dengfeng Yao^{1,2,3*}, Ziwen Meng¹

¹Beijing Key Laboratory of Information Service Engineering,
Beijing Union University, Beijing

²Sichuan Provincial Key laboratory of Philosophy and Social Science
for language Intelligence in Special Education, Sichuan, Leshan

³Lab of Computational Linguistics, School of Humanities,
Tsinghua University, Beijing

{annzy,tjtdengfeng,ziwengM}@bnu.edu.cn, jandp@pku.edu.cn, syjing628@126.com

Abstract

With the growing demand for large-scale data processing of Chinese Sign Language (CSL), establishing comprehensive and standardized phonological annotation protocols has become increasingly urgent. Unlike spoken languages, CSL is a visual-spatial language whose multi-channel features—encompassing handshape, location, palm orientation, movement, as well as facial expressions and torso actions—are both complex and lack a unified annotation convention. This deficiency has long constrained corpus construction and the development of automated analysis techniques. Guided by sign

* 通讯作者

©2025 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

基金项目: 国家社会科学基金一般项目(21BYY106); 国家语言文字工作委员会一般项目(YB145-25); 教育部人文社科基金一般项目(23YJA740013); 国家自然科学基金重点项目(62036001); 四川省自然科学基金面上项目(2024NSFCS0520); 特殊教育语言智能四川省哲学社会科学重点实验室课题重点项目“手语音韵和形态标注加工规范——基于手语语料库的标注实践”(YYN-2024-2)

language linguistics and phonological theory, this study proposes a systematic annotation and processing specification tailored for CSL phonology. The specification consists of two parts: a set of principles that define annotation granularity, delineate the boundaries of annotation units, and clarify hierarchical structuring, and a set of guidelines providing concrete examples and operational instructions for annotating multi-channel features. The implementation of this specification offers essential support for systematic annotation of multi-channel CSL features, thereby promoting advancements in CSL recognition, translation, generation, and education platforms, and accelerating the standardization and normalization of CSL information processing.

Keywords: Chinese Sign Language , Phonological Annotation , Multi-Channel Features , Standardization , Corpus Construction

1 引言

随着人工智能的发展,自然语言处理技术在多个领域取得进展,但手语信息处理未能同步推进(Yin et al., 2021)。其中一个关键问题在于手语语料库规模普遍较小,且缺乏经手势切分、词性标注和知识挖掘等深度加工的资源,限制了其在自然语言处理中的应用(姚登峰et al., 2015)。此外,语料建设仍高度依赖人工标注,缺乏统一的标注规范,影响处理效率和一致性。

中国手语是中国境内聋人群体广泛使用的语言形式,调查显示中国大陆有大量聋生和聋人使用手语。在推进信息无障碍建设的背景下,《第二期国家手语和盲文规范化行动计划(2021—2025年)》提出加快中国手语的规范化和信息化建设。这项工作有助于回应聋人群体的信息需求,支持特殊教育和社会融合。手语信息处理可促进教育、文化传播与社会参与,也需要从语言学角度深入研究手语的音系、形态等结构特征,为构建手语数据库和语料库提供基础支持。

中国手语属于视觉—手势语言范畴,依托手形、运动、位置与手掌朝向等手动特征,并结合面部表情、体态、口型等非手动特征进行信息表达(邱云峰et al., 2018)。其语法高度空间化和同时化:双手可在同一时刻表达不同语义角色,表情与体态则承担语气、时体等辅助功能。统计显示,约50%手势在汉语中难以找到一一对应的词汇,且方言差异显著,手势变体随地区、年龄及教育背景呈连续谱分布。这些特点导致传统线性Gloss注释体系难以充分呈现中国手语的空间与多信道特征,限制了识别、翻译与生成系统的效果发挥。当前国际公开手语语料库规模普遍偏小,缺乏覆盖多信道信息的深度标注数据,数据瓶颈已成为制约手语AI发展的核心问题(郑璇, 2024)。

在梳理国内外手语标注研究的基础上,本文提出一套适用于中国手语信息处理的多信道音系学标注规范。由于该领域研究尚处起步阶段,本文聚焦于音系学标注,形态学等部分计划后续展开。规范制定参考了语言学与信息处理中的标注方案(俞士汶et al., 2000),结合作机构的手语标注经验,充分考虑中国手语的音系特性与多信道需求。该规范适用于自动处理、语料库语言学研究等多个场景。在技术实现方面,规范结合自动标注与人工校验,提升了效率与准确性。该规范的建立为手语信息处理提供基础支撑,并可拓展至相关应用开发和资源建设。

2 手语标注规范研究现状

手语,作为一种与汉语、英语等有声语言截然不同的自然语言,其核心特征在于空间性。手语通过手势、面部表情和身体姿态在三维空间中的动态变化来传递信息,这种独特的表现形式使得手语具有多信道的特性,不仅在手势形态上具有丰富的变化,还包含了空间位置、动作路径和节奏等重要元素,这些元素共同构成了手语信息的完整表达。这种语言模式与传统线性、时间序列性的有声语言截然不同,对手语的研究和处理提出了更多的挑战,也为其标注和识别带来了复杂性。

目前,手语标注的研究与应用主要集中在图像处理与语言学两个领域。在图像处理领域,手语识别的需求催生了众多手语数据集的建立。这些数据集多是为了满足图像处理需求而生成的,标注方法也围绕视觉特征展开。RWTH-PHOENIX-Weather 2014T数据集(Koller et al., 2015)的建立即代表了这一趋势。该语料库在手语视频上进行包括转写、词语类型、手语句子边

界、相应德语句子边界和翻译在内的多层级标注。这种从图像对应注释标注推动了深度学习在手语识别与翻译领域的进步(Pei et al., 2019; Camgoz et al., 2018)。这类方法虽然在手语识别精度上取得了一定的进展,但大多数模型仅在一定程度上结合了语言模型,能够部分改善语义理解。然而,这些研究往往缺乏语言学家的深入参与,手语语义层面的深入解析仍显不足。至今为止,业界尚未出现被聋人群体广泛认可的商业化手语识别或翻译系统,即便有部分初步的系统推出,仍因功能局限和语义准确度问题而难以满足实际需求。

在手语语言学领域,语言学家为了研究的需要,制定了一系列手语标注规范,如澳大利亚手语语料库标注规范(Johnston and De Beuzeville, 2016)和英国手语语料库标注规范(Cormier and Fenlon, 2014)等。这些规范在手语研究的语法、句法和词汇层面起到了重要作用,为学术研究提供了坚实基础。然而,这些标注规范的设计更多是为传统语言学服务,其重点在于学术研究中的精确性和科学性,对数据的精细标注尤为重视,而未考虑到信息处理领域的实际需求。例如Public DGS Corpus(Prillwitz et al., 2008)、ECHO Corpus(Crasborn et al., 2008)、Auslan Corpus(Johnston, 2010)等项目多利用ELAN(Wittenburg et al., 2006)、iLEX(Hanke, 2002)、SignStream(Neidle et al., 2001)等多层次转录工具,或采取SignWriting(Sutton, 2010)、HamNoSys(Hanke, 2004)等特定符号系统,力求对手语表达进行精确而细致的标注。波士顿大学国家手语与手势资源中心语料库NCSLGR(Belissen et al., 2020)正是此类精细语料库的典型代表,虽然规模有限,但多层次标注使其成为研究者深入探讨手语内部特征的宝贵资源。这种学术导向的标注方式通常成本高昂,耗时费力,因为它们往往需要大量的人工参与和细致的语言学分析,而缺乏自动化和规模化的处理能力,难以满足手语识别和翻译系统在大规模数据处理中的效率需求,这对于推动手语的数字化和智能化研究构成了障碍。这种差距使得手语标注在学术研究与技术应用之间形成了一个鸿沟,限制了手语研究成果的实际转化与推广。

整体上看,从语言学的角度研究手语,不仅需要关注手语作为一种视觉语言的独特表达方式,还应着眼于其深层语义的解析,将研究层面从单纯的手语识别技术提升至语言学的语义理解层次。这意味着,研究手语不能仅仅局限于手势动作的提取和匹配,而应深入探讨手语的音系、形态、句法以及语义等各个语言学层面。特别是语义理解,它是实现手语识别系统准确传达手语使用者意图的关键环节。然而,目前针对手语的系统性研究和标注规范仍然处于起步阶段(张艳琼 et al., 2023),这种状况严重制约了手语研究成果的实用化进程和技术转化效率。因此,亟需制定一套针对手语的、系统且科学的标注规范。这套规范不仅应涵盖手语音系学的基本原则,还需具体细化至手语在语义表达中的细节特征,能够全面反映手语的独特性和多样性。

因此,开发一套兼顾语言学研究与信息处理需求的手语标注规范,已成为该领域的重要方向。该标注规范旨在支持学术研究、语料库建设及自动标注工具的开发。标注体系依据音系学原则和细则构建,用于描述手语的基本单位、特征及其组合规律,以提升标注的一致性与可执行性。此外,规范化标注有助于降低成本、提高效率,并为手语语料库自动构建及识别技术应用提供基础支持。该工作有望为手语研究及其在教育、公共服务等领域的应用提供支撑。

3 中国手语音系学标注原则

3.1 理论基础

音系学作为研究语言中最小对比单元的学科,在有声语言研究中侧重分析语言的声音组成与内部结构(李葆嘉, 2025)。对于有声语言而言,这些对比单元是声音;对于手语而言,音系学意义上的对比单元则表现为更小的手势构成要素。手势通常包含手动特征(Manual),即手形、运动、位置、手掌朝向,以及非手动特征(Nonmanual)五大基础成分,这些基本构成单元即是手语中的最小语音单位,也称为音韵参数(倪训博 et al., 2009; 张吉生, 2016; 顾笙韵 and 张吉生, 2017)。从生理基础来看,听人依靠口语,其中肺、横膈膜、气管、声带、口腔、鼻腔、咽喉共同作为发音器官;而聋人依靠手语,通过身体各个部位的配合完成表达。手语根据“发音部位”不同,可分为手动与非手动特征:手动部分包括从肩部到指尖的区域(即手形、位置、手掌朝向和运动方式),而非手动部分则指除手和手臂之外的其他身体部位,如脸部与躯干等,用以传达面部表情、身体动作等更丰富的语义信息(姚登峰 et al., 2015)。当进一步细分这些特征时,还可界定主手与副手的关系,右手手使用者以右手为主手、左手为副手,左利手则相反。

从理论依据来看,手语音系学强调将手语表达分解为若干最小有区别意义的单位,并结合

多信道特征实现语义的传达。这一理论框架为标注体系的构建提供了基础：将手动与非手动特征有机整合，使标注不仅反映手语的“音韵参数”，更能在空间与时间维度再现手势的细微变化。与传统有声语言的音系学研究类似，中国手语音系学标注应遵循语言学分析的基本逻辑：由小至大，从最小对比单元逐步构建更高层级的语义结构。在确定语言的基本单位之后，接下来的关键环节是划定语言单位的边界。汉语中可通过给汉字标注拼音并进行分词操作来识别词汇边界；手语中则借助音韵参数的变化来确定手势的起讫位置。对于汉语，可通过拼音标注和分词操作将词汇或句子切分，明确词语的范围；对于手语，则通过对各项音韵参数的标注与变化分析来定位每个基本手势的开始与结束，从而完成手语的标注和切分。

音韵参数可视为类似口语中音位的“次词汇”单位：若将口语中不同音位组合成音节，再由音节构成词汇、短语、句子，那么手语中则由各类音韵参数组合形成手势。为分析手势的最小单元，Stokoe(1980)曾提出以“位置、手形、运动”三个方面来转写手势的“Stokoe系统”。然而，该体系在描写细节与序列关系方面不够精确，因此Liddell和Johnson(1993)又提出“运动-保持模型”（Movement-Hold Model），将手势的基本构成区分为运动段与保持段，以序列方式组合。这一模型在理论层面强调手势既具备序列性，也能在某些时段保持相对稳定，从而更细致地呈现手语在时序上的结构。通过这样的思路，手语内部的手形、位置、朝向、运动方式和非手动特征皆可被更清晰地辨析和整合，与有声语言的元音、辅音等结构相对应。

在明确了手语的基本结构和切分方法后，便可进一步审视手语标注在实际应用中的研究进程。前面从语言学层面阐述了手语的内部成分及其与有声语言的对应关系，也以运动-保持模型为例展示了切分手势单元的思路和手段。这些理论与模型为手语的自动处理与理解提供了坚实基础，而如何在实际标注过程中有效利用这些理论模型，正是接下来需要探讨的重点。

3.2 音系学标注原则

为构建科学、系统且适应不同应用场景的中国手语音系学标注体系，有必要在明确手语的本质特征与语言学基础的前提下，制定相应的音系学标注原则。这些原则不仅应扎根于手语音系学理论与通用语言学原则，也须充分考虑中国手语的语言特性与使用环境。通过确保标注的一些原则，可为后续的认可、翻译与生成等研究环节提供高质量数据支撑。本部分首先简述原则的缘故，然后以“水果”实例加以说明，可见表 1¹。

特征信道	时间点	
	00:00:01.160–00:00:01.600	00:00:01.602–00:00:02.681
手势片段		
词汇注释 (Gloss)	水	果
主手手形	i/1 字形	半圆形
副手手形	u 字形	半圆形
主手位置	头-下巴	躯干-中部左侧
副手位置	躯干-中部右侧	躯干-中部右侧
主手运动方式	直线	直线
副手运动方式	未发现副手运动	未发现副手运动
主手掌朝向	掌心向内	掌心向右
副手掌朝向	掌心向内	掌心向左
面部表情	圆唇	圆唇
躯干动作	未发现身体动作	未发现身体动作

Table 1: 手语“水果”的ELAN人工标注情况（ELAN使用视频作为标注材料，这里为了便于理解，使用了图像序列）

语义独立性原则 当手语的最小音韵单元（特定的手形、位置、运动方式或手掌朝向组合）不再是多个更小特征简单叠加的意义总和，而在整体上表现出稳定的、相对独立的语义功能时，应将此组合视为一个可标注的基本单位。这确保标注体现手语内部特征的语义成熟度。在

¹本文所取语料均已取得手势者授权

本例中，“水果”被拆分为“水”和“果”两部分手势，分别对应特定的手形、位置、朝向及运动方式。如果该组合长期稳定地表达“水果”这一词汇，并非仅是“水”与“果”机械相加，那么在标注时，就应认定它们携带了相对独立的语义功能，整体可视为一个“语义成熟”的手语词汇。

结构紧密性原则 若某一手势构造的内部组合极为紧密、在时空上呈现高度定型化，不易随意拆分或置换内部手形、位置、朝向等基本参数，则应将此手势结构整体作为一个单独标注单位。这样有助于区分自然手语中稳定出现的“音韵词素”与临时性、可变动的手势组合。“水”和“果”这两部分的内部要素（手形、位置、掌向等）如果在手语使用者群体中呈现高度固定，一旦更换或拆分就无法表达原先含义，说明其具备较高的结构紧密性。在标注时，可以将“水”或“果”这两个基本手势的组合视为若干相对独立的“音韵词素”，而不必进一步把内部各要素过度拆解。

主手优先原则 在区分主手（利手）与副手时，若遇有标注难点、难以明确某一动作由哪只手主导，则倾向将主手执行的手势特征作为优先标注的基础单位。中国手语中主手常承担核心意义的表达，对其他手的协同动作则可在语境中灵活增补。这一原则体现中国手语常规使用中的主副手角色分化。该示例对“水”和“果”分别列出了主手与副手的手形、位置和朝向。如在实际标注过程中，对副手（如“u字形”、“半圆形”）缺乏充分把握时，可先依照主手（如“i/1字形”或“半圆形”）进行优先标注，确保记录了核心信息。如果副手动作重要，则可在同一时间片段中补充标注或在注释层说明。

动态可调粒度原则 在确保标注标准统一与清晰的前提下，可根据研究目标对标注颗粒度适度调整。例如，在基础研究中，为考察中国手语中最小音系单元的构成，可选择较为精细的分解标注；而在实际应用（如机器翻译、教学实践）中，可选择较为整体化的标注方式，以提高系统处理效率。在“水果”示例中，已经对主手、副手、位置、运动方式等做了区分；但若对一些未能确定的动作（标记为“未发现”）需要更细微的拆分，可以在后续精细研究中深入补充。若是为机器翻译或教学演示，维持现有较为“简化”的标注也足以表达主要语义。

非手动特征补充原则 手语意义往往来自多信道特征的协同作用。对于面部表情、躯干动作等非手动特征，在不改变手势基本结构标注逻辑的前提下，可在必要时单独标注或附加标记。如果非手动特征为特定语义的必要条件，则应将其与手动特征整合在同一标注层次中，从而为手语语义理解提供更全面的依据。该示例中记录了面部表情为“圆唇”，躯干动作则标为“未发现身体动作”。若在后续研究发现“圆唇”对“水果”的语义或语气有特殊影响，则可将此特征与手动动作放在同一层次进行整合，以突出“圆唇”在表达“水果”时的辅助意义。同理，若躯干动作在更多情境中扮演着不可或缺的语义功能，也应将其纳入统一标注体系加以记录。

上述原则的制定与实施得益于稳固的理论指导和成熟的技术手段。一方面，手语语言学、语言类型学以及手语音系学研究为标注标准的确立提供了坚实的理论支撑；另一方面，诸如OpenPose、MediaPipe等成熟的关键点检测与姿态估计技术，能够为标注原则的实施提供必要的客观、定量数据支撑。通过理论与技术的双重驱动，这些标注原则不仅能够支持现有研究的开展，也能在未来的数据积累与算法优化中发挥持续作用。

中国手语音系学标注原则的提出，既是对手语本质特征的尊重，也是面向实际应用需求的有力回应。通过明确的标注标准体系，有助于为中国手语的深入研究、自动处理和多模态语言应用铺平道路。

4 中国手语音系学特征描述与分类

音系学的多信道特征是中国手语构成与表达的核心元素，手部动作与非手动特征的协调与组合直接影响手语语义的完整与准确。在这一体系中，手动部分和非手动部分相互配合，形成视觉语言的多维表征模式。为深入研究和标注这些内部特征，本研究结合手语音系学理论，并采用关键点提取技术（以OpenPose、MediaPipe为代表）作为获取量化数据的途径，在多次词汇查询、语料分析和参考文献的基础上，对手动与非手动特征进行系统分类与定义。同时，明确标注层次与编码原则，为建立科学、全面且高效的音系学多信道手语标注体系奠定坚实基础。

4.1 手动特征的定义与分类

手语作为一种依赖视觉-空间传递信息的语言，其表达主要由手部动作、面部表情和身体姿态等构成。在这些表达方式中，手部动作起到了核心作用。为了更清晰地分析和描述

手语的音系学结构与表达规律，我们将手语中以手部为中心的动作特征进行定义和分类，这些特征被统称为“手动特征”。本研究利用关键点提取技术生成的身体、面部及手部关键信息点进行标注，这些信息可来自OpenPose(Cao et al., 2017) (BODY25、Face、Hand模型)或MediaPipe(Lugaresi et al., 2019; Bazarevsky et al., 2020; Kartynnik et al., 2019; Vakunov et al., 2020) (Pose、Face Mesh、Hands解决方案)。当检测并提取到两手信息时，标注过程中会区分主手 (H1) 与副手 (H2)，以适应不同个体的利手习惯。考虑到个别情况下，如两个手语使用者对话时，摄像头可能会只关注对话者表情，导致两只手不在拍摄范围内，标注时可设定主手和副手两层，每一层若能检测到相应的手部，则标记为1，反之为0。图 1展示了主手 (H1) 与副手 (H2) 的区分示意图。

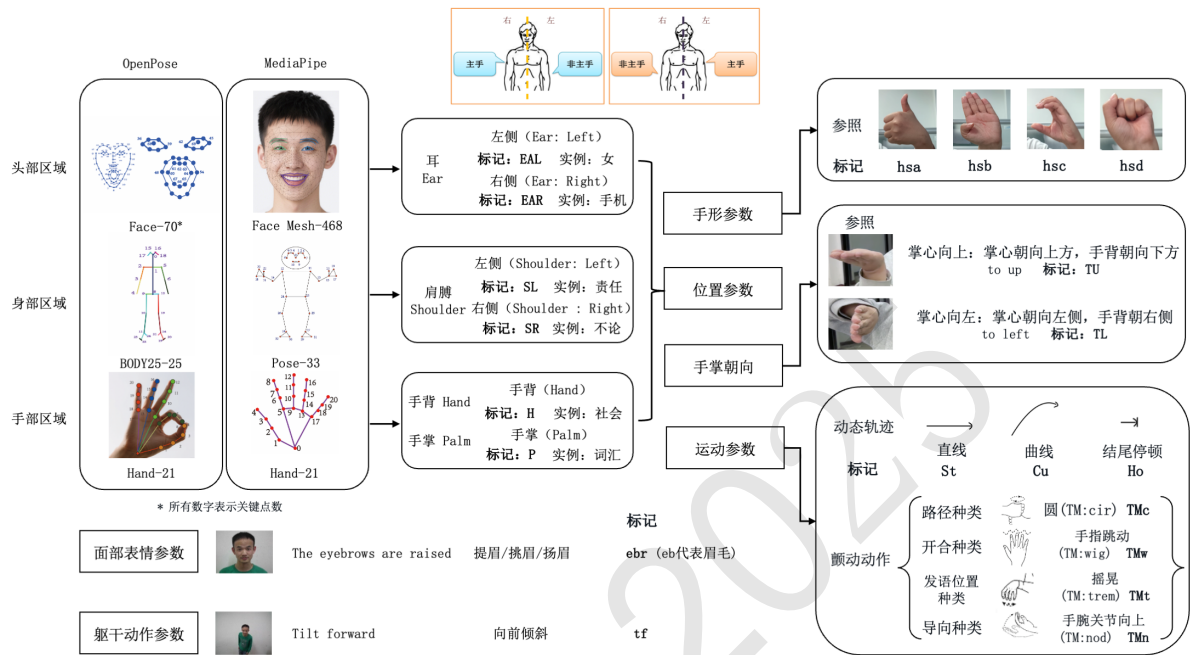


Figure 1: 利手、姿态估计框架以及各信道参数的标注规范部分内容

本研究在构建手动特征标注规范时，参考了早期Stokoe提出的对手语进行音系层面分解的框架。该框架通过一系列缩写符号标记手势内部的构成要素，为手语研究提供了重要的理论基础。此后，众多研究者在此基础上进行了多次扩充与细化，形成了更为广泛采用的若干参数体系。本研究并未完全照搬这些后期形成的参数体系，而是基于早期Stokoe框架，紧密结合中国手语本身的特点，自行建立了面向信息处理的音系学标注体系和规范²，确保了标注体系的科学性、适用性和独特性。

4.1.1 位置参数

在位置参数标注中，本研究以OpenPose框架输出的关键点信息为例子（利用其Face、BODY25和Hand模型），对手部、头部和身部区域进行细分，其关键点的可视化分别为图 1的Head头部图、Body身部图、Hand手部图。手部动作在不同身体参考区域的位置差异为手势的判定与分析提供了重要依据。基于OpenPose和MediaPipe等主流姿态估计技术均能提供覆盖头部、身部、手部这三个总位置的关键点数据（具体数量对比见图 1）的特点，本研究将这三者定义为位置标注的三个基本区域类别。在此基础上，我们进一步定义了一级位置（共包含49个具体位置标签）和二级位置（共包含85个更细化的位置标签）。本规范以二级位置作为位置标注的最小颗粒度，通过在固定参考点基础上扩展二级位置标记的方法，实现精确且细化的多层次位置标注，以提高标注数据的可用性和一致性（部分规范细节见图 1）。经统计，本规范共定义一级位置标签49个，二级位置标签85个。若手势无明确或特定的位置信息需要标注，可标记为“NOL”。

²本研究的音系学标注规范、标注实例及相关资料已上传至GitHub，可通过以下网址访问：<https://github.com/ann-yuan/CSL-PhonoSpec>。

4.1.2 手形参数

手形参数用于描述手部形状的外在特征，包括手指的弯曲程度、相互间的相对位置以及手掌姿态的变化。部分手形在不同位置可能表达相似或相同的意义，因此需要结合语境信息以确保标注的精度。本文参考骆维维(2008)和衣玉敏(2008)等学者对手形的研究，整理并归纳出69种基本手形。图 1展示了7个手势的意义、手形图及其标记，以确保每种手形具备一定的独立性。如果没有展示其手形的变化，可标记为“NOH”。

未来，随着手语研究的不断深入，不排除可能会出现新的手形类型。本文提出的基本手形体系在标注规范制定时已尽可能全面地覆盖现有手语手形，但如在后续研究中确认了新的手形类别，也将在标注规范修订时进行充分评估，并根据实际需求进行完善。

4.1.3 手掌朝向参数

手掌朝向参数用于描述手掌在空间中的指向方向。理想情况下，借助Leap Motion等体感设备的三维数据或多视角视频捕捉深度信息，可以在三维坐标系中精准刻画手掌的朝向变化。然而，目前网络上的手语视频大多由手机或摄像机拍摄，通常不包含三维数据。考虑到大规模采集专用三维设备数据的成本较高且不切实际，本研究基于单一正视角视频进行分析，并在此基础上探索提出了针对三维数据问题的近似解决方案。

OpenPose在此情境下只能输出二维关键点数据。由于缺乏多视角或深度信息，无法实现手掌朝向的三维坐标分析。我们注意到，MediaPipe的Hands解决方案能够从单视角视频中直接估计出手部的三维关键点坐标，这为将来更精确地判断手掌朝向提供了非常有潜力的技术路径，如果未来开发条件允许，我们倾向于采用此类信息。然而，考虑到现有手语视频数据的普遍性（多为普通2D视频）以及在当前研究阶段确保方法普适性和稳定性的需要，本规范在手掌朝向参数的标注上，仍以基于二维平面中关键点相对位置进行近似描述作为核心和基础方法。

这种方法虽无法提供完整的三维朝向信息，但在单摄像头条件下仍具有可操作性和实用价值。实际标注时，可通过手部关键点与人体参考位置的二维坐标，结合向量方向或角度阈值进行近似判定。图 1展示了手掌朝向参数的部分标注规范，目前共有10个主要朝向类型（参考杨峰(2016)、衣玉敏等学者的研究）。其中，“掌心向内”与“掌心向外”等方向需特别注意，在使用当前核心的二维近似方法时，深度信息的缺失是主要限制；即便未来采用估计的三维信息，也需关注其潜在的估计误差，两者都可能需要在标注后期通过人工校对来弥补自动识别的误差。如果没有展示其手掌朝向的变化，可标记为“NOO”。

4.1.4 运动参数

运动参数涵盖手部在动作中的动态轨迹和颤动动作。动态轨迹是指手臂驱动手部运动时的变化，而颤动动作是指手部本身的运动，不依赖手臂，其分类依据动作的物理属性及语言学功能细化为四类：路径种类、开合种类、发语位置种类与导向种类。手势的空间位移、变换速率以及是否存在短暂颤动等动态表现，都需要在标注中明确体现。通过细分动态轨迹(6种)和颤动动作(11种)，见图 1，不仅可以捕捉手语表达的过程性信息，还能呈现不同手势在时空维度上的连贯性和差异性，重点关注每只手在运动中的变化。如果没有展示其运动的变化，可标记为“NOM”。

4.2 非手动特征的定义与分类

非手动特征主要指不涉及手部动作的其他身体特征，包括面部表情与躯干动作。面部表情中，表情肌肉的细微变化能够在视觉层面传递语气、态度及情绪信息，是手语含义的重要补充。然而，现有姿态估计框架的关键点分布往往较为简化，对面部与躯干的细节捕捉尚有局限，这需要在数据处理阶段引入更加精细的模型和识别方法。

4.2.1 面部表情参数

面部表情通过对面部肌肉动态的描述，丰富了手语语义，弥补了仅靠手部动作无法充分表达的意义空缺。图 1展示了面部表情参数部分标注规范，经统计共有7个主表情部位和25个常用表情。如果没有展示其面部表情的变化，可标记为“NOF”。

4.2.2 躯干动作参数

躯干动作参数体现身体在前倾、后仰、旋转等姿态上的变化。通过对躯干动作进行分类与标记，可为手语表达提供更立体的语义支撑，使最终的标注体系更接近手语在实际交流情境中

的表现方式。图 1 展示了躯干动作参数中的 10 个常见动作的一部分。如果没有展示其躯干动作的变化，可标记为“NOB”。

通过系统定义与分类手动和非手动特征，本研究总结并提出了一套基于多信道特征的标注体系。手动特征的标注以 MediaPipe、OpenPose 等关键点提取技术为基础，本研究明确了手形、位置、手掌朝向和运动的标注方法；非手动特征结合面部表情识别与人体动作检测模型，准确描述了面部表情和躯干动作的多维信息。这一体系不仅确保了标注的一致性和全面性，也为手语的自动化标注与语料库建设提供了理论支持。

然而，考虑到手语词汇和句式的多样性与复杂性，本研究未对所有词汇与句子进行详尽分析，而是基于现有特征分类构建了一个开放的标注框架。该框架结合自动标注工具，适用于不同应用场景中的手语参数识别与标注任务。研究过程中如出现新特征或未定义参数，可经评估后作为后续标注规范调整的依据。该设计提升了标注体系的适应性，也为后续手语研究提供了扩展空间。

5 中国手语多信道标注细则

在对中国手语进行多信道标注时，既要确保标注格式与数据结构的高度统一与可扩展性，又要具备对手语多维特征的精细呈现能力。基于语言学理论与实践经验，本研究提出一套多层级标注方案及对应的数据结构，可输出为 XML 等格式，帮助使用者在同一时间基线上分别记录主手、副手、面部表情、躯干动作等信道特征。最终标注结果在结构化存储后，既能支持进一步的算法分析，也能为语言学研究提供可靠依据。

5.1 标注格式与数据结构的定义

为精确呈现手语在多个信道上的同步变化，本研究将主手、副手、面部表情与躯干动作分别作为独立的标注轨道。在标注过程中，每个轨道仅关注某一特征层次，例如主手的手形、位置、运动和朝向，或面部表情中的核心变化点。这种分轨方式能够确保特征标注的层次清晰，并在时间维度上实现不同信道的对齐。

在正式标注前，需要划定手语词或句子的时间范围，以明确标注单位在时间轴上的对应片段。时间范围的划定基于语义单位，即每个完整的手语词汇或句子均被分配特定的时间跨度。在此基础上，若某一手势内部存在显著的阶段性变化，则需进一步细分时间范围。例如，当某一手语动作从“1 字形”过渡到“a 字形”时，可将其划分为两个相邻的时间片段，分别记录每个阶段的手形及其对应的其他特征（如位置、朝向和运动方式）。

标注过程中，每个时间片段的相关信息会以层次化文件格式（如 XML）记录。具体而言，每个时间片段被定义为一个标记单元，包含以下三个主要信息：

时间起止点：记录手势动作的开始与结束时间。

特征标注：包含主手、副手及非手动特征的详细描述，例如主手的手形代码（如“hsa(a 字形)”）、位置代码（如“TMC(躯干：中部中心)”）以及朝向描述（如“TL(掌心向左)”）。

层级结构：每一特征层次对应一个独立的标注层（Tier），所有层在时间轴上实现同步对齐。

5.2 多信道特征标注细节说明

5.2.1 手动特征——位置参数

在标注过程中，主手和副手各自起始动作明确后，需要进一步辨识并记录每只手在动作结束时所处的具体位置。此处所指的位置不仅是手与身体主要部位（如头部、胸部、躯干正中等）的相对关系，也包括手在空间坐标中的停留点。若手语词或句子中包含显著的手势变化，可依据以下三种情形来判断每只手的确切位置：

第一种情形，若某只手在整个动作过程中几乎不动，或仅进行微小调整，则可视作它自始至终停留在同一位置范围内。以主（副）手长时间贴近胸部左侧为例，可以在标注中将其位置统一标记为某个固定的身体参考点或局部区域。通过对该情形的明确记录，能够展现主（副）手在手势表达中的静止或接近静止状态，为分析主手与副手的配合关系提供清晰的参照。如图 2 展示手语“难”，右手为主手，放在额头左边，几乎不动，即可标记主手在前额：左侧。

第二种情形，当手形在某一时刻发生变化时，可将其视为动作时间段的分割节点，并在前后两个时段分别标注手部位置。换言之，一旦检测到“手形由 X 变为 Y”的转换瞬间，就先记录

手势变化前的手部位置，然后在新时段开始处标明变化后的手部位置。这种分段方法有助于在标注层次上精确反映手势阶段性变化的过程，也使得后续对手形、位置与时序的关联分析更为便利。如图 2 展示手语“住院”，右手为主手，有两个手势，分别放在左面颊和额头眉心，即可标记主手在不同手形的所在位置为面颊：左，前额：正中心（眉心）。

第三种情形，若一只手沿着简单、连贯的路径从A点移动到B点，而动作过程本身无其他特征变化（如手形或朝向都未改变），则可分别记录起点A和终点B的位置即可。标注者可以在时间轴上将此移动视为一个完整的连续动作，将起点位置与终点位置分别标示出来。若研究对手部的具体路径或中途细微停顿感兴趣，则可依据需要进一步细分；但在多数情况下，仅记录起点和终点就能清晰呈现手部从A移动到B的空间路径。如图 2 展示手语“快”，右手为主手，从A点移动到B点的动态轨迹为直线运动，分别A点为躯干外面和B点为躯干的中部右侧，即可标记主手在不同手形的所在位置为躯干：外→躯干：中部右侧。

5.2.2 手动特征——手形参数

标记每个手的手形，若某个手形在整个动作持续时段内保持不变，标注者只需在标注文件中为该时间区段统一记录相应的手形即可。然而，若手形在动作中出现明显的转变，例如从“一字形”过渡到“拳握形”，则应在时间轴上将该过渡时刻作为分割节点，并为前后两个阶段分别标注手形。假设在1.0秒到1.8秒的动作中，主手原本是“一字形”，而在1.4秒左右开始逐渐收拢成“拳握形”，则可将该时段细分为两个片段：1.0秒至1.4秒标注“一字形”，1.4秒至1.8秒标注“拳握形”。如果研究者更关心手形过渡过程中的细微差异，亦可进一步增设更多时间切分点，对过渡阶段做更精细的时序标记。

在实际使用中，还需考虑到个体差异和生理抖动等因素。由于不同人在打同一个手势时可能呈现稍有差异的手指弯曲度或手掌张合度，标注方面应在保证语义区分度的前提下进行必要的聚合，避免为微小变动过度拆分时段，导致标注结果碎片化并难以使用。

5.2.3 手动特征——手掌朝向参数

在手语表达中，手掌朝向对于区分不同含义或句式起到重要作用。本研究考虑到大多数视频的拍摄角度都为正视角，因而将手掌朝向参数限定在十种主要类别，如“掌心向上”“掌心向内”“掌心向左”等。这种相对简化的分类方法既能在语言学研究和标注实践中达到可用的辨析度，又避免了因深度信息缺失而导致朝向判断过度细分的情况。

在标注时，每当手形发生变化，只需记录变化之前的手掌朝向即可。比如，若某只手在最初使用“1字形”时掌心朝左，而后在1.4秒左右转换为“拳握形”，手形变化但朝向无变化，则只需在记录该时刻前的手掌朝向即可；从手形发生改变之时起，新的朝向将与新的手形绑定在一起，形成另一个时段的标注单元。如果在时长较短的连续运动中手掌朝向有显著变动，仅需标明运动起点与终点的朝向即可；中途略微的偏转或角度修正通常无需详细拆分，除非此变化对整体语义或句型产生了明显影响。如图 2 展示手语“白菜”中的“菜”，右手为主手，有两个手形，分别为o字形和5字形，其运动方式为颤动动作，即手指放开，其手掌朝向为掌心向上。

5.2.4 手动特征——运动方式参数

在绝大多数单视角（正视角）视频中，深度与侧面信息相对缺失，因而对于运动方式参数的标注往往只能概括地记录手势在平面上的大幅度移动或颤动。与手掌朝向一样，此处并不对运动方式做过细的分段。

在记录动态轨迹时，可依据手部从一个位置移动到另一个位置的路径及幅度加以区分。若动作幅度较大，且可清晰观察到手势在视频帧中从A点明显地转移至B点，则可在标注文件中简要指出该移动对应的轨迹类型，例如“弧形”“直线”或“曲线”等，可见参考图 2 的动态轨迹。通过这一方式，即可反映出手势的主要方向与轮廓形态，而无需对中间过程的细微摆动或轻度偏移做过度拆分。若研究需求更关注特定情境下的轨迹细节，可在后期基于时间轴做更精细的二次标注。

在颤动动作的判定中，若某只手在整个时段中保持大体静止位置，但手形或手掌朝向在局部做出小范围、高频率的变化，也可将其视作颤动动作的一种表现形式，可见参考图 2 的颤动动作。例如，当手位于躯干侧方近似静止，却在腕关节或指尖出现快速摆动时，可将此情形标记为“手腕摆动”或“手腕转动”运动。这样做既能在宏观层面呈现手部运动的主要走向，也能在细微处捕捉到手势内部高频率的小幅变动。同时如果遇到动态轨迹和颤动动作同步进行，参考

图2的手语“快”，可以看出动态轨迹是直线运动，但是仔细看手掌朝向变了，说明它也在做颤动动作，为手腕转动，可以标记为直线运动/手腕转动，动态轨迹在前，颤动动作在后，如果其他手语只展示单个运动方式，只需要标记一个即可。

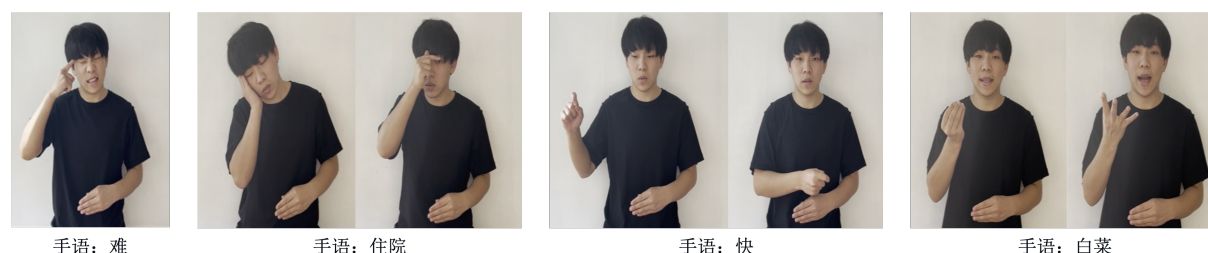


Figure 2: 每个手语所对应的信道变化

5.2.5 非手动特征——面部表情参数和躯干动作参数

在手语表达中，面部表情和躯干动作虽然通常被视为辅助或修饰性信道，然而在特定语境下亦能显著影响手语的语义与语气。为提高标注的完整度，本研究同样将这类“非手动特征”纳入多信道标注体系。

在实际的标注过程中，若某段时间里面部表情仅保持自然状态，未出现诸如张口或抿嘴等较显著动作，亦未见明显的眉眼提升、嘴角上扬等情绪外露，标注方面便可在相应时间片段里记录“NOF”的信息。相似地，若躯干在整个动作过程中近乎静止，未有明显前倾、后仰、左右摆动等姿势变化，同样可记为“NOB”。这样做不仅能精简标注过程，也使得标注文件更为明晰易读。

若研究者对此类非手动特征有更深入的需求，例如需关注极细微的面部表情或微小幅度的上身动作，则可以在后期或不同标注层次上追加相应的精细标注。无论是否针对这些细微特征进行详细刻画，通过设置“NO-”这一标记，可确保在主手和副手特征相对突出的情境下，标注工作依旧遵循统一的多信道原则，并在必要时留有拓展空间。

6 标注规范的初步应用与实践反馈

当前，手语语料资源建设普遍面临缺乏自动标注工具的瓶颈，大量语料仍停留在“生语料”阶段，制约了其规模化处理与计算建模能力的提升。本研究提出的中国手语音系学标注规范，系统定义了适用于自动处理的标注单元与结构层级，为手语语料的结构化加工与自动化标注提供了方法依据。该规范以多轮语料观测与标注实践为基础构建，在国内尚属首次，在国际范围内亦属于开创性工作，具有较强的理论前瞻性与工程适应性。

为验证规范的可用性，我们结合人工校对与自动标注的协同策略，开展了原型系统开发与小规模实测。具体而言，系统基于关键点检测与姿态估计框架（如OpenPose），提取视频中手动特征（包括手形、位置、手掌朝向与运动方式）及非手动特征（如面部表情、躯干姿态等），并结合时序分析与特征对齐机制，辅助判断特定手势片段的起止位置、构形变化及信道间同步关系。初步实现的原型性自动标注模块已可支持多信道信息的结构化识别，具备一定的程序化能力。

系统已在乐山师范学院、北京联合大学、清华大学等多所高校进行了小规模语料测试。实践发现，部分信道（尤其是手形）的识别准确率仍然较低（约为30%），主要受制于类别区分细、样本分布不均及训练数据不足等问题。这一结果体现了手语标注任务的复杂性，也突显了规范在流程设计与结构控制方面的重要作用。为提高后续处理效率，该系统引入深度学习模型进行特征抽取与概率推断，以缓解逐帧人工标注的工作强度，并为标注与模型之间的迭代优化提供接口支撑。

此前手语语料的加工多集中于词汇级Gloss注释，缺乏音系层级的结构化资源，严重影响了语料的再利用能力与分析深度。作为补充，本规范面向信息处理任务，聚焦音系学要素的多信道并行表达，初步填补了长期存在的结构性缺口。配套的标注系统及工作流程为构建标准化语料奠定了基础，也为后续模型训练、语料共享与资源评价提供了规范支撑。未来我们将在规范指导下完成小规模语料的精标工作，并在配套论文中系统发布工具设计、实验方案与评估结

果（如关键点检出率、标注准确率及人工校对工时占比等），以全面验证规范的工程实用性与科学有效性。

总体来看，作为一项规范性研究，本工作聚焦于构建面向中国手语信息处理的音系学标注体系，明确了语料加工的基本单元、结构规则与信道映射方式，为推动手语语言资源标准化提供了方法学支持。在语言信息处理领域，“以规范为起点”的方法路径是高质量语料库建设的前提。手语因其视觉依赖、多信道并发与空间结构表达等复杂特性，在规范设计、系统开发与工具评估方面面临更高挑战，本研究即为在语言建模与计算实现之间寻求深度融合的一次探索性尝试。

7 结语

本研究基于中国手语音系学的理论基础，结合多信道标注原则与结构建模方法，系统构建了面向信息处理的音系学标注规范。该规范明确了手动与非手动特征的基本单元与结构层级，规定了5条标注原则与6个信道划分，并设计了细致的分类标记体系，通过ELAN工具与XML格式实现了多层次、多信道的统一编码与加工流程。以音系结构为纲的标注实例进一步提升了规范的可操作性与示范性。

本规范已在多所高校的语料库建设中投入应用，初步验证了其在结构化标注、标注一致性控制和语料组织方面的实际效用。配套自动标注模块已完成原型开发，并结合姿态估计框架（如OpenPose、MediaPipe与MMPose）开展了小规模测试，揭示了手形等信道在标注准确率上的挑战，后续将通过算法优化与多模态补全进一步提升标注完整性与系统稳定性。

作为一项规范性研究，本文提出的标注体系为中国手语语言资源的标准化加工提供了方法基础，不仅有助于语料资源的规模化建设与共享，也为后续的手语识别、翻译与生成等信息处理任务奠定了结构层级上的数据支撑。未来工作将聚焦于标注流程优化、质量评估体系构建与配套工具开放，以期推动中国手语语言资源迈向系统化、标准化与智能化的发展路径。

参考文献

- Andrey Vakunov, Chuo-Ling Chang, Fan Zhang, George Sung, Matthias Grundmann, and Valentin Bazarevsky. 2020. Mediapipe hands: On-device real-time hand tracking. In *Workshop on Computer Vision for AR/VR*, volume 2, page 5.
- Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, et al. 2019. Mediapipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172*.
- Carol Neidle, Stan Sclaroff, and Vassilis Athitsos. 2001. Signstream: A tool for linguistic and computer vision research on visual-gestural language data. *Behavior Research Methods, Instruments, & Computers*, 33(3):311–320.
- Kayo Yin, Amit Moryossef, Julie Hochgesang, Yoav Goldberg, and Malihe Alikhani. 2021. Including signed languages in natural language processing. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 7347–7360.
- Kearsy Cormier and Jordan Fenlon. 2014. Bsl corpus annotation guidelines. *London: Deafness Cognition and Language (DCAL) Research Centre of the University College London*.
- Necati Cihan Camgoz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. 2018. Neural sign language translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7784–7793.
- Onno A Crasborn, Els Van Der Kooij, Dafydd Waters, Bencie Woll, and Johanna Mesch. 2008. Frequency distribution and spreading behavior of different types of mouth actions in three sign languages. *Sign Language & Linguistics*, 11(1):45–67.
- Oscar Koller, Jens Forster, and Hermann Ney. 2015. Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. *Computer Vision and Image Understanding*, 141:108–125.

- Peter Wittenburg, Hennie Brugman, Albert Russel, Alex Klassmann, and Han Sloetjes. 2006. Elan: A professional framework for multimodality research. In *5th international conference on language resources and evaluation (LREC 2006)*, pages 1556–1559.
- Scott K Liddell. 1993. Holds and positions: Comparing two models of segmentation in asl. In *Current issues in ASL phonology*, pages 189–211. Elsevier.s
- Siegmund Prillwitz, Thomas Hanke, Susanne König, Reiner Konrad, Gabriele Langer, and Arvid Schwarz. 2008. Dgs corpus project—development of a corpus based electronic dictionary german sign language/german. In *sign-lang@ LREC 2008*, pages 159–164. European Language Resources Association (ELRA).
- Thomas Hanke. 2002. ilex-a tool for sign language lexicography and corpus analysis. In *LREC*.
- Thomas Hanke. 2004. Hamnosys—representing sign language data in language resources and language processing contexts. In *sign-lang@ LREC 2004*, pages 1–6. European Language Resources Association (ELRA).
- Trevor Johnston and Louise De Beuzeville. 2016. Auslan corpus annotation guidelines. *Auslan Corpus*.
- Trevor Johnston. 2010. From archive to corpus: Transcription and annotation in the creation of signed language corpora. *International journal of corpus linguistics*, 15(1):106–131.
- Valentin Bazarevsky, Yury Zhang, Andrey Vakunov, Karthik Tkachenka, Fan Sung, Chuo-Ling Chang, and Matthias Grundmann. 2020. BlazePose: On-device real-time body pose tracking. *arXiv preprint arXiv:2006.10204*.
- Valentin Belissen, Annelies Braffort, and Michèle Gouiffès. 2020. Dicta-sign-lsf-v2: remake of a continuous french sign language dialogue corpus and a first baseline for automatic sign language processing. In *LREC 2020, 12th Conference on Language Resources and Evaluation*.
- Valerie Sutton. 2010. The signwriting alphabet. *Read and Write any Sign Language in the World. ISWA Manual*.
- William C Stokoe. 1980. Sign language structure. *Annual review of anthropology*, pages 365–390.
- Xiankun Pei, Dan Guo, and Ye Zhao. 2019. Continuous sign language recognition based on pseudo-supervised learning. In *Proceedings of the 2nd Workshop on Multimedia for Accessible Human Computer Interfaces*, pages 33–39.
- Yury Kartynnik, Artsiom Ablavatski, Ivan Grishchenko, and Matthias Grundmann. 2019. Real-time facial surface geometry from monocular video on mobile gpus. *arXiv preprint arXiv:1907.06724*.
- Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2017. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7291–7299.
- 顾笙韵 and 张吉生. 2017. 手语音系研究及其理论模型. 外国语(上海外国语大学学报), 40(01):52–65.
- 李葆嘉. 2025. 西方语言学“音位”“音位学”新诠释. 辞书研究, (01):53–76+126.
- 骆维维. 2008. 《中国手语》手形研究. Master’s thesis, 北京师范大学.
- 倪训博, 赵德斌, 姜峰, and 程丹松. 2009. 中国手语音韵标记的建立、实现及其有效性验证. 计算机学报, 32(12):2438–2453.
- 邱云峰, 姚登峰, 李荣, and 刘春达. 2018. 中国手语语言学概论. 中国国际广播出版社.
- 姚登峰, 江铭虎, 阿布都克力木·阿布力孜, 李晗静, 哈里旦木·阿布都克里木, and 夏娣娜. 2015. 中国手语信息处理述评. 中文信息学报, 29(05):216–227.
- 姚登峰, 江铭虎, 鲍泓, 李晗静, and 阿布都克力木·阿布力孜. 2019. 手语计算30年:回顾与展望. 计算机学报, 42(01):111–135.
- 杨峰. 2016. 上海手语音节结构分析. Ph.D. thesis, 华东师范大学.
- 衣玉敏. 2008. 上海手语的语音调查报告. Ph.D. thesis, 复旦大学.

- 俞士汶, 朱学锋, and 段慧明. 2000. 大规模现代汉语标注语料库的加工规范. 中文信息学报, (06):58–64.
- 张吉生. 2016. 从有声语言音节看手语音节的理据及特点. 中国特殊教育, (06):17–25.
- 张艳琼, 朱兆松, and 赵晓驰. 2023. 面向手语语言学的中国手语词汇多模态语料库构建研究. 数据分析与知识发现, 7(10):144–155.
- 郑璇. 2024. 手语数字人研发现状与思考. 语言战略研究, 9(03):17–28.
- 邱云峰, 姚登峰, 李荣, and 刘春达. 2018. 中国手语语言学概论. 中国手语语言学概论.