

非语言信息的认知加工路径 及其对大语言模型多模态翻译的启示

黄婕

对外经济贸易大学英语学院

<https://chezvivan.github.io>

研究背景与问题



背景

多模态翻译需求急剧增长



历史回顾

从文本为中心到多模态共构意义



研究空缺

非语言信息认知加工路径尚不明确

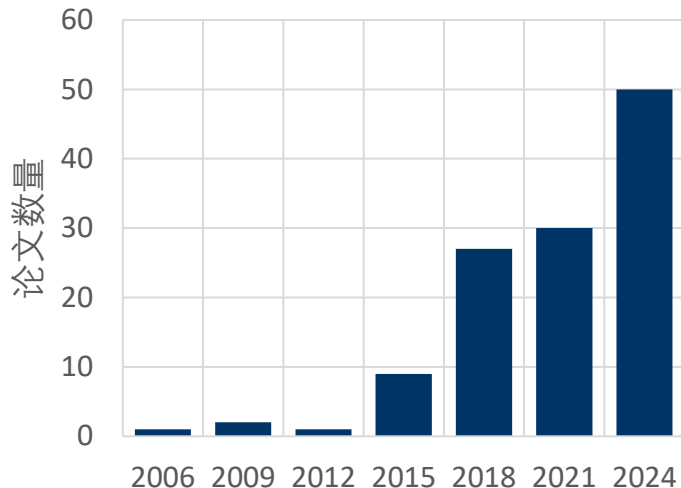


核心问题

非语言信息如何影响翻译认知过程？

研究热度上升

2004-2024 期刊论文含“multimodal translation”主题关键词数量



数据来源: Web of Science (2025-07 检索)

理论框架

认知负荷理论

非语言信息增加认知负担

分散注意力资源

增加工作记忆压力

VS

双重编码理论

非语言信息辅助理解

激活双通道加工

增强记忆与理解效果

本研究假设

非语言信息处理存在最优化认知路径：适度非语言信息辅助而非干扰认知加工

研究方法概述

研究设计

变量类型	详细说明
自变量	非语言信息密度 : 高 vs 低 视频类型: 人物特写 vs 画外音
因变量	眼动指标: 注视时长、注视次数、注视分布 键盘记录: 键盘操作次数、停顿时长与次数 主观评价: NASA-TLX 认知负荷评估

被试: n=30

翻译硕士, 拥有CATTI 翻译资格证书
平均年龄23.53岁, ± 1.26 岁

研究流程



被试招募

招募30名翻译专业学生参与实验



眼动与键盘记录

同步记录翻译过程中的眼动和键盘操作



主观评价

任务完成后填写NASA-TLX认知负荷评估量表

实验材料与设计

高非语言信息密度



近景、表情和身体动作、角色叙事

低非语言信息密度



远景、无身势语、画外音

材料控制策略

来源

科普纪录片片段

控制变量

语言难度指数（可读性指标、词频等）

时长均为28秒±5秒

主题均为科技伦理领域

非语言密度量化方法

身势语 (Kinesics): 手势、身体动作

镜头距离: 人物半身近景、环境远景

声音: 剧情内叙事、第三人称画外音

眼动追踪方法与指标



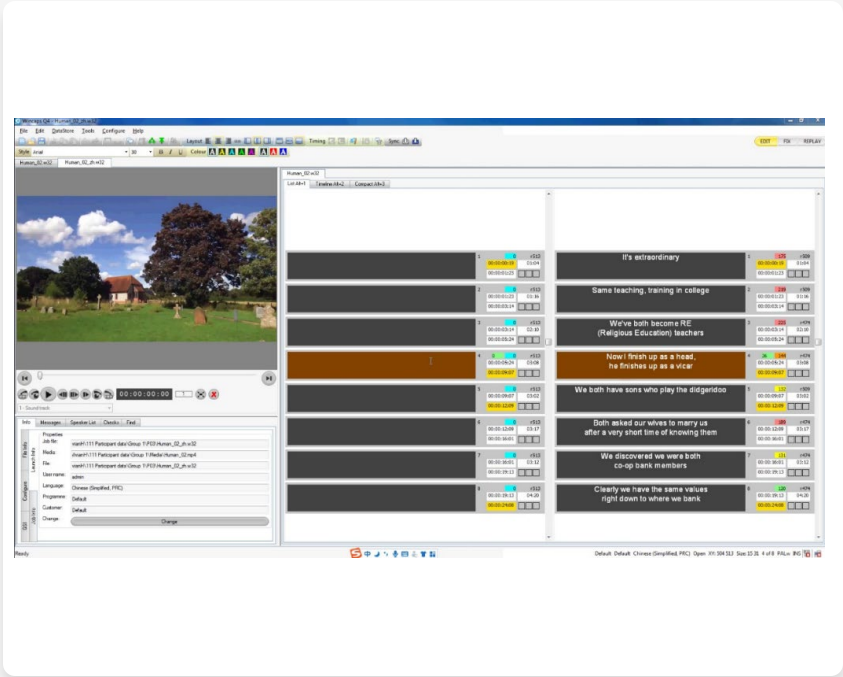
Tobii TX300 眼动仪 (300Hz)

指标	定义	计算方法
注视时长	视线停留在特定区域的总时间	$\sum \text{fixation_duration}$
注视次数	视线在特定区域停留的总次数	$\text{count}(\text{fixation})$
注意力分布	视线在源文本、目标文本、非语言区域的分布比例	$\text{area_fixation} / \text{total_fixation} \times 100\%$

筛选标准: 注视时长 $\geq 100\text{ms}$ 被记为有效注视点, 样本点中有效率高于均值2个标准差的数据纳入分析

键盘记录方法与指标

字幕翻译软件界面



键盘操作指标

指标类别	具体指标	意义
输入与删除操作	总按键次数	反映总体输入工作量
	输入次数	译文生成行为
	删除次数	译文修改行为
停顿指标	总停顿时长	认知中断的总耗时
	总停顿次数	认知中断的频率

重要参数：停顿阈值包括 300ms 和 1000ms，并比对指标结果

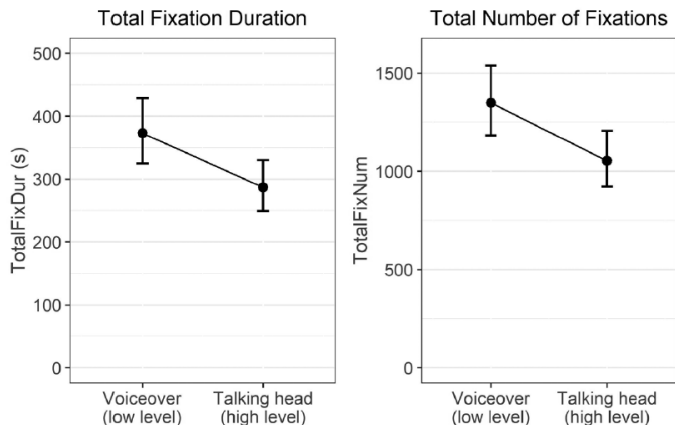
数据处理：使用 Tobii Studio 导出XML格式数据，通过R语言脚本分析处理

眼动数据分析结果

👁️ 总注视时长与次数对比

FIGURE 2

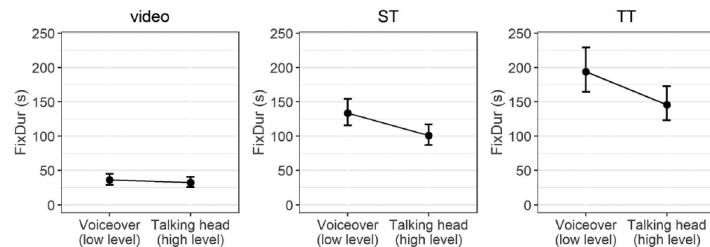
Effect plots of nonverbal information on total fixation duration and total number of fixations



🧠 注意力分配情况对比

FIGURE 3

Effect plots of nonverbal information on AOI-specific fixation duration and total number of fixations



核心发现

高非语言符号信息量条件下，总注视时长减少**22.8%**

高非语言符号信息量条件下，总注视次数减少**22%**

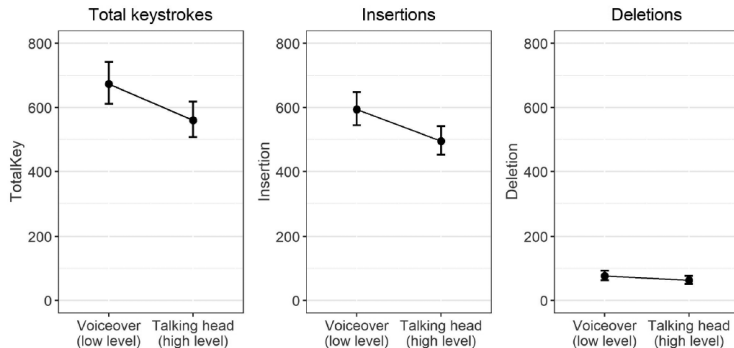
以三个兴趣区逐个对比分析，注视行为减少的来源主要是对**译文区 (24.90%)**和**原文区 (24.55%)**注视行为的减少。

键盘操作数据分析结果

总键盘操作次数、输入次数、删除次数

FIGURE 5

Effect plots of nonverbal information on keystroke operations



主要发现

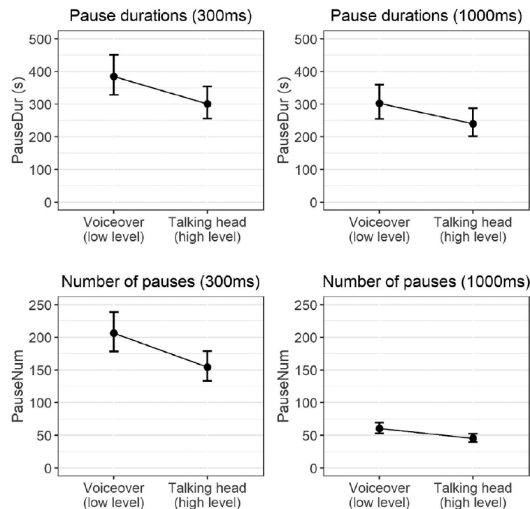
高非语言符号信息量条件下，总键盘操作次数减少**17%**。其中，输入次数减少约**17%**，删除次数减少约**18%**

高非语言符号信息量条件下，停顿时长和次数也减少约**20%**及以上（在300ms和1000ms阈值水平上结果一致）。

停顿时长、停顿次数

FIGURE 6

Effect plots of nonverbal information on pause durations and number of pauses



解释

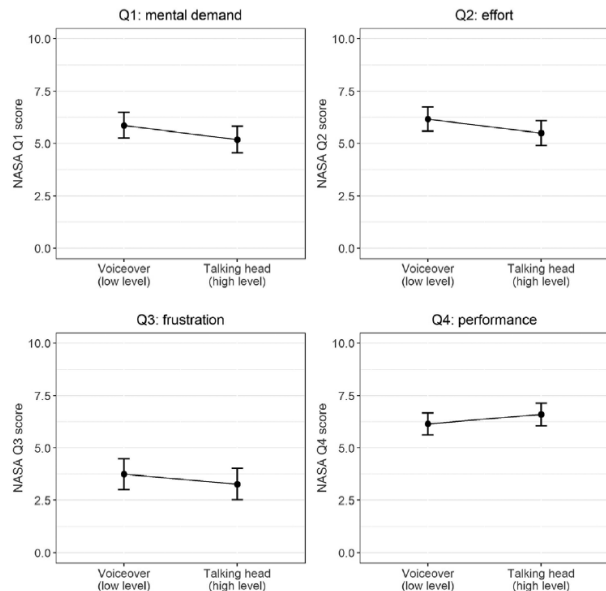
适度增加的非语言符号信息能够**降低认知负荷**，减少翻译过程中的犹豫和停顿，提高翻译效率。这与双重编码理论预测一致，表明非语言信息在翻译过程中起到辅助作用，而非干扰作用。

主观评价结果

NASA-TLX 主观负荷评估

FIGURE 7

Effect plots of nonverbal information on NASA-TLX index results



数据解读

脑力需求程度、努力程度、受挫指数、表现水平

NASA-TLX 量表表示例

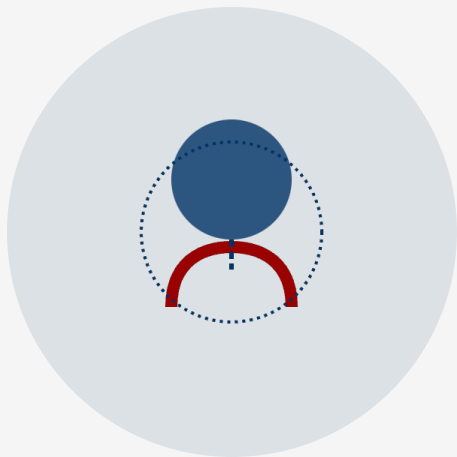
The Adapted NASA Task Load Index for Measuring Translation Difficulty (cf. Sun and Shreve 2014)

- 1. Mental Demand** How mentally demanding was the task?
0 1 2 3 4 5 6 7 8 9 10
Low High
- 2. Effort** How hard did you have to work to accomplish your level of performance?
0 1 2 3 4 5 6 7 8 9 10
Low High

译者主观反馈解读

结果表明，当原文视频中含有较多的非语言符号时，译者自评的认知负荷较低，且译文质量更优；当原文视频中的非语言符号较少时，译者认为自己的认知负荷较高，且译文表现更差。

核心发现与讨论



非语言-语言信息协同处理

- 1 适度的非语言信息**降低认知负荷**，促进翻译过程，支持双重编码理论
- 2 译者采用的是**双通道并行处理**策略，而非单一通道串行处理
- 3 但是非语言信息中，究竟是什么具体因素（比如图像、音频中的某些信息点）对于降低译者认知负荷起到了最关键的作用？这仍有待进一步研究和实验验证。

这些发现与最新多模态认知研究一致，如 Baadte and Meinhardt-Injac (2019) 和 Tardel (2021) 的研究均表明多模态信息处理能够形成相互增强的效果。

对大语言模型多模态翻译的启示



语料库构建

平衡的多模态语料库

包含多层次非语言信息密度

收集真实人类翻译过程数据



训练策略

认知启发训练

模拟人类双通道并行处理

分阶段注意力训练策略



评测方法

认知导向评测

融合人类认知负荷指标

增加多模态理解评分



架构设计

情境感知注意力

模拟人类选择性注意

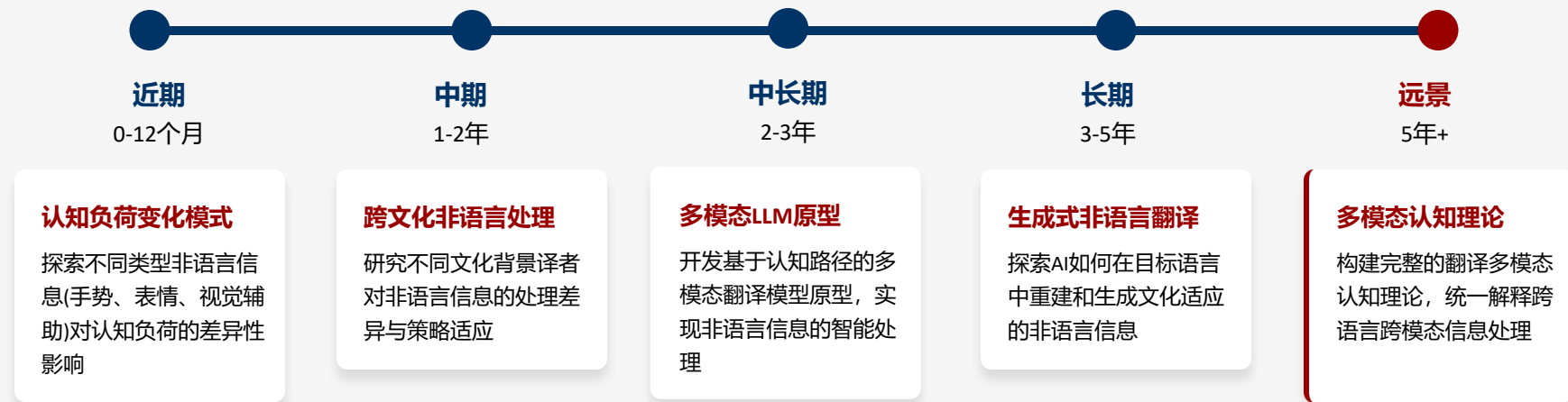
关键语义节点强化

动态适应非语言密度变化



注意力机制

未来研究方向



协同研究方向

认知神经科学

fMRI研究非语言信息处理的神经基础

计算语言学

多模态表征学习与跨模态对齐

翻译教学

基于非语言认知的翻译教学改革

感谢各位的聆听与讨论。

本研究得到对外经济贸易大学新进青年教师科研启动项目
(编号: 23QD24) 的资助, 即将出版于国际期刊 Meta, 69(3).

欢迎交流与合作!

黄婕



chezvivan@outlook.com



<http://chezvivan.github.io>