

数据驱动的初级汉语课堂优化研究： 语言使用比例可视化与生成式 AI 的结合 (A Data-Driven Approach to Optimizing Beginner Chinese Classrooms: Integrating Language-Use Proportion Visualization with Generative AI)

徐勤
(Xu, Qin)
京都大学
(Kyoto University)
xu.qin.4f@kyoto-u.ac.jp

砂冈和子
(Sunaoka, Kazuko)
早稻田大学
(Waseda University)
ksunaoka@waseda.jp

摘要：本研究以日本的大学初级汉语课堂的两段教学录音为样本，利用我们自主开发的语音可视化应用 Voice-to-Text App 进行分析。结果显示，该应用在处理中日语码混合的课堂录音时，具有较高的转写准确率与操作便利性，能迅速识别课堂中的 L2（汉语）使用比例。与传统的课堂语言行为评估框架相比，该系统在效率、可操作性与教师自主分析能力方面均表现出显著优势。在此基础上，研究将 APP 生成的数据结合两种类型的提示词（prompt），输入 ChatGPT，由生成式 AI 提供课堂改进建议，以探讨 AI 在汉语课堂设计中的潜在优势与应用局限。结果发现，AI 能够在语言形式重组与表层推理层面提出较为合理的课堂优化方案，但由于缺乏深层认知能力与创新语法概念的生成能力，难以提出具有启发性的教学设计。

Abstract: This study investigates the optimization of beginner-level Chinese language classrooms in Japan through a data-driven approach that integrates speech visualization and generative AI. Classroom recordings were analyzed using a self-developed voice-to-text app, which automatically transcribes classroom recordings of mixed Chinese-Japanese classroom discourse and then visualizes the proportion of L2 (Chinese) use. The app demonstrates high transcription accuracy and operational convenience, offering significant advantages in efficiency, usability, and teacher autonomy compared with traditional classroom language analysis frameworks. Based on the app-generated data and two types of instructional prompts, ChatGPT was used to generate feedback and suggestions for classroom improvement, in order to explore how AI could be applied in Chinese language class design and identifying its limitations. The findings reveal that while AI can effectively propose revisions at the linguistic and surface-level reasoning stages, it lacks deeper cognitive and creative capacities necessary for generating pedagogically insightful designs.

关键词: 数据驱动的汉语教学, 课堂优化, 语音转写 App, 语用比例可视化, 生成式人工智能

Keywords: Data-driven Chinese language teaching, Classroom optimization, Voice-to-Text App, Visualization of L1/L2 usage ratio, Generative AI

1. 引言

1.1 研究目的与背景

经济合作与发展组织 (OECD) 在 Learning Compass 2030 中提出以“学习者能动性 (student agency)”为核心的教育理念, 主张学习是学习者主动建构意义的过程, 而非被动接受知识。知识与技能被重新界定为“认识论知识 (epistemic knowledge)”, 即学习者能理解、反思并应用的能力 (OECD, 2019 a; OECD, 2019 b)。该框架促使教育从“知识传递”转向“能力建构”, 强调学习过程的动态性与反思性。

然而, 日本的大学第二外语教学仍主要采用以语法大纲为中心的“知识传递型教学”模式。课堂通常由教师主导, 学生缺乏使用目标语的机会。受笔试评价体系的影响, 教学往往重形式而轻意义, 第一语言使用比例偏高, 导致学习动机与成效均不理想。近年来, 日本文部科学省的多次问卷调查结果¹显示, 超过半数的大学生对外语学习成效持否定态度, 由此引发了对学习意义与教学方法的持续质疑 (文部科学省, 2020; 2022; 2023)。

本研究开发了一款语音可视化应用——Voice-to-Text App (以下简称 APP), 可自动转写课堂语音并可视化呈现 L1/L2 的使用比例, 从而帮助教师进行课堂语言使用的自我诊断与反思。此外, 将 APP 生成的结果输入生成式 AI, 由 AI 对课堂互动进行分析并提出改进建议, 使教师借助数据驱动的外部反馈, 实现持续的自我省察与教学重构。

本文第二节介绍 APP 的功能, 第三节展示课堂应用效果, 第四节探讨与 AI 结合的可行性, 第五节总结研究结果。

¹ 该调查的有效回答率不足一成, 样本代表性相对有限, 但调查对象覆盖了日本全国范围内的国公立与私立大学及短期大学的学生。由于该调查由国家教育机构组织实施, 其结果具有较高的权威性, 因而可能会对今后的外语教育政策产生一定影响。

1.2 传统课堂语言行为评估框架及其局限

为改进外语课堂教学而广泛采用的分析框架主要包括：FLINT（Foreign Language Interaction：Moskowitz, 1971）、FOCUS（Foci for Observing Communications Used in Settings：Fanselow, 1977）、和 COLT（Communicative Orientation for Language Teaching：Spada & Fröhlich, 1995）等。这些框架通过系统分析课堂语言行为，帮助外语教师进行自我诊断、改进教学策略，并为教师培训提供参考。

然而，欧美开发的课堂分析系统多以交际导向型课堂为前提（Patsy & Spada, 2021），在以读写活动为主的日本外语课堂中，其适用性仍受限制（飯野厚, 2009; 飯窪真也等, 2020）。此外，传统的课堂分析框架通常要求对整节课进行录像与转写，并依据预设类别进行人工编码。此过程不仅耗时费力，还需要一定的专业知识，因而难以实现实时分析。例如，COLT 原版包含 13 个主要类别及 40 余个子类别（Spada & Fröhlich, 1995）；FLINT 系统则以 15 个功能类别评估教师语言与反馈的教育功能（Moskowitz, 1971）。尽管这些框架结构严谨，但因分析成本过高，仍难在教师与研究群体中被广泛推广（Pellerin et al., 2024）。作为数字化应对方案，Mobile COLT 等工具相继出现（石塚博規等, 2021）。然而，此类工具仍依赖专业分析与大量时间投入，难以满足日常教学改进需求。因此，亟需开发能有效降低分析负担、使教师便捷且实时获取课堂分析结果的新型工具。

2. APP 的开发目的与历程

近年来，自动语音识别（Automatic Speech Recognition, ASR）技术的进步显著降低了语音转写的成本与时间，为缓解上述问题提供了新的契机。

Ferraro et al. (2023) 采用单词错误率（WER: word error rate²）评估了多种开源 ASR 工具（如 Mozilla DeepSpeech³、Conformer⁴、HuBERT⁵、SpeechBrain⁶、WhisperX⁷、SpeechStew⁸）与商业 ASR 服务（如 Amazon Transcribe⁹、Microsoft

² WER 衡量的是模型转录错误的单词占参考文本总单词数的百分比，详见 Ferraro et al. (2023): “[T]he metric measures the percentage of words that are incorrectly transcribed by the model relative to the total number of words in the reference transcript”。

³ 工具信息详见：<https://github.com/mozilla/DeepSpeech>。

⁴ 工具信息详见：<https://github.com/sooftware/conformer>。

⁵ 工具信息详见：<https://github.com/facebookresearch/fairseq/tree/main/examples/hubert>。

⁶ 工具信息详见：<https://speechbrain.github.io/>。

⁷ 工具信息详见：<https://github.com/m-bain/whisperX>。

⁸ 论文参考：<https://doi.org/10.48550/arXiv.2104.02133>。该工具目前尚未公开官方代码。

⁹ <https://aws.amazon.com/transcribe/>。

Azure Speech to Text¹⁰、Google Cloud Speech API¹¹、IBM Watson Speech to Text¹²) 在七个常用数据集上的语音转文本的性能, 其结果表明, 在大多数评测数据集上 (包括 LibriSpeech, CommonVoice, WSJ 和 CHiME), 开源 ASR 工具的表现优于商业 ASR 服务¹³。

商用 ASR 服务通常采用按处理分钟数计费或按需付费的定价模式, 长期或大规模使用会产生较高的经济成本, 且个性化定制能力有限。而开源 ASR 工具通常可供用户免费使用、成本低, 并支持个性化定制和本地部署。其中, 以 Whisper 为代表的开源 ASR 模型不仅支持多语种识别与翻译, 还能自动检测语种并生成带时间戳 (timestamp) 的转写文本, 在准确率和跨语言性能上已超越了大部分商用 ASR 服务。该模型以 30 秒音频片段为单位进行训练, 其长音频转写策略是通过对连续 30 秒的音频片段进行逐段转录, 并结合模型预测的时间戳信息进行窗口平移, 从而实现对长音频的高效转写 (Radford et al., 2023)。鉴于日本汉语课堂录音语料中频繁出现中日语码切换, 以及课堂语言分析的实际需求, 本研究最终选择 Whisper 作为核心语音转写模型, 并将其集成至自主开发的 APP 中。

2.1 针对多语码转换的自动语音识别模型

语码转换 (Code-switching, CS) 是指在同一段话语中交替使用两种或两种以上语言的现象 (Mustafa et al., 2022)。近年来, 在外语教育中, 学者们将学习者母语 (L1) 的语码转换 (CS) 视为教师互动的资源, 并对其有效性进行评价 (Macaro, 2009; Myers, 2002)。但日本的汉语课堂过度依赖母语 (L1), 从而导致提升 (L2) 语言运用能力的教学目标难以实现 (砂冈和子等, 2023a), 即便是在英语课堂中, 这一问题依然存在 (田崎敦子, 2006)。另外, 多语码转换仍然对识别的准确率构成挑战, 并且后续的编码环节仍存在较高的技术门槛 (砂冈和子 & 徐勤, 2023b)。

Whisper 是 OpenAI 于 2022 年 9 月发布的多语言自动语音识别 (ASR) 模型, 提供 tiny、base、small、medium、large 五种规模。模型规模越大, 识别精度越高, 计算资源与处理时间的需求也相应增加。此后, OpenAI 于 2022 年 12 月推出 large-v2, 并在 2023 年 11 月发布性能进一步提升的 large-v3。徐勤 & 砂冈和子 (2024) 将 large-v3 与 “Pyannote.audio”¹⁴ 的话者分离功能结合, 成功应用于包含中日语码转换的汉语课堂录音, 实现了文本转写与说话人区分, 为多语码课堂的语音数据分析提供了有价值的范式。

¹⁰ <https://azure.microsoft.com/en-us/products/cognitive-services/speech-to-text/>。

¹¹ <https://cloud.google.com/speech-to-text>。

¹² <https://www.ibm.com/cloud/watson-speech-to-text>。

¹³ 详见 Ferraro et al. (2023): “[O]ur analysis also highlights that open-source solutions outperform paid services for most datasets, including LibriSpeech, CommonVoice, WSJ, and CHiME.”

¹⁴ 可参考项目主页: <https://pyannote.github.io/pyannote-audio/>

2024 年 10 月, OpenAI 又发布了 large-v3 的优化版本——Whisper large-v3-turbo (以下简称“turbo”)。与 large-v3 相比, turbo 在识别精度略有下降的同时显著提升了转录速度¹⁵。在此基础上, 砂冈和子 & 徐勤 (2025) 开发了基于 Whisper 多版本 (tiny、base、small、medium、large-v2、large-v3、turbo) 的语音转写 Web 应用——APP¹⁶ (见图 1)。该应用基于 Python 与 Flask 构建, 支持音频上传、多模型选择与时间戳转写, 并集成 Matplotlib 进行可视化处理, 可自动统计并展示课堂中汉语 (L2) 与日语 (L1) 以及教师授课时偶尔出现的英语的使用比例 (见图 3, 图 4), 结果可供用户下载并自动保存为 Word 文档。

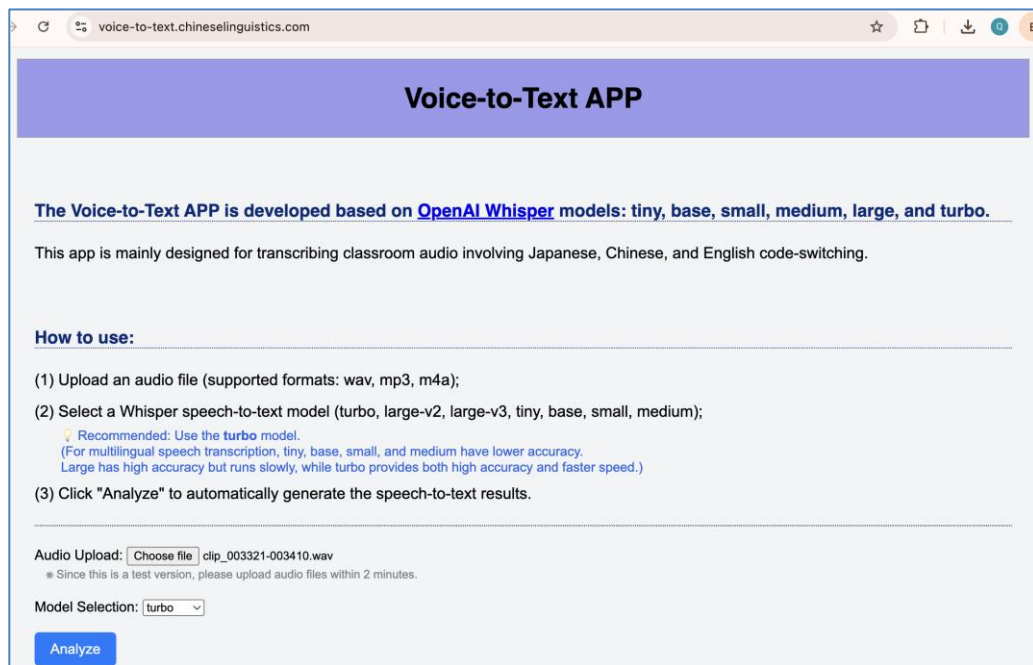


图 1 Voice-to-Text APP 的页面

3. 基于可视化分析的课堂流程优化研究

3.1. 通过 APP 实现课堂发话的可视化

与传统外语课堂观察工具相比, 该系统在自动化与效率上具有显著优势。仅需数秒即可将发话音频转写为文本, 识别精度可达 90% 以上 (徐勤 & 砂冈和子, 2024; 砂冈和子 & 徐勤, 2025)。其操作无需编程能力与专业知识, 也无需对外公开教学内容, 即可实现近实时的课堂自我诊断, 在保障隐私与降低技术门槛的前提

¹⁵ 可参考 large-v3-turbo 的测评结果: https://medium.com/@bnjmn_marie/whisper-large-v3-turbo-as-good-as-large-v2-but-6x-faster-97f0803fa933。

¹⁶ 该 APP 目前仍处于测试阶段, 目前公开测试版可通过以下网址访问: <https://voice-to-text.chineselinguistics.com/>。后续功能及界面尚有进一步调整与完善的可能。

下，减轻教师与学生的认知负荷，并支持教师开展自主、持续且常态化的课堂分析与教学改进。

若依托 FLINT 系统把时长 60 分钟的课程进行分析，则需要一名具备中日双语能力的研究助理耗时约 4 个月、累计工作时间超过 300 小时（曲明&砂岡和子，2024）。图 2 展示了基于 FLINT 分类框架，对 2022 年 1 月 11 日于日本某大学开设的一年级初修汉语课程（以下简称[2022 课堂]）的片段进行人工分析的结果（原始音频 0:34:10–0:37:07，约 3 分钟）。尽管人工转写在精确度上或许优于自动转写，但其所需时间至少为自动转写的二十倍，分析成本极高。相比之下，本研究开发的 APP 无需外部协助，教师即可在课堂结束后即时生成并查看分析结果，在效率、可操作性及教师自主性方面均具有显著优势。此外，APP 具备自动区分母语（L1）与目标语（L2）的功能，其性能明显优于 FLINT 和 COLT 系统。

開始	結束	経過時間	時長	行為分類	發言人	參與方式	轉寫文本
0:34:10	0:36:36	0:02:26	146	321	教師	線上+線下	はい、これは理解しやすいところだと思いますけれども、「誰々がどこにいる」「何がどこにある」という文ですので、主語は人間、もしくは物になりますよね。それから、後ろは在です。で、これから後ろは場所がきます、場所は地名のときもあれば、名詞が場所として使われることもあります。@@@とか@@@とかは地名ですよね、(中略)名詞は場所として使うとき、先も簡単に説明しましたんですけども、皇子は本来では、机で、物の名前ですけども、これは場所として使われる時は、上もしくは裡どれか一つね、(中略)中国語は、書くときは、名詞が場所として使われる時、この名詞の後ろに上もしくは裡をつける必要があります。ここは覚えてくださいね。で、これを、復習する、復習というか、これさらに確認するために、一番下に、「名詞の場所化」と言うところがあるので、こういうように使えますよ。フレーズ、名詞と裡、名詞と上、一緒に使う例があげられています。
0:36:36	0:36:42	0:00:06	6	311	教師	線上+線下	では、じゃあ、一人一個づつね。まず、「冷蔵庫の中」を読んでください。
0:36:42	0:36:45	0:00:03	3	310	教師	線上+線下	山本 大郎！
0:36:45	0:36:48	0:00:03	3	314	教師	線上+線下	沉默 (老師等待學生回覆的時間)
0:36:48	0:36:52	0:00:04	4	310	教師	線上+線下	山本 さんいますか。いないようで
0:36:52	0:36:55	0:00:03	3	310	教師	線上	いないようで、鈴木和夫。
0:36:55	0:36:58	0:00:03	3	330	學生*	線上	はい。
0:36:57	0:37:00	0:00:03	3	310	教師	線上	山本さん、山本さんね
0:37:00	0:37:03	0:00:03	3	311	教師	線上	はい、「冷蔵庫の中」を読んで欲しい。
0:37:03	0:37:06	0:00:03	3	330	學生**	線上	冰箱裡。
0:37:06	0:37:07	0:00:01	1	313	教師	線上	非常 好ね、冰箱裡ね。

图 2 FLINT 分析例¹⁷（部分）

下文选取[2022 课堂]中的一个发话片段（实例 1，简称[2022 课堂]），以及另一所日本的大学汉语课堂发话片段（实例 2，简称[2025 课堂]），将二者导入 APP 进行分析¹⁸。

3.2. 实例 1：[2022 课堂]发话内容的语音转写

实例 1 为[2022 课堂]中，教师总结当日语法重点的片段。APP 将原始音频（00:52:01-00:52:35，共 34 秒）按时间戳（如图 3 Transcription with timestamps）

¹⁷ “行为分类”系指事先设定的发话类别。[321] 等编号为分析者对各类别所赋予的编号。
“参与方式”：由于之一堂课为混合授课形式，故区分为“线上（远程参与）”与“线下（课堂参与）”。

¹⁸ 截至 2025 年 11 月 5 日，目前的 APP 测试版每次可处理的音频文件上限约为 2 分钟，因此在分析前需对原始录音进行剪辑。

自动转写为文字，并展示课堂中的语种使用比例（Language Ratios）。ASR 技术将课堂中的发话细节真实地转写为文字。虽然极短的语气词有时可能被忽略，但整体上几乎所有语音内容都得到了准确转写。

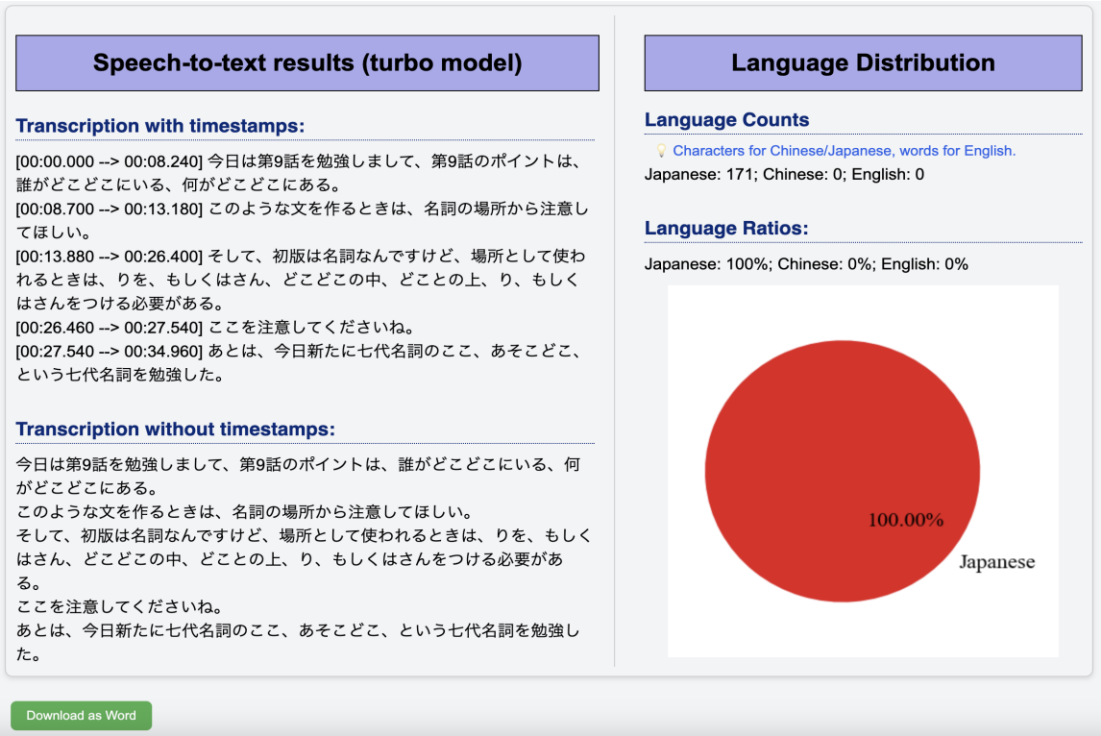


图 3 实例 1[2022 课堂]转写部分（带时间戳）

表 1 实例 1 的人工校对（不带时间戳）

今日は第 9 課を勉強しまして、ええと、第 9 課のポイントは、ええ、誰々がどこどこにいる、何々がどこどこにある。このような文を作るときは、名詞の場所化を注意してほしい。そして、书包は名詞なんですけど、場所として使われるときは、里を、もしくは上、どこどこの中、どこどこの上、里、もしくは上をつける必要がある。ここを注意してくださいね。あとは、ええ、今日新たに指示代名詞のここ、あそこどここという指示代名詞を勉強した。

注：带下划线的文字为 APP 转写错误或遗漏部分。灰色底纹的黑体字为中文（下同）。

表 2 实例 1 文本的中译

今天我们学习了第 9 课。嗯，这一课的重点是，嗯，表示“谁在某处”“什么东西在某处”的句型。造这种句子时，要注意名词的“场所化”。比如，“书包”是名词，当它表示场所时，需要加上“里”或“上”，“在某处里”“在某处上”。要加上“里”或“上”。请大家注意这一点。另外，嗯，今天还学习了新的指示代词——“这里”“那里”“哪里”等用法。

结果显示，该段授课语言几乎全部为日语。APP 检测的语种比例为：日语（L1）100%，汉语（L2）及英语均为 0%（见图 3）。实际上，该片段中包含 5 个

汉语词语（见表 1 与表 2 中带底纹的黑体字），真实的汉语（L2）使用比例约为 3%。这是因为 Whisper 在处理句内语码转换（code-switching）时的识别效果不理想，常将句中嵌入的外语误判为整句的主要语种，从而导致 L2 使用比例被低估。例如，在表 1 中，日语语境中的汉语词语“そして、书包は名詞なんですけど”被误识为“そして、初版は名詞なんですけど”（文中下划线部分为笔者标示的识别错误，下同）；汉语“里”“上”分别被转写为日语读音相近的平假名“り”“さん”。此外，即便是日语部分，涉及专业术语时也容易出错，如“場所化”“指示代名詞”分别被误识为“場所から”“七代名詞”。

语码转换不仅可能导致机器识别错误，也可能对人类的听辨与理解造成干扰（Xiao & Park, 2021; Amrate & Tsai, 2025）。因此，外语教师在课堂中进行语码转换时应有策略，同时应尽量避免使用过多高难度术语，可将其转化为更易理解的表达，以确保学生能够真正听懂并理解课堂的内容。虽然 APP 的语音识别存在一定误差，但能即时将语种比例可视化，这为教师检查语言使用比例、调整语码转换策略提供了依据。APP 自动生成的时间戳不仅呈现语料的时间分布，还可用于分析教师发话的节奏与密度。例如，在实例 1 中，单句话语的最长持续时间约 13 秒，最短为 1.08 秒（图 3），这表明教师发话单位较短、语速较快，整体节奏紧凑。在此之前，教师已就“名词的处所化”进行了约 146 秒的讲解（见图 2 第一行），随后点名 9 名学生依次朗读课文例句并检查发音（图 3 展示了其中两名学生的互动）。其中 3 名学生未作答，实际仅 6 名学生各朗读一句，平均每句约 20 秒，其余学生仅处于聆听状态，缺乏发言机会，也失去了与教师进行意义协商的契机。[2022 课堂]的其他片段亦呈现类似情况。

总体来看，该课堂以教师讲解为主，虽有师生问答环节，但多停留在知识再现层面，缺乏交际性互动（图 2）。FLINT 分析结果进一步验证了这一倾向（曲明 & 砂岡和子，2024）：

（A）教学主导性强——教师发话次数占总量的 58%，时长占 67%，均高于学生的 42% 与 33%。

（B）缺乏支持学生主体性的发话——直接性发话（如指示、说明、订正）显著多于间接性发话（如提问、称赞、鼓励、重述改正），在次数和时长上分别高出约 28% 与 50%。

值得注意的是，无论是 FLINT 还是 APP，均只能在定量层面揭示课堂结构性问题，尚无法提供如何重构互动或激发学生能动性的具体策略。即便如此，教师仍可依据这些分析结果回顾自身教学行为，检查课堂互动状况，并据此调整策略，以促进教学改进与质量提升。

3.3. 实例 2: [2025 课堂]发话内容的语音转写

实例 2 取自日本的一位大学汉语教师（母语为汉语）于 2025 年 9 月 27 日为初级学习者进行的约 15 分钟模拟课堂片段，主题为讲解“是……的”句式。APP 自动转写了其中约 1 分 32 秒的片段，结果显示课堂语种比例为：L2（中文）42.18%，L1（日语）57.45%，英语 0.36%（见图 4）。整体来看，[2025 课堂] 的教师在讲解中注重意义协商，并设计了使学生能够在真实语境中操练的语言活动（见表 3），因此目标语（汉语）的使用比例相对较高。尽管模拟课堂中没有学生互动环节，但讲解流畅、结构清晰。然而，教学内容仍主要停留在“是……的”句式的知识再现阶段，缺乏引导学生将语法形式与交际情境相结合的教学设计。

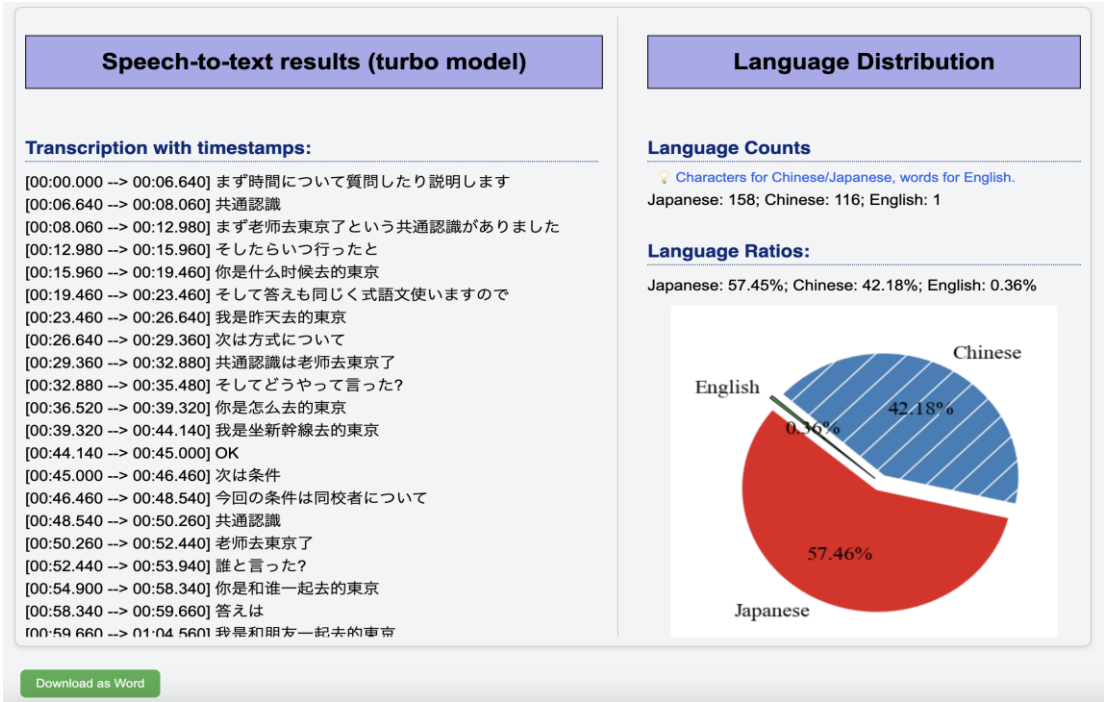


图 4 实例 2 [2025 课堂]（带时间戳）

表 3 实例 2 的人工校对（不带时间戳）

まず時間について質問したり説明します。共通認識。まず老师去東京了という共通認識がありました。そしたらいつ行ったと。——你是什么时候去的東京。でそして答えも同じく式的構文使いますので。——我是昨天去的東京。次は方式について。共通認識は老师去東京了。でそしてどうやって言った？——你是怎么去的東京。私は坐新幹線去的東京。OK。次は条件。今回の条件は同校者について。共通認識。——老师去東京了。誰と言った？——你是和谁一起去的東京。答えは我是和朋友一起去的東京。で次は場所について。共通認識は。——老师買了東京香蕉。先ほど皆さんの教室で。お土産ありますよって。——老师買了東京香蕉。じゃあどこで買った？。——東京香蕉是在哪買的？東京駅ですね。——東京香蕉是在東京站买的。

表 4 实例 2 文本的中译

首先,我来提问或说明“时间”这一部分的内容。这里有一个共通认知:老师去了东京。于是我们可以问——“什么时候去的?”——你是什么时候去的东京?那么,回答时同样使用“是……的”句式,比如:我是昨天去的东京。接下来是“方式”。共通认知仍然是“老师去了东京”。那么,可以问“怎么去的?”——你是什么时候去的东京?回答:我是坐新干线去的东京。然后是“条件”。这次的条件是“同行者”。共通认知仍是“老师去了东京”。于是可以问“和谁去的?”——你是和谁一起去的东京?回答:我是和朋友一起去的东京。那么,接下来是关于“地点”的部分。共同认识是:老师买了东京香蕉。刚才在大家的教室里,我说“有礼物啊”,——老师买了东京香蕉。那,在哪里买的呢?东京香蕉是在东京哪里买的?——在东京站。——东京香蕉是在东京站买的。

对于以日语为母语的学习者而言,“是……的”句式以及实例 1 中出现的“处所名词”等结构,均属于新的语法概念(王亚新, 2021; Pan & Liu, 2023)。若仅依赖形式性讲解与“共通认知”“处所化”等抽象术语,难以实现“可理解性输入(comprehensible input)”。部分以中文为母语的教师未能充分利用日语这一学习者的母语资源进行对比,导致学习者难以从语境中理解并认同“汉语母语者为何常用这些句式”,从而缺乏意义建构的动机。

这种依循教材进度、以知识传授为中心的授课模式,在其他课堂中亦屡见不鲜。如何避免陷入此类“惯性化教学”的盲点,仍有赖于教师的省察与反思能力(白水始等, 2021)。目前,APP 虽能在定量层面揭示教学问题,但尚无法提供具体的课堂改进策略。为弥补这一不足,本研究进一步尝试借助生成式人工智能,自动生成课程优化建议,以探讨其在汉语课堂设计中的潜在优势与应用局限。

4. 生成式 AI 辅助课堂改进的效能评估

生成式人工智能通过分析海量文本的规律,基于上下文预测最合适的下一个词语或表达,从而生成回答。因此,许多面向第二语言习得的语言学习应用程序及提示词设计,均依托 AI 算法的优势,涵盖了翻译与摘要、误用检测、文本补全、习得难度预测等多种功能(Shan et al., 2024; 连维琛等, 2024)。在本研究中,我们对 ChatGPT 的期待并不限于这些表层的语言输出功能,而是希望其能够提供关于教学方法与课堂设计的启发性建议。

已有研究指出,明确且精确的提示(prompts)能够显著提升 AI 输出的针对性与质量(Poole & Coss, 2024)。然而,要实现课堂改进指令的明确化,教师需对自身的教学过程具有深入的理解、评估与调适能力,并能够将这种元认知觉察进行语言化与概念化(Mizumoto, 2023)。本研究利用 AI 的目的,主要在于协助非语言教育专业背景的教师反思并诊断其教学实践。考虑到部分教师难以自行提出教学改进的关键词,我们设计了两种类型的提示词:(A)开放式提示与(B)结构化提示。(A)类提示词根据前述 FLINT [2022 课堂] 分析中揭示的两大问题——即“如

何促进学生更具主体性、积极参与与互动”——进行探索性设计。(B)类提示词则要求根据具体条件(据情况可增减条件项)提出相应的教学改进方案。两类提示词均要求提升课堂中的汉语使用比例,具体提示内容见表5。带波浪下划线的部分为作者有意强调的重点,“//”符号之后所附的课堂文本为未经修改、未校正的 APP 自动转写原文,便于授课教师直接使用。正如后文所示, AI 已对其中的转写错误部分进行了基本修正。

表 5 提示类型与具体提示示例¹⁹

提示类型	具体提示例	生成结果号码
(A) 开放式提示	以下是一位汉语教师在课堂上讲解(如处所词组/“是……的”句式;根据本课内容填写)语法点时的一段课堂发话文本。 <u>将其改写为能够促使学生发挥主体性并积极参与互动的课堂形式。同时将课堂中汉语的使用比例控制在 30%至 40%之间。</u> //(此处附上该课堂的 APP 转写文本)	[实例 1 (A) - 1] [2022 课堂]
		[实例 2 (A) - 2] [2025 课堂]
(B) 结构化提示	以下是一位汉语教师在课堂上讲解(如处所词组/“是…的”句;根据本课内容填写)语法点时的一段课堂发话文本。 <u>依据下列四项条件提出教学改进方案:</u> 1)引导学生通过母语(日语)的比较来深化汉语语法理解;2)避免使用抽象术语,确保输入内容易于理解;3)鉴于学生处于初级水平,课堂练习宜以简短回答或选项反应为主;4)将课堂中汉语的使用比例控制在 30%至 40%之间。//(此处附上该课堂的 APP 转写文本)	[实例 1 (B) - 1] [2022 课堂]
		[实例 2 (B) - 2] [2025 课堂]

(A)类提示要求 AI 分别将 [2022 课堂] 与 [2025 课堂] 改写为师生互动更频繁、汉语使用率更高的课堂脚本,整体难度较低。(B)类提示较(A)类多出三个条件,内容更具体、更严密。以下将展示在(A)与(B)两种不同提示词条件下所得回应的主要特点,并对比两种提示条件下的输出结果,以检验 AI 生成建议的适切性与启发性。鉴于生成式 AI 在多次运行中可能产生不同的输出结果,本文选取了 ChatGPT 5.0 于 2025 年 9 月 15 日至 10 月 26 日期间生成的数十个版本中的若干示例进行呈现与分析。本次提示词与回答均以汉语进行;若提示词改为日语或英语等其他语言, ChatGPT 亦可生成结构基本一致的结果。

¹⁹ 尽管指令中要求中文使用比例保持在整体约三成。因 GPT 无法进行精确的数值计算,有时会超过三成。

4.1 AI 生成能力评估

总体而言,无论是(A)类提示还是(B)类提示,AI在改进课堂设计方面均展现出一定优势,不仅能提供多维度的互动模式,还能够提出较为适切的语言使用比例方案。相较之下,AI在语法讲解的生成方面仍显不足,往往仅对输入的提示内容进行表层重组,或提供来自网络检索的简短句例。这些句子既缺乏真实交际语境,也难以体现语法功能的深层关系(陶红印,2025)。AI在教学改进建议中未能展现出具有启发性的教学构想,也难以生成新的潜在概念,甚至偶有错误出现。下面将分别介绍并分析AI方案的检验结果。以波形框标示的ChatGPT方案,除对前后说明部分及个别换行位置进行了适度调整外,其余内容均与原文保持一致。

4.1.1 AI 辅助课堂设计的效能

经过多次检验后,从下列[实例1(A)-1]【A】【B】与[实例2(A)-2]【C】的结果可见,ChatGPT除语法讲解部分外,均能持续生成符合提示要求的方案。具体而言,【A】通过看图说话、全班齐读,【B】以小游戏活动,【C】则以伙伴互问等方式,均被改写为增强课堂互动的教学脚本。与原课堂几乎完全由教师单向讲授的情况相比,这些改写有效提升了学生的课堂参与度与教学效果。值得注意的是,教学改进示意均为ChatGPT自动生成的描述,可见AI具备提出课堂改进方案的能力。

【A】[实例1(A)-1]通过看图说话、问答、全班齐读等方式,增强了师生间的互动。

教师:大家好!今天我们学习第九课。今天的重点是——“谁在哪儿”“什么在哪儿”。(黑板书写:谁在哪儿?什么在哪儿?)请看图片。(展示教室图片)这是谁?——对,是老师。老师在哪儿?(指向学生)小李,你说说看。

学生A:老师在教室里。

教师:很好!“老师在教室里。”——对不对?(全班回应)那,“书”呢?“书在哪儿?”

学生B:书在桌子上。

教师:非常好!大家一起来说:“书在桌子上。”

(全班齐读)

【B】[实例1(A)-1]以小游戏提高学生的反应速度,增强课堂互动。

【练习活动】

教师:现在我们玩一个小游戏——“找一找”。

我说一个东西,你们说它在哪儿。比如——“铅笔”在哪儿?

(学生环视图片或实物,回答)

学生C:铅笔在书包里!

教师:很好!那,“手机”呢?

学生 D: 手机在老师的手上! 😄
教师: 哈哈, 对! 非常好!

【C】[实例 2 (A) -2]用图片提示、学生间互问等形式, 提升学生的课堂参与度。

教师:现在换一个问题。还是“老师去东京了”。那, 怎么去的? (提示用图片或实物) 大家问我吧!
学生:老师, 您是怎么去的东京?
教师:很好! 我是坐新干线去的东京。那你呢?
(学生间互问) “你是怎么去的东京?”——“我是坐飞机去的。”
(教师小结) 👉“是……的”也可以强调方式。

4.1.2 AI 能把握语言使用比例

无论是 (A) 类还是 (B) 类提示, ChatGPT 从生成的教学示例均将汉语使用比例控制在 35~40% 之间, 且兼顾了可理解性与实践性。可见, AI 在设计外语教学方案时, 能够较好地掌握语言使用比例的适切性, 并能主动提供教师课堂常用语。尽管部分语句 (如“谁在桌子上?”参见【D】) 并非常用句, 但在维持汉语使用比例、不增加课堂难度方面, 仍具有实用价值。

【D】[实例 1 (B) -1] 改进要点: 教师讲解或规则说明使用日语 (60~70%); 教师的提问、提示和练习指令使用汉语 (30~40%); 学生的回答以汉语短句为主。语言分配示例如下:

教学环节	教师主要语言	汉语比率
导入・比较	日语+汉语例句	約 30%
示范讲解	汉语	約 40%
练习活动	教师问答・学生短答 (汉语)	約 40%
总结・回顾	日语	約 30%
教师用的固定句 (保持汉语比率而不增加难度): “请看——”“谁在桌子上?”“对, 不错!”“再说一遍。”“很好, 大家一起说——”		

4.2 语法认知生成的缺位

如上所见, 在数据驱动的条件下, 生成式 AI 在面向初级水平学习者的课堂设计与语言使用比例控制方面展现出一一定优势。然而, 其在语法教学生成方面仍存在局限, 表现出认知生成的缺位。AI 所生成的汉语句式往往刻板且简短, 常缺乏真实交际语境, 难以体现语法功能的深层关系。有时还生成脱离教学语境的内容, 导

致句子显得突兀、不自然,甚至偶尔出现判断错误。下面【E】方案中的“谁在哪儿?”、“学生在操场”、“猫在椅子上”等句子缺少谓语成分,名词也多未搭配量词,显得不够自然。应用“(量词)+谁/什么+在哪儿+做什么”的句式,如“(两个)学生在操场上(跑步)”、“(一只)猫在椅子上(睡觉)”等表达更为自然。此外,【F】方案中,教师以“这句话对吗?”来让初级学生判断句式正误,语气略显突兀,不够自然。在“是……的”结构中,除强调主语外,“是”在口语或非正式语境中常被省略,因此AI无需刻意纠正“这句话还少一点儿‘是’”之类的表达。下面所举的方案中,笔者以**斜体粗字体**加*号标注的部分,表示包含应答方式在内,均为不自然或错误的汉语表达(下同)。

【E】[实例1(A)-2]处所词组的教学改进方案

【练习一: **谁在哪儿?**】

教师:请大家看第二张图。(图片: * **学生在操场**, **猫在椅子上**)我问——谁在操场?

(指名) 学生A: * **学生在操场**。

教师:很好!那猫呢?

(学生回答) 学生B: * **猫在椅子上**。

教师:好!大家一起说一遍。

(全班齐读)

【F】[实例2(A)-2]「是……的」句的教学改进方案

① 导入(情境设定)

教师(日语+少量中文):今天我们来学习一个非常有意思的句型——“是……的”。比如:“老师去东京了”。* **这句话对吗?**

(学生: **对!**)

可是,如果我想问“什么时候去的?”,该怎么说?

学生:(思考后)“你什么时候去的东京?”

教师:很好!* **但是**——(微笑) **这句话还少一点儿“是”**。听我说:“你是什么时候去的东京?”大家一起来说一遍。

学生(齐读):你是什么时候去的东京?

另外,尤其是在(B)类提示的四项条件中,AI难以有效实现“引导学生通过日汉比较加深语法理解”与“避免使用抽象术语以确保可理解输入”这两项教学要求。其生成脚本多停留在语言形式与表层推理层面,缺乏教学构想的启发性,未能形成新的教学概念(参见【G】-【I】)。理想的回答不仅应引导学生注意母语与目标语的差异,还应超越单纯的比喻式说明,能够以语言的形式阐明“日汉表达概念的差异”。例如,汉语一般名词不同于日语,不具“场所性”,因此需加方位词才能表示地点;又如,在讲解“是……的”句式时,AI应进一步说明其与汉语时态概念的关联,从而帮助学生理解“是……的”是汉语中常用且必须掌握的句式。

ChatGPT 在语法认知生成方面的不足, 主要源于其仅依赖输入文本(如 APP 转写语料)进行表层重组与检索式生成, 缺乏深层语法认知与教学构想能力, 因而难以提出具有启发性的教学方案。另外, AI 的优势主要体现在编写适合不同水平班级的例句、提供带拼音的词汇表、制作音频与图像教材以及设计测验等方面, 从而为课堂改进提供有力支持(胡文菊, 2021; Zhang, 2021)。鉴于篇幅所限, 其他功能不再赘述。

【G】[实例 1 (B) -1]以中日对照, 促进学生理解处所词组的差异。

改进要点: 首先呈现中日对照, 让学生“自己发现”句式结构的差异。

例: 汉语: 老师在教室里。

日语: 先生は教室にいます。

→ 教师提问: “日语中只用‘に’, 但汉语中说‘在……里’, 为什么呢?”

让学生自己发现汉语中需要“地点+方位词”的结构。

意图: 不是让学生死记语法形式, 而是通过对母语与目标语差异的比较, 引导学生从意义层面加深理解。

【H】[实例 2 (B) -2] 教学策略: 教师先呈现中日对照句, 帮助学生发现结构差异。

例如: 汉语: “你是什么时候去的东京?”

日语: “あなたはいつ東京に行きましたか?”

教师引导学生观察:

👉 日语中只用“行きました”就能表示过去,

👉 而汉语则使用“是……的”来表达“过去+强调”的意义。

活动设计: 在黑板上并列书写两种语言, 教师提问学生: “哪里不同?” “感觉有什么差别?” 引导学生以意义为焦点进行比较, 而非仅关注形式。

目的: 让学生以母语结构为线索, 理解汉语表达的独特性(即焦点化功能)。

【I】[实例 1 (B) -2] 不要频繁使用“名词+助词”或“方位词”等抽象术语, 而应借助教室中常见的图片、实物和动作来帮助学生理解。

教师可以在黑板上书写并配合图示:

- 老师(人)在教室里
- 猫(东西)在桌子上
- 书(东西)在包里

→ 一边指着图示, 一边提问学生: “谁? 在哪儿?” 以此引导学生主动发话。

比喻性说明示例: “‘在’是告诉位置的魔法词, ‘里·上·下’就像指示地方的小手指”。通过这种形象化的说明来避免使用抽象术语。

5. 总结

5.1 研究结论

本研究围绕自动转写与生成式 AI 在汉语课堂中的应用, 构建并验证了一套集音频转写、L1 / L2 使用比例可视化与课堂诊断功能于一体的 APP。实证结果表明, 该 APP 在处理中日语码混合的课堂录音时, 具有较高的转写准确率与操作便利性, 显著降低了人工分析成本, 为教师即时掌握课堂语言分布、优化教学策略提供了量化依据。

在案例分析中, 教师可借助 APP 快速了解课堂中 L2 使用比例及师生的互动情况。当 APP 结果页面显示 L2 使用偏低或师生互动不足时, 教师可结合生成式 AI, 将量化结果与课堂情境相结合, 生成针对性更强的教学改进方案, 从而推动数据驱动的课堂优化实践。该机制不仅为教师能动性的培养提供了新路径, 也为语言教育研究带来了数据化与智能化的新方向。即便缺乏专业分析背景的教师, 也能依据系统提供的数据做出基本的教学优化判断。

然而, 如第四章所述, ChatGPT 仍受限于创新语法概念的生成。其“智能”缺乏元认知能力的原因, 一方面与大型语言模型 (LLMs) 的架构设计有关, 另一方面则在于 LLMs 尚未能有效吸收语法研究的最新成果, 尤其是汉语本体研究与第二语言习得 (SLA) 研究之间的衔接不足, 导致 AI 缺乏形成创新语法概念的优质知识资源。鉴于人工“智能”尚未成熟, 教师应在教学情境中保持主体性, 通过自主决策、反思与创新教学设计, 引导课堂改进。与此同时, 积极关注 SLA 研究的最新进展, 在努力提升 L2 使用比例的同时, 充分利用学生的母语资源, 在当前的教学实践中尤为关键。

5.2 研究局限与未来展望

首先, 在方法层面, 本文以实例 1 与实例 2 的 APP 输出结果为基础, 结合预设提示词输入至 ChatGPT 生成课堂改进草案, 此过程主要基于假设设定而展开。未来研究有必要将该工具实际引入课堂场景, 系统收集并分析教师与学习者的反思性数据, 以开展更具实证意义的验证。

其次, 在技术层面, 如第 3.2 节案例分析所揭示, APP 在界面交互、功能扩展以及多语种混合语料的识别精度方面仍有不足。目前尚不支持视频直接上传与转写, 也缺乏对不同发言人的区分功能。在语言识别机制上, 基于字符范围的归属判定方法易高估汉语比例, 尤其在日语文本中大量使用汉字时, 可能导致偏差。

最后, 在应用层面, 尽管该工具已公开提供使用, 但实际用户规模仍然有限, 且尚未实现与生成式 AI 的自动联动功能, 这在一定程度上限制了其推广与应用。

为评估该 APP 的功能实用性并明确后续改进方向, 我们邀请 5 名教师²⁰在试用该 APP 后填写问卷。因篇幅所限, 以下仅呈现用户反馈的关于功能便利性、不便之处以及改进建议的反馈。

在功能便利性方面, 用户普遍认可该 APP 在转录中日语码混合音频转录中高便利性(如可直接上传音频进行分析、可视化 L1/L2 使用比例)与较高的准确性, 认为其在节省人工处理时间、提高课堂教学分析效率方面具有一定的应用价值。

用户反馈的功能不便主要集中在五个方面: (1) 交互界面的设计相对简单; (2) 对 APP 内可选语音识别模型的说明不足; (3) 无法直接上传视频文件进行语音转写; (4) 转写结果未区分不同的发言人; (5) 英日语码混合场景下的语音识别精度偏低。

用户提出的改进建议主要包括三个方面: (1) 扩展相关功能: 如支持视频文件直接上传与转写、转写结果标注不同发言人, 以及支持简体/繁体中文切换等; (2) 增强界面交互体验: 包括统一界面语言风格、在模型选择界面增加说明及模型推荐提示、丰富网页布局设计及转写结果的呈现方式; (3) 优化多语种混合(如英日混合)以及句间语码转换场景下的语音识别精度。

结合用户的反馈和建议, 以及本研究开发该 APP 的主要目的与当前存在的主要问题, 未来研究拟从以下四个方面进行优化, 以提升 APP 的适用性与推广价值。

(1) 提升响应速度, 在保证识别精度的前提下缩短模型加载与转写时间。目前, 模型加载耗时较长, 语音转写的响应速度较慢。Whisper 模型本身的参数量较大, 尽管系统通过预加载机制 `whisper.load_model` 避免每次请求时重复加载模型的问题, 但整体转写过程仍存在延迟, 尤其是使用 `large-v2` 或 `large-v3` 这两种大模型时, 加载和转写时间明显增加。为兼顾识别精度与速度, 建议用户选 `turbo` 模型进行语音转写。

(2) 集成说话人分离功能, 结合 Whisper 与 “`pyannote.audio`”, 实现对课堂互动的细粒度分析。徐勤 & 砂冈和子 (2024) 通过 Whisper `large-v3` 与 “`pyannote.audio`” 的组合实现语音转写和说话人分离: 即由 Whisper 生成带有时间戳标注的初始文本, 再由 “`pyannote.audio`” 对音频进行话人分离。未来计划集成该功能, 以拓展“谁在说什么”的细粒度口语语料分析需求。今后将在 APP 中集成该功能, 在语音转写后的结果中区分不同的说话人。

(3) 优化汉日混合语料的语言归属判定算法, 减少比例偏差(见表 3 与表 4), 提升中日汉字识别与语种归属识别的精度。当前版本的 APP 主要通过字符范围进

²⁰ 参与本研究的 5 位教师均为女性, 均从事语言教育工作。其中, 4 位分别在日本的不同大学教授汉语或英语, 1 位在中国的高中教授日语。她们的教龄分布为: 5 年以下 3 位, 6-10 年 1 位, 20 年以上 1 位。除 1 位教师提供了口语考试录音作为研究材料外, 其余均提供了本人实际课堂的教学录音用于分析。

行语言归属划分, 其中汉字被自动归类为汉语, 平假名、片假名被归类为日语。这会导致语言判断偏差, 尤其在处理日语中存在大量汉字的文本时, 汉语的比例易被高估。未来计划进一步优化汉日混合语料文本的语言归属判定, 以提升语言识别的准确率。此外, 进一步提升系统在英日混合及句间语码转换场景下的识别准确性, 也将有助于其在更多语言环境中的推广应用。

(4) 扩展支持用户自定义分析时间段的功能, 便于片段式语料研究。当前 APP 默认转写整个音频。未来计划在 Web 界面添加自定义时间区间的选项, 使用户在上传音频后可自行指定音频需要分析的起止时间段, 以拓展语言研究中常见的片段式语料分析需求。

如前所述(详见 1.1 章节), 文科省的调查结果显示, 日本的大学外语学习者普遍对自身的语言掌握程度感到不满意, 这在一定程度上表明教师可能未能准确把握学生的实际理解水平。传统上, 教师多依赖“客观测试”来检验学生的理解程度, 但若测试仅限于词汇和语法的辨别性能力, 生成式 AI 的表现往往优于学生 (Mizumoto, 2023)。在技术快速发展的背景下, 语言教学应更加注重认知能力的培养 (He & Lin, 2021)。因此, 外语教师更应主动反思日常教学实践, 改进教学设计, 以确保提升学生的真实理解水平与学习成效。本研究开发的 APP, 旨在为教师提供发现课堂问题的反思契机, 从而支持持续的教学改进与优化。

致谢: 本研究得到日本学术振兴会科学研究基金 (JSPS KAKENHI: 课题编号 24K04091、24K16129) 资助。在此谨致谢忱。本文系根据2024年6月22日于TCLT12 (The International Conference and Workshops on Technology and Chinese Language Teaching) 上由砂冈和子与徐勤共同发表之报告〈多语码汉语教学课堂中的话者分离与文本转录——Whisper与“Pyannote.audio”的应用研究〉修订而成。

参考文献

- Amrate, M., & Tsai, P. (2025). Computer-assisted pronunciation training: A systematic review. *ReCALL*, 37 (1), 22-42. <https://doi.org/10.1017/S0958344024000181>
- Fanselow, J. F. (1977). Beyond Rashomon: Conceptualizing and describing the teaching act. *TESOL Quarterly*, 11 (1), 17-39. <https://doi.org/10.2307/3585589>
- Ferraro, A., Galli, A., La Gatta, V., & Postiglione, M. (2023). Benchmarking open source and paid services for speech to text: an analysis of quality and input variety. *Frontiers in Big Data*, 6, 1210559. <https://doi.org/10.3389/fdata.2023.1210559>
- He, F., & Lin, C. (2021). Supporting online Chinese narrative writing pedagogy through metacognitive writing process and approach: A design-based research. *Journal of Technology and Chinese Language Teaching*, 12(1), 117-137.
- Hu, W.-C. (2021). The theoretical foundation of virtual reality assisted language learning and its application in TCSL. *Journal of Technology and Chinese Language Teaching*, 12(2), 66-85. [胡文菊. (2021). 虛擬實境科技運用於語言學習的理論背景與華語教學範例. *科技与中文教学*, 12(2), 66-85.]

- Iino, A. (2009). Review of classroom observation systems -focusing on FLINT, COLT, and FOCUS, *Bulletin of Seisenjogakuin Junior College*, 27, 13-29.
http://purl.org/coar/resource_type/c_6501 [飯野厚. (2009). 語学授業観察法の概観—FLINT, COLT, FOCUS に焦点をあてて. 清泉女学院短期大学紀要, 27, 13-29. http://purl.org/coar/resource_type/c_6501]
- Ishizuka, H., Koshie, M., Sakurai, Y., Kamada, R. & Kubo, M. (2021). An attempt to develop foreign language teaching by the use of classroom analysis tool which can provide immediate feedback — By using mobile COLT. *Journal of Hokkaido University of Education (Humanities and Social Sciences)*, 72(1), 69-78. [石塚博規, 越江麻衣, 櫻井靖子, 鎌田亮祐, & 久保稔.(2021). 即時フィードバック可能な授業分析ツールによる外国語授業改善の試み: Mobile COLT を用いて. 北海道教育大学紀要 (人文科学・社会科学編), 72(1), 69-78.]
- Lian, W., Zheng, M., & Xu, J. (2024). A comparative study on development models of AI Chinese language partners based on the ERNIE large language model. *Journal of Technology and Chinese Language Teaching*, 15(2), 35-53. [连维琛, 郑明鉴, & 徐娟 (2024). 基于文心大模型的 AI 中文语伴开发模式对比研究. 科技与中文教学, 15(2), 35-53.]
- Macaro, E. (2009). Teacher use of codeswitching in the second language classroom: Exploring “optimal” use. In M. Turnbull & J. Dailey-O’Cain (Eds.), *First language use in second and foreign language learning* (pp. 35–49). Multilingual Matters. <https://doi.org/10.21832/9781847691972-005>
- Ministry of Education, Culture, Sports, Science and Technology. (2020). *Results of the FY2019 national student survey (pilot implementation): Data report*. [文部科学省. (2020). 令和元年度「全国学生調査（試行実施）」結果【資料編】] https://www.mext.go.jp/content/20200618-mxt_koutou01-000001987_03.xlsx
- Ministry of Education, Culture, Sports, Science and Technology. (2022). *Results of the FY2021 national student survey (second pilot implementation): University edition — Data report*. [文部科学省. (2022). 令和3年度「全国学生調査（第2回試行実施）」結果（大学）【資料編】.] https://www.mext.go.jp/content/20221110-koutou01-000001987_1.xlsx
- Ministry of Education, Culture, Sports, Science and Technology. (2023). *Results of the FY2022 national student survey (third pilot implementation): University edition — Data report*. [文部科学省. (2023). 令和4年度「全国学生調査（第3回試行実施）」結果（大学）【資料編】] https://www.mext.go.jp/content/20230712-koutou02-000001987_2.xlsx
- Mizumoto, A. (2023). Data-driven learning meets generative AI: Introducing the framework of metacognitive resource use. *Applied Corpus Linguistics*, 3(3), 100074. <https://doi.org/10.1016/j.acorp.2023.100074>
- Moskowitz, G. (1971). Interaction Analysis-A New Modern Language for Supervisors. *Foreign language annals*, 5(2), 211-221. <https://doi.org/10.1111/j.1944-9720.1971.tb00682.x>
- Mustafa, M. B., Yusoof, M. A., Khalaf, H. K., Rahman Mahmoud Abushariah, A. A., Kiah, M. L. M., Ting, H. N., & Muthaiyah, S. (2022). Code-switching in

- automatic speech recognition: The issues and future directions. *Applied Sciences*, 12(19), 9541. <https://doi.org/10.3390/app12199541>
- Myers, S. (2002). *Contact linguistics; bilingual encounters and grammatical outcomes*. Oxford University Press.
- OECD. (2019a). *OECD future of education and skills 2030/2040*. <https://www.oecd.org/en/about/projects/future-of-education-and-skills-2030.html>
- OECD. (2019b). *The OECD learning compass 2030*. <https://www.oecd.org/en/data/tools/oecd-learning-compass-2030.html>
- Pan, V. J., & Liu, C. (2023). Focus constructions involving *shi* in Mandarin Chinese. *Languages*, 8(2), 103. <https://doi.org/10.3390/languages8020103>
- Patsy, M. L., & Spada, N. (2021). *How languages are learned* (5th ed.). Oxford University Press.
- Pellerin, M., Ishizuka, H., & Cervantes, V. F. (2024). Language teaching supervision for a new era: AICOLT system's journey to AI-driven innovation. *Proceedings of the International CALL Research Conference*, 2024, 213–218. <https://doi.org/10.29140/9780648184485-32>
- Poole, F. J., & Coss, M. D. (2024). Can ChatGPT reliably and accurately apply a rubric to L2 writing assessments? The devil is in the prompt(s). *Journal of Technology and Chinese Language Teaching*, 15(1), 1-24. <https://doi.org/10.35542/osf.io/3r2zb>
- Qu, M., & Sunaoka, K. (2024). Pedagogical analysis of Chinese hyflex classrooms: A focus on teacher' and students' talk. *The Journal of Modernization of Chinese Language Education*, 13(2), 14-24. [曲明, & 砂岡和子. (2024). 汉语 Hyflex 课堂教学分析—教师与学生发言行为的分析. *中文教学现代化学报*, 13(2), 14-24.] <https://waseda.repo.nii.ac.jp/records/2005755>
- Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. In *International conference on machine learning*, 28492-28518. <https://doi.org/10.48550/arXiv.2212.04356>
- Shan, L., Pan, Z., & Weidman, R. (2024). Integrating task-based language teaching and generative AI: Design, implementation, and evaluation of the CFLingo Platform for Chinese learning. *Journal of Technology and Chinese Language Teaching*, 15(2), 1-34.
- Shinya, I., Moegi S., Hajime S., & Kimihiko H. (2020). Reality of interaction among teachers and researchers in lesson study. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, 27(4), 461-486. [飯窪真也, 齊藤萌木, 白水始, 堀公彦. (2020). 授業研究における教師と研究者の相互作用のリアリティ. *認知科学*, 27(4), 461-486.] https://www.jstage.jst.go.jp/article/jcss/27/4/27_2020.043/_pdf/-char/ja
- Shirouzu, H., Iikubo S., & Saito M. (2021). Birth, growth, and future issue in learning sciences: In search of science of learning that supports practices. *The Annual Report of Educational Psychology in Japan*. 60, 137-154. [白水始, 飯窪真也, & 齊藤萌木 (2021). 学習科学の成立, 展開と次の課題: —実践を支える学びの科学を模索して. *教育心理学年報*, 60, 137-154.]

- Spada, N., & Fröhlich, M. (1995). *COLT Communicative Orientation of Language Teaching observation scheme: coding conventions and applications*. National Centre for English Language Teaching and Research, Macquarie University.
- Sunaoka, K., Wang, S., Sugie, S., & Xu, Q. (2023a). Code-switching in Chinese language classes: Inclusive membership and L2 acquisition optimization. *Proceedings of the 72nd National Conference of the Chinese Language Society of Japan* (pp. 253-257). [砂岡和子, 王松, 杉江聡子, & 徐勤. (2023a). 中国語授業の Code-Switching—包摂的メンバーシップと L2 習得最適化. *日本中国語学会第 72 回全国大会予稿集* (pp.253-257).]
- Sunaoka, K., & Xu Qin. (2023b). Speaker diarization and text transcription in Chinese classrooms containing multilingual code-switching: Applied study of Whisper and Pyannote.audio. *Proceedings of the 12th International Conference and Workshops on Technology and Chinese Language Teaching (TCLT12)*, 61-71. (砂岡和子, & 徐勤. (2023b). [多语码汉语教学课堂中的话者分离与文本转录—Whisper 和 Pyannote.Audio 的应用研究. *第十二届国际汉语电脑教学研讨会论文集*, 61-71.]
- Sunaoka, K., Xu Qin. (2025). Self-analysis of foreign language classes using a multilingual Voice-to-Text app and AI: Development of the multilingual Voice-to-Text app. *Proceedings of the 31st Annual Meeting of the Association for Natural Language Processing (NLP2025)*, 1799-1804. [砂岡和子, & 徐勤. (2025). 多言語音声転写アプリと AI による外国語授業の自己分析—Multilingual Voice-to-Text App の開発. *言語処理学会第 31 回年次大会 (NLP2025) 予稿集*, 1799-1804.]
- Tao, H. (2025). Interactive competence: The integration of interactional linguistics and language teaching and its practice in teaching Chinese as a second language. *Journal of International Chinese Teaching*, 03, 2-16. [陶红印. (2025). 互动能力: 互动理论研究与语言教学的结合及其在汉语二语教学中的实践. *国际汉语教学研究*, 03, 2-16.]
- Tasaki, A. (2006). An overview of studies on code-switching: For analysis of communication in multilingual societies. *Language, Culture, and Japanese Language Education: Special Supplementary Issue — Frontiers in Second Language Acquisition and Education Research*, 54-84. [田崎敦子. (2006). コードスイッチング研究の概観: 多言語社会のコミュニケーション分析に向けて. *言語文化と日本語教育(増刊特集号, 第二言語習得・教育の研究最前線)*, 54-84.]
- Wang, Y. (2021). Pragmatic conditions and functions of the Chinese “shi...de” construction. *Bulletin of the Institute of Human Sciences, Toyo University*, 23, 17-37. [王亚新. (2021). 汉语“是…的”句的语用条件和功能. *東洋大学人間科学総合研究所紀要*, 23, 17-37.]
- Xiao, W., & Park, M. (2021). Using automatic speech recognition to facilitate English pronunciation assessment and learning in an EFL context: pronunciation error diagnosis and pedagogical implications. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, 11(3), 74-91.

- Xu Q., & Sunaoka, K. (2024). Speech transcription and speaker separation with multiple language codes: Advanced automatic speech recognition with Whisper + Pyannote.audio. *Proceedings of the 30th Annual Conference of the Association for Natural Language Processing*, 3149-3154. [徐勤, & 砂岡和子. (2024). 複数言語コードを含む発話転写と話者分離: Whisper+Pyannote.audio による自動音声認識の高度化. *言語処理学会第30回年次大会論文集*, 3149-3154.]
- Zhang, S. (2021). Integrating augmented reality into a task-based thematic language teaching unit. *Journal of Technology and Chinese Language Teaching*, 12(2), 29-48.