

ARTICLE



Teaching pronunciation online: A sociomaterial study of Swedish as a second language instruction in adult education

Andreas Lagerlöf, Linköping University

Abstract

This paper examines how pronunciation instruction is enacted in online Swedish as a Second Language (SSL) within Municipal Adult Education (MAE). Using Actor–Network Theory (ANT), classroom observations and teacher interviews trace how human and non-human actors (e.g., teachers, students, platforms, cameras, microphones, recordings) shape instruction. Five analytic scenes show that: (1) close-up camera/microphone affordances amplify articulatory modeling; (2) prosody practice becomes an auditory, distributed activity coordinated via audio files and chat functions; (3) corrective feedback is configured by one-speaker-at-a-time interfaces and headphone listening; (4) assessment is redistributed over time; and (5) domestic soundscapes and connectivity contingencies unevenly condition what is hearable and assessable. In the paper, pedagogical implications grounded in these scene-based analyses are articulated. Pronunciation instruction is thus shown to emerge as a sociomaterial practice, enacted through the actors and infrastructures of the online classroom.

Keywords: online pronunciation instruction; computer-assisted language learning (CALL); Actor–Network Theory (ANT); feedback; assessment

Language(s) Learned in This Study: Swedish

APA Citation: Lagerlöf, A. (2026). Teaching pronunciation online: A sociomaterial study of Swedish as a second language instruction in adult education. *Language Learning & Technology*, 30(1), 1–16.
<https://doi.org/10.64152/10125/73689>

Introduction

“Watch closely how my lips move when I say /u/ and /y/.” The teacher’s face fills the screen. Her lips move slowly and deliberately, framed by her hands to accentuate articulation. This close-up scene—mediated by a webcam, delivered via an educational platform, amplified through a microphone, and streamed through students’ headphones and speakers—illustrates how pronunciation instruction becomes materially and technologically configured in online instruction. Captured during an observation of a Swedish as a Second Language (SSL) online lesson in Municipal Adult Education (MAE), this moment encapsulates the rearticulated nature of teaching pronunciation in online environments.

As more adult second language (L2) instruction migrates online, teachers’ pedagogical practices are shaped through e.g., video conferencing programs, asynchronous recordings, chat functions, and learners’ domestic settings (Gruber et al., 2023). In research literature, pronunciation is considered a crucial aspect of L2 learning (Derwing & Munro, 2005; Leis, 2025) and remains a central component of L2 intelligibility and communicative success, yet studies show that it is often marginalized in digital environments, where reading and writing tend to be prioritized (Bergdahl & Hietajärvi, 2022). Recent studies highlight how online tools—ranging from audio players to AI-powered speech analysis—afford both opportunities and constraints for L2 pronunciation instruction (Godwin-Jones, 2023). Yet more remains to be known about how pronunciation instruction unfolds in online instruction, not as a static

method, but as a co-produced practice involving several actors, such as people, technologies, spaces, and interactions. This study contributes to ongoing conversations about how digital environments shape L2 pronunciation instruction and teacher/student interaction, especially in relation to embodiment, feedback, and technological affordances (Bahari, 2023; Šimáčková & Podlipský, 2023; Wang et al., 2024).

The study contributes to a growing body of research that views Computer-Assisted Language Learning (CALL) environments not as neutral containers for language teaching and learning, but as sites where teaching is continuously negotiated through technological and embodied entanglements (Pischetola et al., 2021; Wigham & Satar, 2021).

Aim and research questions

The aim of this study is to explore how pronunciation instruction in SSL online teaching within MAE unfolds. The study traces enacted practices and the designable conditions under which pronunciation work becomes hearable, reviewable, and assessable. The study is guided by the following research questions:

1. How do actors of the online classroom shape the embodied practices of L2 pronunciation instruction?
2. How do the affordances of technological actors reconfigure modeling, feedback, and assessment practices in online L2 pronunciation instruction?

Literature Review

Research on second language pronunciation is extensive, but several strands have been assessed as particularly relevant for this study. In this section, multimodally oriented research is first reported, followed by studies on augmented reality (AR) videoconferencing, computer-assisted pronunciation training (CAPT), and automatic speech recognition (ASR). Next, studies concerning teachers' online pronunciation teaching practices are presented, and finally, sociomaterially oriented studies on online teaching.

Research on embodiment and visual cues highlights how facial movements, gestures, and bodily presence in multimodal interaction make pronunciation visible and thereby possible to reproduce in authentic interactional situations (Farran & Morett, 2024). This becomes particularly relevant in online classrooms, as a study by Sewell et al. (2023) indicates that the presence of a speaking face has positive effects on the comprehensibility of spoken language—results pointing to the importance of mediating the teacher's body in such a way that students can clearly see the teacher's face. In a study by Wen et al. (2023), AR videoconferencing was shown to improve L2 students' pronunciation production, particularly of difficult segments, by reducing cognitive overload and directing attention toward relevant articulatory gestures.

Research on CAPT and ASR has mapped out how digital technologies can support modeling, practice, and feedback. A study by Martin (2020) shows that L2 students studying in online formats do not make clear progress without targeted pronunciation training. Furthermore, Silpachai et al. (2024) demonstrate that “good enough” machine feedback can be pedagogically effective. Taken together, the studies above highlight the tension between individual learner autonomy and teacher-led online teaching: while students may gain independence through self-directed practice, the absence of support risks turning pronunciation into a solitary rather than an interactive activity. The design of instruction and the sequencing of tasks thus become central in translating the potential of technology into actual learning outcomes. Visual feedback via spectrograms and articulatory visualizations, for example, can help students identify and correct errors (see e.g., Amrate & Tsai, 2025; Bliss et al., 2018). In addition, AI-based tools such as conversational bots are being tested for practice activities and explanations of pronunciation phenomena (Ji et al., 2024; Mompean, 2024). While these innovations point to increased efficiency, they raise questions of authenticity and engagement, since practicing with a bot or a spectrogram differs

substantially from practicing with a peer or a teacher in real time. These studies demonstrate both the opportunities and limitations of digitalized pronunciation instruction, thereby reinforcing the understanding that teachers continue to play decisive roles as carriers of presence and as directors of online pronunciation teaching—especially since earlier studies suggest that oral components tend to be sidelined in online teaching (Bergdahl & Hietajärvi, 2022; Foote et al., 2016).

Teachers' pronunciation teaching practices involve not only audio sound and decisions about what is mediated on screen, but also interaction, encouragement of participation, and crucial judgments about when and how to give feedback. These practices are difficult to replace, and even the most advanced tools often require teacher mediation to become pedagogically meaningful. Regarding pronunciation instruction, technology enables close-ups, sound amplification, and recordings that can strengthen modeling and practice (Cucchiarini et al., 2000; Lee et al., 2015; Leis, 2025; Toyama & Hori, 2025). However, there is also a risk that pronunciation becomes marginalized in digital learning environments, where reading and writing tend to be prioritized (Bergdahl & Hietajärvi, 2022; Butler & Starkey, 2024; Dalman, 2025).

Research with sociomaterial perspectives has shown how teaching is shaped in the interplay between embodiment, technology, and space (Mörtsell, 2024; Stenliden & Sperling, 2024; Valasmo et al., 2023), and how online teaching is always dependent on both human and material actors in the assemblages they constitute (Bangou & Waterhouse, 2021; Pischetola et al., 2021). Based on the above-mentioned studies, it is emphasized that online teaching is never merely a matter of technology in itself, but the outcome of networks involving people, technologies, pedagogical aspects, and institutional conditions. Taken together, these studies show that online L2 pronunciation teaching is shaped not only by pedagogical design and technological affordances, but also by the situated interplay between teachers, students, and tools. This study builds on these insights by applying a sociomaterial perspective to investigate how pronunciation teaching is enacted in digital classrooms in adult education.

Actor-Network Theory

ANT provides analytical tools and an approach for understanding social phenomena as effects of interconnected networks of relationships between both human and non-human actors (Fenwick & Edwards, 2010). In ANT, actors are not only people, but also objects, technologies, environments, and other elements that have the power to act within the network, that make a difference (Latour, 2005). These actions and effects are analyzed based on the principle of symmetry, which means that humans and things are analyzed equally, without giving greater significance to any actor in advance. Actors should also, as far as possible, be described using the same type of language, regardless of whether they are human or non-human.

ANT emphasizes actions that arise through *assemblages*, emerging from actors connecting. Assemblages hold together through relations; the whole exceeds the sum of its parts (Müller & Schurr, 2016). These assemblages are not fixed structures but are constantly formed and re-formed as interaction emerges. In this study, online teaching related to pronunciation in SSL education at MAE includes, for example, technologies, language, and human participants, who together co-produce teaching practices and enact how the L2 subject takes shape regarding pronunciation.

From an ANT perspective, material elements are not treated as passive objects but are understood as integrated and agentic components of the educational process (Sørensen, 2009). A central concept from ANT used in this study is *enactment*, which refers to how different entities come into being through interactions within the network (Mol, 2010). It captures how something is done in relationships - in this case, how pronunciation teaching emerges through the interplay between actors. The concept of enactment emphasizes that reality is not merely represented but enacted in practice, through relationships and actions in networks. Both human and non-human actors are thus analyzed and identified through the effects they produce, and how actions are generated in relation. This aligns with Latour's (2005) notion that actors act by making a difference in the relational assemblages in which they are embedded. ANT

offers a lens on how online L2 pronunciation teaching can be understood by giving voice to the components of the online educational environment.

Methods and Context of the Study

MAE in Sweden is regulated by the Swedish National Agency for Education and includes courses at basic level (SFI), elementary level, and upper secondary level, with all levels offering SSL. In line with policy demands for flexibility, a large part of MAE in Sweden is organized as distance and online education (Mufic, 2023).

Purposive sampling was used to capture variation across municipalities and program levels. Ten certified SSL teachers participated in the study; each was observed and interviewed twice between February 2023 and February 2024. [Table 1](#) provides an overview of the participating teachers, institutional contexts (four municipalities and five MAEs), and educational levels:

Table 1

Teacher ID, Municipalities and Educational Levels

Teacher ID	Municipality type	Educational level and MAE
T1	Large city municipality	Upper secondary level, MAE A
T2	Medium-sized municipality	Elementary and upper secondary levels, MAE B
T3	Medium-sized municipality	Elementary and upper secondary levels, MAE B
T4	Medium-sized municipality	Elementary and upper secondary levels, MAE B
T5	Medium-sized municipality	Basic level (Swedish for Immigrants), MAE C
T6	Medium-sized municipality	Basic level, MAE C
T7	Medium-sized municipality	Basic level, MAE C
T8	Medium-sized municipality	Basic level, MAE D
T9	Medium-sized municipality	Elementary and upper secondary levels, MAE E
T10	Medium-sized municipality	Elementary and upper secondary levels, MAE E

Note. All teachers were teaching SSL in synchronous online formats.

The observations included both observations of single lessons (14 instances), as well as extended teaching days (6 instances). They were conducted either with the researcher physically sitting next to the teacher (six instances) or by participating via screen (14 instances). This allowed the researcher to follow how pronunciation teaching online unfolded both in teachers' physical work environments and within the online classroom.

Online lessons were primarily conducted via Google Meet and Zoom and routinely involved webcam video, audio playback, chat, screen sharing, shared documents (Google Docs/itslearning), and built-in recording (itslearning), technologies that analytically were treated as actors within the assemblages of online instruction (cf. Sørensen, 2009). Observations ranged from single lessons (~80 minutes) to full days (up to 8 hours), yielding ~90 hours of data. Because pronunciation work was interwoven with regular instruction rather than isolated, fieldnotes targeted moments where pronunciation became pedagogically salient in interaction, including modeling, feedback, and breakdowns/disruptions (cf. Bangou & Waterhouse, 2021; Pischetola et al., 2021). Following each observed lesson, semi-structured interviews (40–90 minutes; total ~466 transcript pages) were conducted focusing on pedagogical choices, affordances/limits of tools, feedback strategies, and reflections (Brinkmann, 2016) on teaching L2 pronunciation online.

Data Analysis

The analytic work moved through repeated cycles between observation notes, interview transcripts, and analytic memos. In a first phase, the entire dataset was read multiple times to identify situations in which pronunciation became pedagogically salient, explicitly or implicitly—for example through modeling, speaking activities, corrective feedback, technical disruptions, or assessment practices. These sequences were open coded with attention to actions, actors, and relations (e.g., *camera proximity*, *headphone listening*, *audio delay*, *replay listening*, *muted microphones*, *collegial assessment*). In this phase, human and non-human actors were coded symmetrically, in line with ANT (Latour, 2005).

In a second phase, codes were compared and regrouped to identify recurring sociomaterial configurations. Here, assemblages were mapped—that is, how actors such as teachers, students, cameras, microphones, platform functions, sound environments, and recordings interacted in specific instructional situations. Assemblage mapping was used to examine how pronunciation instruction was enacted through relations among actors, rather than as isolated techniques or variables.

In a third phase, analytic scenes were constructed. Each scene was developed by comparing and synthesizing several empirical sequences into analytically distinct, yet empirically grounded, ways in which pronunciation instruction took shape online (e.g., close-up modeling, distributed prosody work, focused corrective feedback, time-shifted assessment, infrastructural vulnerability). The selection of the five scenes was based on their recurrence across teachers, their representation of different pedagogical functions in pronunciation instruction, and their illustration of distinct sociomaterial arrangements. At the same time, each scene is anchored in specific classroom events and teacher accounts. The scenes are thus analytically constructed around concrete observations and related interview excerpts from individual teachers. In this way, the scene-based analyses retain a close connection to situated teaching practices while highlighting recurring enactments across the material.

In a fourth phase, the scenes were analyzed with attention to *enactment* (Mol, 2010), that is, how pronunciation instruction emerged through relational effects among actors. This involved a closer examination of how, for example, webcam framing, platform audio logics, recording functions, and domestic environments participated in shaping what could be modeled, heard, repeated, and assessed. As one example, observation notes describing how a teacher leaned toward the webcam, students' microphones were muted, and several students used headphones were initially traced as *visual amplification*, *auditory isolation*, and *silent practice*. In later analytic stages, these were not treated as separate features but as effects of an assemblage in which camera, microphone, headphones, and bodies worked together—an assemblage that became central in the construction of Scene 1.

Throughout the process, analytic decisions were documented in memos, forming an audit trail. Interview data were used to probe, elaborate, and critically examine patterns identified in the observations.

The AI tool ChatGPT was used for translation from Swedish to English during the process of writing this paper, however, the tool was not used for generating text or analysis, only for translation purposes. An overview of the methods applied in the study is presented in [Appendix](#).

Rigor, Trustworthiness, and Reflexivity

Trustworthiness was strengthened through triangulation of observation and interview data, ongoing memo writing, and peer debriefing of coding and scene construction. Analytic decisions were documented in an audit trail. The author's prior experience in adult education was treated both as a resource and as a site for reflexive work, through continuous memos in which interpretations were critically examined.

Ethical Considerations

The study involved ten MAE teachers in the subject of SSL and focused solely on their teaching practices. Procedures followed the Swedish Research Council's guidelines (2017): informed consent, confidentiality, and secure data handling. All teachers were informed about the study and provided

written consent and could withdraw at any time. Students present during observations were informed about the study and provided verbal consent to participate; no personal data were recorded, and observations would have been discontinued if any student declined participation (none did). This study focuses on teachers enacted practices; direct measures of learner outcomes were thus not collected. Claims about learning are therefore limited to process-level opportunities and constraints evidenced in classroom activity and teacher accounts.

Results

The section below presents five empirically grounded analytic scenes from online SSL classrooms. Each scene is anchored in specific observations and interviews (identified by teacher IDs) and is analytically constructed from several empirical sequences. Together, the scenes show how pronunciation work is organized through recurring sociomaterial configurations of instruction, tools, and conditions for audibility and participation, ranging from close-up articulatory modeling to time-shifted assessment and infrastructural breakdowns.

Scene 1 – Close-up Modeling of Articulation via Camera/Mic

During a segment of an online pronunciation lesson, the teacher (T8) leans close to the camera. The lesson is focused on the phenomenon of long and short vowels. In a demonstration, the teacher moves even closer so that her lip shaping is visible to the students. She shows how to form the mouth for the sound [ɑ:], producing the sound slowly and clearly. She uses her hands to emphasize the movement, cupping them around her mouth and pointing to the position of the lips. Some students wear headphones and stay silent, others appear to be practicing the sound, but this is not audible to the teacher or to other students since all student microphones are muted. The teacher is therefore unable to provide feedback on these students' pronunciation. "I am braver with using my facial expressions here than in the classroom," the teacher says in a follow-up interview. She describes how elements such as camera optics, microphone capture, and not being physically in the same room as the students make her feel able to "exaggerate a bit more." She continues:

Here, [it's as if] every student ends up at the very front of the classroom if they want. Sometimes [in a physical classroom] someone ends up at the back behind a pillar, or there's drilling from the floor below. Now they [students] can choose more freely. I also have a somewhat thin, ordinary voice, but now I can manage better and everyone can hear me. (T8)

Analysis

Pronunciation instruction here emerges as a practice enacted through an assemblage of bodily expressions (lips, voice, hands), technologies (camera, microphone, headphones), and spatial configurations that are specific to the online environment (cf. Fenwick & Edwards, 2010). The empirical description shows how the teacher repeatedly leans toward the webcam so that her lips dominate the screen, slows down her articulation, and frames her mouth with her hands—actions that work in concert with the webcam's close-up framing and the microphone's selective amplification. These actors do not merely "mediate" instruction; they shape the teacher's scope for pedagogical action. The teacher's comment that she feels more comfortable using pronounced facial expressions online points to how camera proximity and audio capture reconfigure embodiment and make exaggerated articulatory modeling both possible and pedagogically salient.

This aligns with research emphasizing visual modeling and embodied interaction in pronunciation instruction (e.g., Munro & Derwing, 2011), while also showing how such practices take on practice-specific forms in online environments. Articulatory modeling emerges as an effect of a sociomaterial assemblage in which webcam framing, audio amplification, bodily positioning, and spatial arrangements together shape what can be seen, heard, and made available for modeling.

Scene 2 – Prosody Training as an Auditory, Distributed Practice

“Listen carefully. Where do you hear the long sound?” The teacher (T2) speaks into the microphone of her headset, and her voice is mediated through the students’ speakers and headphones. On the screen—within the learning platform the teacher is sharing via the videoconferencing program—a word list, small video windows showing the teacher and the three students present, and a short set of instructions in the chat. The students are to identify where the stress falls in a word—on either the first or the second syllable. “Here’s a new word.” A click is heard as the teacher plays the next audio file: “banan” ([baˈna:n]; “banana”). Chat notifications pop up as students type either “1” or “2” to indicate which syllable they think is stressed. One student answer incorrectly, but the teacher does not comment. She plays the audio again and says, “Try saying the word quietly to yourself—feel what sounds right.” After a brief pause, her voice returns: “Think about the rhythm. Long sounds take up space. We want to get Swedish word stress right.” She exaggerates a few words, first close to the mic, then farther back. “Now let’s listen to each other. Say the word you wrote in the chat, I want to hear your stress.” A student turns on the mic and says “[baˈnan],” with the stress on the second syllable. The teacher responds, “Good! Do you hear the difference?” Several mics activate and some laughter is heard. A final audio clip plays, and the teacher concludes: “I’ll share the link so you can keep practicing at home. Don’t forget to work on the rhythm—it’s the key to fluency.”

Analysis

In this scene, pronunciation emerges through a tightly structured sequence of actions: playing audio files, silent listening, typing in the chat, quiet self-repetition, and short oral turns. The teacher repeatedly alternates between playing recordings (“Here’s a new word”), inviting students to respond in the chat (“1” or “2”), and prompting embodied listening (“Try saying the word quietly to yourself—feel what sounds right”). Prosody work is thus organized through cycles of mediated listening and brief moments of speech, rather than through extended oral production. At the same time, the chat becomes a parallel interactional channel where students can display perception without speaking, making stress judgments visible even when microphones remain muted. These observed cycles—playing an audio file, responding in the chat, and prompting quiet self-repetition—ground the analytic claim that prosody instruction is distributed across audio playback, written responses, and embodied self-rehearsal. When students are finally invited to turn on their microphones and “listen to each other,” short oral contributions are layered onto an already established auditory pattern. Pronunciation is enacted here through a distributed configuration in which audio files, chat entries, headset-mediated voice, and silent bodily repetition together shape how online prosody practice takes form.

Scene 3 – Individual Corrective Feedback

“Wait—can you say that word again?” The teacher’s (T7) voice interrupts the student’s reading aloud. The student pauses and searches the text: “etablering” ([e.taˈble:rɪŋ], “implementation”), they offer. When a student speaks, the picture of the student’s face is enlarged so that it covers most of the screen, while the other students’ pictures stay small on the right-hand side of the teacher’s screen. The teacher leans in toward the screen, her face in close-up. “Almost. Listen closely: e-ta-ble-ring.” She breaks the word down into syllables, articulating clearly and slowly. The student repeats, first hesitantly, then more fluently. “Good. And do you know what it means?” They continue reading. When the next unfamiliar word—“resonemang” ([re.sø.nɛˈmaŋ], “reasoning”)—comes up, the process repeats. The teacher pauses, sounds it out, asks for the meaning, and explains how to pronounce words. It is a rhythmic interplay: read, pause, listen, understand, move on. During the entire four-minute sequence, the teacher stops at six different words to give direct pronunciation feedback. Each correction is followed by the student’s attempt at the correct pronunciation and further feedback from the teacher. The online classroom is quiet except for their voices.

Afterward in an interview, T7 says: “I find it easier to give feedback this way. It becomes so focused. The student is wearing headphones, sees only me, we’re in it together.” She explains how she consciously

uses stress and intonation to enhance linguistic clarity. “When I only see one student at a time, it’s easier to hear exactly what they need.”

Analysis

Pronunciation instruction emerges as a focused and responsive practice, enacted through a tightly organized sequence of reading, interruption, modeling, repetition, and confirmation. The empirical description shows how each feedback episode follows a recurring cycle: the student reads aloud, the teacher interrupts (“Wait—can you say that word again?”), articulates the word in segmented syllables, the student repeats, and the teacher confirms before the reading resumes. Across the four-minute sequence, this cycle is repeated six times. Pronunciation feedback thus takes the form of short, rhythmic micro-interactions, making corrective work locally concentrated and interactionally dense.

These feedback episodes are materially shaped by the platform interface. When the student speaks, their video tile is automatically enlarged and occupies most of the screen. At the same time, the teacher leans toward the camera, bringing her face into close-up. This screen configuration produces a visual field in which the student’s voice and the teacher’s face dominate the interactional space, while other participants are visually and audibly backgrounded. The teacher’s feedback is therefore not only heard but also seen; mouth movements, syllable segmentation, and facial emphasis become perceptually salient resources in the correction sequence. These observed features—teacher interruption, syllable-by-syllable articulation (“e-ta-ble-ring”), repeated student attempts, and the speaker-focused screen framing—ground the analytic claim that what counts as “feedback” is co-produced by bodily action, platform logics, and audio technology. The teacher’s interview comments—“The student is wearing headphones, sees only me, we’re in it together” and “When I only see one student at a time, it’s easier to hear exactly what they need”—directly echo these observed arrangements. Her sense of focused engagement is grounded in concrete material conditions documented in the scene: one active speaker, reduced background sound, enlarged faces, and headphone-mediated listening. Feedback emerges here as relational effects (cf. Mol, 2010) of teacher, student, text, camera, microphone, and interface working together.

The teacher’s face, repeatedly brought into close-up, functions as a pedagogical instrument. Facial movements and slowed articulation become resources for guiding the learner’s production, aligning with research on the importance of visual and embodied modeling in pronunciation instruction (Munro & Derwing, 2011). At the same time, the practice is reconfigured by the online setting: individualized feedback is sustained through interface logics that privilege one-speaker-at-a-time talk and visually foreground the current speaker. In this scene, the reduction of competing sounds and images, rather than their absence, creates the conditions for the dense, moment-by-moment feedback work that McCrocklin (2019) associates with effective pronunciation instruction. The concentrated feedback encounter is thus produced not despite digital mediation, but through the specific ways the platform rearranges participation, perception, and turn-taking (cf. Gacs et al., 2020).

Scene 4 – Recordings and Collegial Feedback across Time

After a read-aloud session on a literary text, the students are asked to write answers to a set of reading comprehension questions available on the learning platform *itslearning*. In the final task, they are supposed to write a new ending to the text. If they do not finish the task during class, it becomes homework for next time. At the end of the lesson, the teacher (T6) says: “Next week you will present your texts orally. I will record everything so I can give you feedback on your pronunciation afterwards.”

A week later, the oral presentations of the students’ texts begin. “I’m starting the recording now,” says the teacher before the students read their texts. They read their prepared texts slowly and carefully, seemingly conscious of the recording. The teacher listens in silence and takes notes without giving feedback in the moment. After the lesson, she downloads the recordings onto her computer. Some she listens to by herself, while others—where she wants help with co-assessment—she goes through together with a colleague. Together they pause, rewind, and discuss where the stress falls incorrectly, where vowel length needs to be extended, and which words signal comprehension problems related to the students’

interlanguage and transfer errors. Later the teacher writes individual feedback based on the assessment of each student's recorded speech and uploads it to the student's profile on the learning platform. In an interview, she reflects:

As a teacher it is valuable that we can record each presentation and review it afterwards, because it can be quite difficult if you have, say, five oral presentations in a classroom to remember everything and give fair feedback that really helps each student afterwards. (T6)

Analysis

In this scene, pronunciation assessment is reorganized around the possibility to record, store, and replay students' oral production. The empirical description shows how the teacher explicitly announces the recording ("I'm starting the recording now"), withholds immediate feedback during the presentations, and instead takes notes while students read. Assessment work is thus deliberately displaced from the live interaction to a later moment, when the recordings can be revisited. After the lesson, the teacher downloads the files, listens to them repeatedly, and—in some cases—invites a colleague to co-listen. The assessment process therefore unfolds across several stages: live performance, recorded storage, individual replay, collegial review, and finally written feedback uploaded to the platform.

These observed practices ground the analytic claim that the built-in recording function enables assessment distributed over time and space. Pronunciation is no longer only evaluated in the fleeting moment of classroom talk, but becomes an object that can be paused, rewind, compared, and discussed. The teacher's description of stopping, rewinding, and jointly identifying misplaced stress, insufficient vowel length, and transfer-related pronunciation problems shows how the recording function materially supports a more analytic mode of listening. What counts as "feedback" here is thus produced through a chain of actions: pressing record, downloading files, replaying segments, listening together, writing comments, and uploading them to individual student profiles.

The teacher's reflection that recordings make it easier to give "fair feedback that really helps each student" is directly tied to these documented practices. Her sense of fairness is not an abstract belief but is anchored in concrete material operations: the possibility to return to the same utterance several times, to compare students' productions, and to involve a colleague in evaluative decisions. In this way, technological affordances (record, replay, storage) interact with pedagogical concerns (clarity, fairness, support) to reconfigure how assessment is done.

The scene thus illustrates how pronunciation teaching becomes recordable and, by extension, co-assessable. Documentation and assessment are not carried by a single actor but by an actor-network (cf. Latour, 2005) of software, recordings, teachers, colleagues, students, and platform infrastructures. Within this network, students' speech is transformed from an ephemeral classroom event into a durable object that can circulate across times, spaces, and professional relations. Pronunciation becomes temporally multiple: first performed, then replayed, then collaboratively interpreted, and finally re-presented to students as individualized feedback. What assessment "is" and "does" is therefore reconfigured through these enactments.

Scene 5 – Infrastructural Contingencies

During an online lesson, students log in from different locations. Some appear to be at kitchen tables, others in office environments. One student has their camera off, while another is wearing headphones. When one student starts reading aloud, the sound glitches; their voice becomes a metallic echo before suddenly disappearing. The teacher (T1) asks the student to repeat themselves. "You glitched a bit there, could you try again?" she says, but the connection remains unstable.

In another window, a student turns around quickly when a child calls out in the background. They apologize, and the teacher encourages them to continue reading. In the background, a child drops something, a scream is heard, and the student immediately turns off both video and sound. "If it's quiet at home and the internet works, it's great," T1 says in an interview, "but if not, everything falls apart."

Several teachers (T1, T2, T5, T9, T10) describe having to adjust their feedback according to the technical conditions. T5 mentions that when a student's microphone produces static or when the internet connection is weak, it becomes difficult to determine whether the issue lies with pronunciation or with the technology, and teachers adapt to this. "Sometimes you don't know if it's an audio glitch or actually a pronunciation problem," T5 says in an interview. From observations of T9 and T10, some students have professional headsets and strong connections, which gives them an advantage in class, while others participate with basic equipment or from noisy environments. In one example, T9 chose to let students record their presentations afterward because, she says: "[...] you couldn't hear properly during the lesson – too much interference." These recordings were submitted through the learning platform and reviewed by the teacher in her office, where technical problems were no longer a factor.

Analysis

This scene foregrounds how pronunciation instruction becomes infrastructurally contingent through concrete disruptions and adjustments. The empirical description shows how audio glitches interrupt a student's reading, how background noise from a child leads a student to mute both sound and video, and how the teacher repeatedly has to ask students to repeat themselves. These moments make visible how domestic soundscapes, internet stability, microphones, and headsets participate in shaping what can be heard, sustained, and responded to in pronunciation work.

In ANT terms, these elements are actors that enter the assemblage of instruction (cf. Fenwick & Edwards, 2010; Latour, 2005). The metallic echo, the dropped connection, the screaming child, and the muted camera are not external "disturbances" to an otherwise stable pedagogy; they directly reorganize interaction by interrupting turns, silencing students, and producing uncertainty about what is being heard. The teacher's comment—"Sometimes you don't know if it's an audio glitch or actually a pronunciation problem"—is grounded in these observed breakdowns and captures how agency becomes distributed across bodies, devices, and infrastructures (cf. Sørensen, 2009). The ability to judge pronunciation is here materially dependent on routers, microphones, home acoustics, and bandwidth.

The data also show how teachers respond pedagogically to these infrastructural conditions. In the example where T9 asks students to submit recorded presentations after the lesson because "you couldn't hear properly during the lesson," assessment is actively relocated from the unstable live environment to a quieter, more controlled setting. This shift can be read as an infrastructural adaptation: pronunciation assessment is reorganized around other times, places, and material arrangements. Through this move, the teacher attempts to re-stabilize what can be heard and evaluated, redistributing assessment across different spatiotemporal configurations (Mol, 2010).

Taken together, the observed audio breakdowns, household interruptions, and teachers' accounts of uncertainty ("audio glitch or actually a pronunciation problem") ground the analytic claim that online pronunciation instruction is an infrastructurally vulnerable practice. It is enacted through fluctuating configurations of home environments, network quality, and unequal access to equipment. Some students' professional headsets and stable connections afford clearer participation, while others' noisy settings and basic devices constrain what becomes intelligible and assessable. Rather than treating these factors as external obstacles, the scene shows how they are woven into the teaching itself. Pronunciation instruction here emerges as a situated, relational achievement—continuously negotiated through the interplay of teachers' actions, students' domestic contexts, and the material capacities of digital infrastructures (cf. Bangou & Waterhouse, 2021; Valasmo et al., 2023).

Discussion

The five empirical scenes enable an analysis of what characterizes online pronunciation instruction and show how the teaching of SSL in MAE is shaped by enacted interaction between several actors of the online classroom. The analysis shows how specific human–nonhuman configurations make possible (or inhibit) articulation focus, prosody attention, and time-shifted assessment. The study foregrounds how

seemingly mundane elements—such as webcam positioning, audio latency, recording features, and chat functions—produce the affordances and limitations of pronunciation instruction enacted online. The analysis draws attention to the distributed nature of pedagogical presence and control, adding nuance to accounts that position the teacher as the primary orchestrator. By showing how actors of the online classroom enact what can be modeled, heard, repeated, and assessed in pronunciation instruction, the study extends earlier research on technology-enhanced pronunciation teaching and CALL (e.g., Cucchiaroni et al., 2000; Lee et al., 2015; Godwin-Jones, 2023; Leis, 2025). The study also complements CAPT/ASR research by shifting attention from tool effectiveness to the sociomaterial arrangements and enactments through which pronunciation pedagogy comes into being in practice (Martin, 2020; Silpachai et al., 2024).

Online pronunciation instruction is analyzed as a network in which teachers, students, technology, and physical environments together enable, or hinder, pronunciation instruction. Recurring in the material is how the teacher's body is transformed and made visible in new ways in the online environment, affecting pronunciation instruction. In this sense, the findings align with research showing that articulatory cues and embodied demonstration can support phonetic awareness and comprehensibility, while also specifying how such cues are technologically reconfigured when mediated through webcam framing and audio capture (Sewell et al., 2023; Farran & Morett, 2024; Wen et al., 2023). The camera and microphone allow specific parts of the teacher's body – often mouth movements and facial expressions – to be isolated and amplified, becoming central tools in pronunciation instruction. One teacher describes how they are “more confident” with facial expressions and are therefore able to reinforce modeling in ways that do not always occur in physical classrooms. The online classroom's affordances change how the teacher's body is mediated and therefore lead the teacher to use her body (e.g., lip formation in producing vowels) more explicitly. At the same time, the teacher appears as an active moderator and manager of turn-taking. In the online setting, more explicit management of dialogue is required since many of the informal turn-taking cues present in physical classrooms are lost. Teachers describe taking greater responsibility for distributing speaking time fairly, keeping track of who has spoken, and ensuring that silences are not too long or uncomfortable. This resonates with prior online-teaching research showing that interactional order and participation management tend to become more explicit in videoconferencing, particularly when platform logics privilege one-speaker-at-a-time talk (Wigham & Satar, 2021; Gacs et al., 2020).

Another key finding is how feedback and assessment of pronunciation are transformed. By recording oral presentations, teachers can listen multiple times, analyze in detail, and collaborate with colleagues in the assessment process. This could also be done in a physical classroom setting; however, since the instruction takes place in a digital environment, teachers only need to press a button to start a recording. Students also get the opportunity to listen back to their own performance and reflect on their pronunciation. The possibility to replay and revisit learner speech mirrors and extends research on technology-supported feedback practices in pronunciation learning, while showing how “recordability” becomes an infrastructural condition, for fairness and analytic attention in everyday teaching (McCrocklin, 2019; Silpachai et al., 2024). Further, the technology of the online classroom can make feedback a collegial practice but also alter the temporal and spatial aspects of when, how, and where assessment takes place. Students' home environments and technical conditions become decisive factors for the quality of pronunciation instruction. When the home sound environment is calm and the technology works as expected, online instruction is described as efficient and clear. However, when the internet falters or background noise breaks through, both the teacher's ability to hear pronunciation clearly and the student's ability to concentrate might be affected. Teachers manage these challenges by using strategies such as allowing students to submit recordings after the lesson or offering flexible feedback when technical problems hinder real-time interaction. This finding also speaks to sociomaterial accounts of online education where domestic environments and technical infrastructures are treated as active participants in what teaching can become, rather than external “noise” to be controlled (Bangou & Waterhouse, 2021; Pischetola et al., 2021; Valasmo et al., 2023).

Moreover, the findings resonate with and extend prior research that acknowledges the multimodal and embodied nature of pronunciation teaching (e.g., Šimáčková & Podlipský, 2023), while also highlighting how these embodied practices are conditioned by platform-specific constraints and technological rhythms. In this sense, the study contributes to a growing sociomaterial understanding of digital language pedagogy (e.g., Bangou & Waterhouse, 2021; Pischetola et al., 2021; Silpachai et al., 2024), emphasizing that teaching is always enacted through specific configurations of people, tools, time, and space. It does not suggest that such dynamics are absent from in-person contexts, but rather that they may become visible—or differently configured—when teaching emigrates online. By tracing how feedback, modeling, and assessment are materially enacted, the article invites researchers and practitioners to consider how technologies participate in shaping pedagogical interaction. These insights may inform future design, teacher training, and critical engagement with the affordances and constraints of online language teaching.

Pedagogical Implications

Pedagogical implications below are grounded in the analytic scenes and should be read as situated possibilities rather than prescriptive models. The findings show that online pronunciation teaching is highly sensitive to how bodies and voices are framed through cameras and microphones. Close camera positioning, selective amplification, and headphone listening shape what aspects of articulation become perceptible and pedagogically usable. This underscores the importance of treating visual framing and audio configuration as integral to pronunciation pedagogy, rather than as technical backdrops. Further, pronunciation work is sustained through tightly organized micro-sequences that combine listening, silent rehearsal, written responses, and short oral turns. Such distributed arrangements suggest the pedagogical value of designing activities that move flexibly across modalities. Platform features that support rapid shifts between listening, writing, and speaking emerge as central resources for structuring pronunciation practice.

Feedback practices are shaped by both interface logics and recording infrastructures. Short, focused exchanges enable fine-grained corrective work, while recording functions allow feedback and assessment to extend beyond the live lesson through replay, analytic listening, and, at times, collegial review. This highlights the pedagogical potential of integrating synchronous feedback with replay-based and time-shifted assessment practices.

Finally, the study makes visible how pronunciation instruction is infrastructurally contingent. Home soundscapes, connectivity, and access to equipment condition what can be heard, interpreted, and assessed. Flexible pedagogical designs that allow movement between synchronous and asynchronous formats, together with platform support for device checks and alternative submission modes, are therefore central to sustaining equitable pronunciation work online. Across these dimensions, online pronunciation instruction emerges as an accomplishment of both pedagogy and interface, where small platform conditions participate in configuring what can be modeled, heard, repeated, and assessed.

Limitations of the Study

Findings derive from Swedish MAE settings and from configurations typical of Google Docs/Google Meet/Zoom/itslearning and so transfer to other systems may vary. The study privileges enacted classroom practices over direct outcome testing; claims about learning are theoretically and empirically grounded but not causal. Connectivity and device heterogeneity constrained a subset of recordings; while treated analytically as part of the phenomenon, this also reduced audio granularity in those scenes. These issues are mitigated through observations and interviews, replay-based analysis, and transparency about research contexts.

Final Reflections

This study contributes to ongoing discussions in CALL, L2 pronunciation education, and sociomaterial research by illustrating how central pedagogical practices are shaped through material conditions and

infrastructural arrangements. It builds on and complements the work offering a situated empirical perspective on how pronunciation instruction is enacted, made visible, and rendered audible in online settings. The teacher's role appears in a new light, not just as a provider of input or evaluator of output, but as a moderator of participation, coordinator of rhythm, and translator between material conditions and pedagogical goals. At the same time, the study highlights the fragility of these enactments, revealing how inequalities in infrastructure and access materially interfere with the teaching of something as embodied and perceptual as pronunciation. This sensibility is consistent with research showing that online language teaching conditions participation and oral work unevenly, and that infrastructural variation shapes what becomes hearable, assessable, and pedagogically actionable (Gacs et al., 2020; Bergdahl & Hietajärvi, 2022).

Ultimately, this study underscores that effective pronunciation instruction in online environments cannot rely solely on technique or content design. It must be attuned to the sociomaterial conditions through which instruction is enacted and disrupted. In this space, technology functions as a co-teacher, a gatekeeper, and sometimes a barrier. Understanding these dynamics is key to designing equitable and pedagogically sound L2 education in online settings. Design explicitly mobilizes these affordances to stage and sustain pronunciation work, while teacher presence and relational work remain crucial. Future work may further explore how platform features can be aligned with embodied teaching routines, and how infrastructural inequities shape learners' opportunities to participate in pronunciation work across educational online settings.

References

- Amrate, M., & Tsai, P. H. (2025). Computer-assisted pronunciation training: A systematic review. *ReCALL*, 37(1), 22–42. <https://doi.org/10.1017/S0958344024000181>
- Bahari, A. (2023). Affordances and challenges of technology-assisted language learning for motivation: A systematic review. *Interactive Learning Environments*, 31(9), 5853–5873. <https://doi.org/10.1080/10494820.2021.2021246>
- Bangou, F., & Waterhouse, M. (2021). New materialist perspectives on language education. *The Canadian Modern Language Review*, 77(4), 291–301. <https://doi.org/10.1080/10494820.2021.2021246>
- Bergdahl, N., & Hietajärvi, L. (2022). Social engagement in distance-, remote-, and hybrid learning. *Journal of Online Learning Research*, 8(3), 315–342. <https://doi.org/10.70725/885512cuowjr>
- Bliss, H., Abel, J., & Gick, B. (2018). Computer-assisted visual articulation feedback in L2 pronunciation instruction: A review. *Journal of Second Language Pronunciation*, 4(1), 129–153. <https://doi.org/10.1075/jslp.00006.bli>
- Brinkmann, S. (2016). Methodological breaching experiments: Steps toward theorizing the qualitative interview. *Culture & Psychology*, 22(4), 520–533. <https://doi.org/10.1177/1354067X16650816>
- Butler, L., & Starkey, L. (2024). OK Google, help me learn: an exploratory study of voice-activated artificial intelligence in the classroom. *Technology, Pedagogy and Education*, 33(2), 135–148. <https://doi.org/10.1080/1475939X.2024.2311779>
- Cucchiari, C., Strik, H., & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *The Journal of the Acoustical Society of America*, 107(2), 989–999. <https://doi.org/10.1121/1.428279>
- Dalman, M. (2025). Second language (L2) learners' perceptions of online-based pronunciation instruction. *Languages*, 10(4), Article 62. <https://doi.org/10.3390/languages10040062>

- Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, 39(3), 379–397. <https://doi.org/10.2307/3588486>
- Farran, B. M., & Morett, L. M. (2024). Multimodal cues in L2 lexical tone acquisition: Current research and future directions. *Frontiers in Education*, 9, Article 1410795. <https://doi.org/10.3389/educ.2024.1410795>
- Fenwick, T., & Edwards, R. (2010). *Actor-network theory in education*. Routledge. <https://doi.org/10.4324/9780203849088>
- Foote, J. A., Trofimovich, P., Collins, L., & Urzúa, F. S. (2016). Pronunciation teaching practices in communicative second language classes. *The Language Learning Journal*, 44(2), 181–196. <https://doi.org/10.1080/09571736.2013.784345>
- Gacs, A., Goertler, S., & Spasova, S. (2020). Planned online language education versus crisis-prompted online language teaching: Lessons for the future. *Foreign Language Annals*, 53(2), 380–392. <https://doi.org/10.1111/flan.12460>
- Godwin-Jones, R. (2023). Emerging spaces for language learning: AI bots, ambient intelligence, and the metaverse. *Language Learning & Technology*, 27(2), 6–27. <https://doi.org/10.64152/10125/73501>
- Gruber, A., Matt, E., & Leier, V. (2023). Transforming foreign language education: Exploring educators' practices and perspectives in the (post-) pandemic era. *Education Sciences*, 13(6), Article 601. <https://doi.org/10.3390/educsci13060601>
- Ji, H., Han, I., & Park, S. (2024). Teaching foreign language with conversational AI: Teacher-student-AI interaction. *Language Learning & Technology*, 28(2), 91–108. <https://doi.org/10.64152/10125/73573>
- Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford University Press.
- Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, 36(3), 345–366. <https://doi.org/10.1093/applin/amu040>
- Leis, A. (2025). How speech-to-text technology affects pronunciation gains and self-confidence in EFL learners. *Computer Assisted Language Learning*. Advance online publication. <https://doi.org/10.1080/09588221.2025.2534498>
- Martin, I. A. (2020). Pronunciation development and instruction in distance language learning. *Language Learning & Technology*, 24(1), 86–106. <https://doi.org/10.64152/10125/44711>
- McCrocklin, S. (2019). ASR-based dictation practice for second language pronunciation improvement. *Journal of Second Language Pronunciation*, 5(1), 98–118. <https://doi.org/10.1075/jslp.16034.mcc>
- Mol, A. (2010). Actor-network theory: Sensitive terms and enduring tensions. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 50(1), 253–269.
- Mompean, J. A. (2024). ChatGPT for L2 pronunciation teaching and learning. *ELT Journal*, 78(4), 423–434. <https://doi.org/10.1093/elt/ccae050>
- Müller, M., & Schurr, C. (2016). Assemblage thinking and actor-network theory: Conjunctions, disjunctions, cross-fertilisations. *Transactions of the Institute of British Geographers*, 41(3), 217–229. <https://doi.org/10.1111/tran.12117>

- Mufic, J. (2024). 'And suddenly it's not so flexible anymore!' Discursive effects in comments from school leaders and staff about distance education. *Studies in the Education of Adults*, 56(1), 102–120. <https://doi.org/10.1080/02660830.2023.2249141>
- Munro, M. J., & Derwing, T. M. (2011). The foundations of accent and intelligibility in pronunciation research. *Language Teaching*, 44(3), 316–327. <https://doi.org/10.1017/S0261444811000103>
- Mörtsell, S. (2024). Lesson enactments: Maintenance in everyday educational practice. *Postdigital Science and Education*, 6(2), 595–609. <https://doi.org/10.1007/s42438-023-00401-z>
- Pischetola, M., de Miranda, L. V. T., & Albuquerque, P. (2021). The invisible made visible through technologies' agency: A sociomaterial inquiry on emergency remote teaching in higher education. *Learning, Media and Technology*, 46(4), 390–403. <https://doi.org/10.1080/17439884.2021.1936547>
- Sewell, K., Brown, V. A., Farwell, G., Rogers, M., Zhang, X., & Strand, J. F. (2023). The effects of temporal cues, point-light displays, and faces on speech identification and listening effort. *Plos one*, 18(11), Article e0290826. <https://doi.org/10.1371/journal.pone.0290826>
- Silpachai, A., Neiriz, R., Novotny, M., Gutierrez-Osuna, R., Levis, J. M., & Chukharev, E. (2024). Corrective feedback accuracy and pronunciation improvement: Feedback that is 'good enough'. *Language Learning & Technology*, 28(1), 1–16. <https://doi.org/10.64152/10125/73582>
- Šimáčková, Š., & Podlipský, V. J. (2023). Online guided pronunciation practice helps adult EFL learners improve L2 prosody. *AUC PHILOLOGICA*, 2022(1), 115–129. <https://doi.org/10.14712/24646830.2022.32>
- Stenliden, L., & Sperling, K. (2024). Breaking the magic of automation and augmentation in Swedish classrooms. *Nordisk tidsskrift for pedagogikk og kritikk* [Nordic Journal of Pedagogy and Criticism], 10(1), 15–32. <https://doi.org/10.23865/ntp.v10.6174>
- Sørensen, E. (2009). *The materiality of learning: Technology and knowledge in educational practice*. Cambridge University Press.
- Toyama, M., & Hori, T. (2025). Technology-enhanced multimodal approaches in classroom L2 pronunciation training. *Frontiers in Education*, 10, Article 1552470. <https://doi.org/10.3389/educ.2025.1552470>
- Valasmo, V., Paakkari, A., Sahlström, F., Backholm-Nyberg, Y., & Majors, J. (2023). Invisible participation: A sociomaterial analysis of a synchronous distance lesson during emergency remote teaching. *Digital Culture & Education*, 14(4), 112–132. <https://www.digitalcultureandeducation.com/s/Valasmo-2023.pdf>
- Wang, S., Li, J., & Liang, Q. (2024). Visual reinforcement through digital zoom technology in FL pronunciation instruction. *Language Learning & Technology*, 28(1), 1–26. <https://doi.org/10.64152/10125/73558>
- Wen, Y., Li, J., Xu, H., & Hu, H. (2023). Restructuring multimodal corrective feedback through Augmented Reality (AR)-enabled videoconferencing in L2 pronunciation teaching. *Language Learning & Technology*, 27(3), 83–107. <https://doi.org/10.64152/10125/73533>
- Wigham, C. R., & Satar, M. (2021). Multimodal (inter) action analysis of task instructions in language teaching via videoconferencing: A case study. *ReCALL*, 33(3), 195–213. <https://doi.org/10.1017/S0958344021000070>

Appendix. Method Overview

Element	Summary
Setting & participants	10 certified SSL teachers in Municipal Adult Education (MAE); 4 municipalities; 6 schools; levels: SFI (basic), compulsory, and upper secondary; fully online context.
Data collection period	February 2023–February 2024.
Sampling strategy	Purposive, maximum variation by municipality, program level, and platforms/tools.
Data corpus	Classroom observations \approx 90 hours (from 80-min lessons to full-day sessions), \sim 2 observations/teacher; semi-structured interviews $n=20$ (40–90 min; $M \approx$ 48 min); transcripts \approx 466 pages.
Platforms & tools	Google Meet, Zoom; webcam video; audio playback; chat; screen sharing; shared documents (Google Docs; itslearning); platform-integrated recording (itslearning); headsets/headphones.
Focus of observation	Material enactments of pronunciation instruction: articulation modeling, prosody/rhythm practice, corrective feedback; breakdowns/disruptions and domestic soundscapes.
Interview protocol	Post-lesson, semi-structured (Brinkmann, 2016): pedagogical choices, affordances/limitations, feedback strategies, perceived role shifts; contextual follow-ups via email.
Analytic approach (ANT)	Actor–Network Theory (Fenwick & Edwards, 2010; Latour, 2005): tracing relations and “following the actors”; symmetry of human/non-human; enactment (Mol, 2010); sociomaterial lens on teaching as enacted practice.
Coding & analysis workflow	Iterative cycles: open coding \rightarrow actor identification \rightarrow assemblage mapping \rightarrow scene construction \rightarrow ANT-informed interpretation; decisions documented in an audit trail.
Rigor, trustworthiness, & reflexivity	Triangulation (observations + interviews); analytic memo trail & audit trail; peer debriefing; thick description; reflexive memos addressing the author’s prior experience in adult L2 contexts.
Ethics	Swedish Research Council (2017) principles (informed consent, confidentiality, secure data handling); no teacher/student personal data recorded; right to withdraw.

About the Author

Andreas Lagerlöf is a PhD candidate in Pedagogy with a research interest in adult second language teaching. He has extensive experience teaching Swedish as a second language in municipal adult education as well as at the university level.

E-mail: andreas.lagerlof@liu.se

ORCID: <https://orcid.org/0009-0008-5014-2687>