



ELSEVIER

Contents lists available at ScienceDirect

## Linguistics and Education

journal homepage: [www.elsevier.com/locate/linged](http://www.elsevier.com/locate/linged)

# Building word knowledge through integrated vocabulary explanations in ESL tutorials

Derya Duran\*, Leila Kääntä

University of Jyväskylä, Department of Language and Communication Studies, Athenaeum, Building A, FI-40014, Jyväskylä, Finland

## ARTICLE INFO

### Article history:

Received 4 May 2022  
 Revised 3 April 2023  
 Accepted 4 April 2023  
 Available online xxx

### Keywords:

Vocabulary explanations  
 Pronunciation  
 ESL tutoring sessions  
 L2 learners  
 Conversation analysis

## ABSTRACT

This conversation analytic study explores how a tutor and tutees collaboratively negotiate word knowledge in English as a second language (ESL) tutorials. Specifically, we focus on integrated vocabulary explanations, where explaining the meaning and use of a word is intertwined with teaching its spelling and pronunciation. The data come from 18 hours of dyadic and multi-party tutorials held at an urban community college in the USA. The findings show that for the participants such aspects of word knowledge as orthography, pronunciation, associations, connotations, syntax and grammatical functions form an essential part of vocabulary work. Notably, we argue that negotiations of these integrated vocabulary explanations are realized through multimodal resources, drawing on linguistic, prosodic, embodied and material resources, among others. The findings highlight the unique context of these ESL tutorials that affords individualization and immediacy, enabling tutors to be responsive to the tutees' second language (L2) learning needs. The study also has implications for tutor education and classroom instruction with regard to the development of L2 word knowledge.

© 2023 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

## 1. Introduction

Small-group tutorial activities are a common instructional practice in higher education institutions and many tutoring settings have emerged to assist student learning. A typical context is language labs, i.e., places in which students seeking academic support are offered language-related assistance by first language users of the target language (Vick et al., 2015). The pedagogical goal is to help students, who complete their degrees in a second language (L2), acquire supplementary linguistic support and develop familiarity with the host country, thereby enhancing academic development and contributing to student success (Long, 2013). Today, there are different kinds of English as a second language (ESL) tutoring sessions, such as writing tutorials and conversation-for-learning tutorials. Depending on their focus, the language-related 'tutorial business' (Belhiah, 2009) conducted in these sessions varies from a form of 'free-talk' (e.g., Leyland & Riley, 2021) to having pre-established pedagogical goals (Kasper & Kim, 2015).

The data for our study originated from one such ESL tutoring setting. Yet it differs from other ESL tutorials in two ways. First, the primary goal in our research site was to support the linguistic needs of the L2 learners to improve *all* of their language skills. That is, the sessions aimed to promote the tutees' language com-

petence in speaking, reading, and writing through various instructional practices. Second, although the tutorials lacked formal lesson plans, they had pedagogical goals based on the tutees' specific needs, such as being able to complete homework assignments. A recurrent need addressed in our data is the learning of vocabulary that occurs via extended explanation sequences during which the tutor and the tutees collaboratively negotiate not only the *meaning* and *use* of words but also their *form*, all three constituting key aspects of word knowledge (e.g., Elgort & Nation, 2010; Nation, 2001).

According to Elgort and Nation (2010), who adopt a cognitive psycholinguistic view of L2 vocabulary learning, knowing a new word includes knowledge of collocation, of synonyms and antonyms, of syntactic conventions and of variation based on context and register. In addition, recognizing the way a word is spelled and pronounced is crucial (Nation, 2001). Miller (1999) also proposes that word knowledge involves more than meaning and pronunciation as the situational contexts in which a word is used to convey a particular meaning (contextual knowledge) are of importance in building vocabulary depth. In this regard, learning a new word in L2 entails the attainment of some of this knowledge on the part of the L2 learner. Since knowing a word involves multiple components, one problem with teaching vocabulary is that it can be approached through a variety of different views related to *what* to teach and *how* to teach it (e.g., Nation, 2001; Stoewer & Musk, 2019).

\* Corresponding author.

E-mail address: [deryadrnn@gmail.com](mailto:deryadrnn@gmail.com) (D. Duran).

Set against this background, our study investigates *integrated vocabulary explanations*, where semantics, phonology, orthography, morphology, and syntax are intertwined during L2 vocabulary explanations. As our method, we use Conversation Analysis for Second Language Acquisition (CA-SLA) that approaches language learning as a social practice, observable in the details of interaction (e.g., Kasper & Wagner, 2011). That is, we show how teaching and learning are socially accomplished in situ when lexical items are made interactionally relevant as teachables and learnables (Eskildsen & Majlesi, 2018) by the participants. This entails that vocabulary teaching and learning are analyzed as a context-dependent activity, designed specifically for and with the present participants in and through the emerging interaction. In the case of our data, integrated vocabulary explanations reflect and construct the pedagogical goals of the setting, where the participants engage in negotiating different aspects related to L2 competence for much of the tutoring session. This affords them flexibility in delving into those aspects of word knowledge the tutees make relevant or require for their L2 learning.

To our knowledge, the three aspects of word knowledge, i.e., form, meaning and use, have so far been studied separately within different fields of L2 learning, while there is no research on how they are taught in an integrated manner in naturally occurring instructional interaction (although see Stoewer & Musk, 2019). Although there is by now an extensive body of research on L2 vocabulary explanations in different instructional settings (see below for an overview), such studies have focused on participants' verbal, embodied and material practices of explaining word *meaning* but have to a lesser extent addressed other aspects of word knowledge (cf. Mortensen, 2011; Stoewer & Musk, 2019). On the other hand, there is also some research on pronunciation instruction that has investigated teachers' practices of teaching word pronunciation (see below), but more is needed to fully understand how pronunciation is taught to L2 learners in different instructional settings. By bringing together the previously separate fields of L2 learning research (i.e., L2 tutorials, vocabulary explanations and pronunciation instruction), our study not only provides novel insights into the varied resources participants employ in learning L2 vocabulary, but also offers new insights on how word knowledge is constructed as part of instructional interactions. Specifically, our study seeks to answer the following research questions:

- 1) How do the participants collaboratively negotiate the form, meaning and use of lexical items in an integrated manner?
- 2) What kinds of learning opportunities do integrated vocabulary explanations offer for the L2 learners?

## 2. Literature review

### 2.1. Tutoring discourse as an interactional environment for L2 learning

As mentioned above, there are different kinds of ESL tutoring sessions, including writing-tutorials and conversation-for-learning tutorials, both of which have substantially been investigated. Such studies have provided a well-established account of interactional processes of tutorial discourse (e.g., see Leyland, 2018, 2020, 2021 for writing tutorials; Belhiah, 2009, 2013; Kasper & Kim, 2015; Seo, 2021; Seo & Koshik, 2010 for conversation-for-learning tutorials). Considering all kinds of tutorial sessions, it is crucial that tutors draw on various interactional resources both to manage relations with tutees and to facilitate the establishment of learning opportunities (Long, 2013). Therefore, besides talk, the role of embodied and material resources in creating language learning opportunities has received increasing interest in L2 tutorial contexts (e.g., Belhiah, 2009; Lilja, 2014; Park, 2019;

Seo, 2021; Seo & Koshik, 2010). For example, Belhiah (2013) has investigated tutors' gestures and their interactional roles in vocabulary explanations during reading activities in dyadic tutoring encounters. Three types of functions that gestures serve were identified: (1) reinforcing the meaning of verbal utterances, (2) disambiguating the meaning of lexical items, and (3) establishing gestural cohesion across turns at talk. These observations led the author to argue that recipients also employ gestures as a resource to display alignment and understanding through gesture replication and gesture co-production.

Tutors' embodied conduct may also problematize various aspects of the tutee's talk (Seo & Koshik, 2010) or writing assignments (Park, 2019). For example, Seo and Koshik (2010) investigated a tutor's embodied gesture work that engendered repair in ESL conversational tutoring sessions. The authors describe how facial displays and two types of gestures, i.e., head tilts and head pokes, are used by the tutor as embodied resources for initiating repair, thus establishing mutual understanding. Seo (2021) has also shown how tutees employ gestures with a candidate solution to their word searches, yet produce a mismatch between the solution and the concomitant gestures. The author argues that producing candidate understandings plays an important role in creating learning opportunities for tutees when tutors respond to the discrepancy between their speech and gesture.

Alongside gestures, tutors utilize situated objects in constructing repair and coordinating understanding (see Ro, 2021 for the use of PowerPoint slides to resolve understanding problems and Leyland & Riley, 2021 for the use of notes to initiate deferred correction sequences), and thus manage to pinpoint students' L2 knowledge gaps through a series of embodied and verbal actions along with textual resources. For example, Kim and Cho (2017) examined one L2 writing tutor's use of gestures and physical materials (pen and paper) to scaffold the writing of a student with low L2 proficiency. Their study suggests that the tutors can engage their tutees more by utilizing the materials and accompanying gestures, thus promoting intersubjectivity, and eventually supporting the tutees' writing performance.

While the above-mentioned studies have reported relevant findings in terms of how verbal, embodied and material resources are used to do 'tutorial business' (Belhiah, 2009), they have not specifically addressed participants' orientation to enhance L2 learners' language competence from the perspective of word knowledge. By focusing on a specific type of ESL tutorial setting, of which there is relatively little research on (although see Duran & Jakonen, 2022), our study contributes to previous CA research on small-group tutorial sessions by providing new insights on how the participants collaboratively accomplish in situ the teaching and learning of word knowledge.

### 2.2. Vocabulary explanations as collaborative and multimodal achievements

Teachers' vocabulary explanations have substantially been investigated in instructed L2 settings (e.g., Koç & Ergül, 2023; Käätä et al., 2018; Majlesi, 2014; Mortensen, 2011; Morton, 2015; Sert, 2017; Tai & Brandt, 2018; Tai & Khabbzbashi, 2019) and in L2 tutorial contexts (e.g., Belhiah, 2013; Mori & Hayashi, 2006; Lilja, 2014). Such studies have demonstrated that vocabulary explanations are often unplanned as they emerge from the local sequential contingencies of interaction. They are also collaborative by nature as L2 learners in different ways participate in them, for example, by requesting for an explanation or clarification of word meaning (e.g., Käätä & Kasper, 2018; Sert, 2017; Tai & Brandt, 2018; Stoewer & Musk, 2019) or providing candidate explanations when asked for one by teachers (e.g., Mortensen, 2011). In addition, such studies have shown that vocabulary explanations are multimodal,

i.e., teachers not only use linguistic means but also embodied resources and the spatio-material ecology of the instructional setting.

The verbal practices that have explicitly been reported include, for instance, translating (e.g., [Mortensen, 2011](#); [Morton, 2015](#); [Stoewer & Musk, 2019](#)), providing homonyms ([Morton, 2015](#); [Stoewer & Musk, 2019](#)) and synonyms ([Waring et al., 2013](#)), and paraphrasing ([Stoewer & Musk, 2019](#)). According to [Waring et al. \(2013\)](#), teachers use two types of vocabulary explanations: an analytic approach that draws on verbal and textual resources and an animated approach that entails a variety of multimodal resources, including paralinguistic, gestural and other bodily actions. Teachers in their data contextualize the focal words and use understanding check questions after the vocabulary explanations to secure intersubjectivity. While contextualization in analytic explanations emerges through such practices as placing a word in a sentence, in animated explanations, contextualization occurs through gestures and scene enactments.

The findings on embodied conduct highlight the creative ways teachers employ gestures and different kinds of enactments to explain word meaning. Several studies demonstrate that teachers' gestures are salient elements for L2 learners as they not only enhance students' understanding (e.g., [Kääntä et al., 2018](#); [Sert, 2017](#); [Tai & Brandt, 2018](#); [Waring et al., 2013](#)) but also serve as resources for them when learners recycle and modify the teacher's gestures in different ways to show understanding (e.g., [Belhiah, 2013](#); [Majlesi, 2014](#); [Eskildsen & Wagner, 2013](#); [Smotrova & Lantolf, 2013](#)). In terms of enactments, [Waring et al. \(2013\)](#) have shown how teachers produce scene enactments that include both verbal and gestural components to convey meaning and that are performed either as stand-alone actions or as illustrations of an already given explanation. [Tai and Brandt \(2018\)](#) have also examined how participants in an adult beginning level ESOL classroom in the US employ embodied enactments, including gesture, body movement and verbal resources, to demonstrate an aspect of hypothetical events. The authors maintain that the concept of "embodied enactment" differs from the term "embodied explanations" ([Sert, 2017](#)) as the latter has conceptualized enactments as a supplement to the teacher's verbal responses rather than as "a distinct form of interactional and embodied conduct" ([Tai & Brandt, 2018](#), p. 249). Thus, [Tai and Brandt's \(2018\)](#) study highlights that the use of embodied enactments is an essential practice for teachers to demonstrate the situationally and pragmatically appropriate use of L2 lexical items.

A key difference between the current study and the above studies is its focus on the different aspects of word knowledge that the participants establish as relevant for the tutees' L2 learning. To date, only [Stoewer and Musk's study \(2019\)](#) on advanced English as heritage language lessons has highlighted the different aspects of word knowledge in content-sensitive ways so that form, meaning, and use are present in vocabulary explanations but reflexively developed depending on the interactional contingencies. In the other studies cited, this aspect has not been captured, although they have noted how teachers in addition to explaining the semantic aspect of the target word also attend to its form, often by writing it on the board and/or by carefully articulating its pronunciation (e.g., [Mortensen, 2011](#); [Stoewer & Musk, 2019](#)). As a result, the treatment of integrated vocabulary explanations is still under-investigated, and as the analysis will show, the participants address such issues as knowledge of association, grammatical functions, syntax, and meaning in an integrated manner when collaboratively negotiating word knowledge during ESL tutorials, including issues of spelling and pronunciation to which we turn next.

### 2.3. Pronunciation instruction as part of vocabulary teaching

In our data, the participants recurrently engage in teaching and learning pronunciation of words during the extended vocabulary

explanation sequences, and thus show their orientation to form as an essential part of word knowledge. This is in contrast to previous CA studies on different L2 settings that have shown that pronunciation sequences in interaction put the main business of the interaction on hold (e.g., [Brouwer, 2004](#); [Mori, 2004](#); [Park, 2007](#)). Considering classroom-based studies, there is some research evidence on the role of speech rhythm ([Szczeppek Reed, 2012](#)) and on teachers' embodied actions (e.g., gestures and postures) in pronunciation instruction. [Smotrova \(2017\)](#), for example, argues that embodied actions are powerful teaching tools for making abstract concepts visible and concrete, thereby enhancing students' understanding of pronunciation instructions. Her findings indicate that the teacher's reiterative gestures, i.e. catchments, facilitate the production of such suprasegmental features as word stress and speech rhythm. [Nguyen \(2016\)](#) also shows that teacher's pointing gestures and demonstrations of how and where sounds are articulated help students see and practice target sounds better.

Altogether, previous literature on L2 pronunciation instruction has tended to focus more on basic philosophy of pronunciation and core skills to teach the pronunciation of the target language (e.g., [Brown, 2014](#)), while the basic question of *how* pronunciation teaching is accomplished has largely been unanswered as only few studies to date have addressed it as a local and situated practice in classroom environments, that is, within interactions from a participant-relevant viewpoint ([Kääntä, 2017](#); [Nguyen, 2016](#); [Szczeppek Reed, 2012](#); [Smotrova, 2017](#)). The current study fills in this research gap by demonstrating how pronunciation teaching and learning are collaboratively achieved by the participants in this specific ESL tutorial setting, and how they are part and parcel of the negotiation of word knowledge, not a side sequence that puts the interaction on hold. In the course of the analysis, we will show that teaching and learning to pronounce words are intertwined with knowing how to spell them, which helps the tutees understand how to pronounce the lexical items under focus.

### 3. Data and method

The data are drawn from 18 hours of ESL tutorial sessions video-recorded at an urban community college on the east coast of the United States, involving one English native speaker tutor and seven English non-native speaker tutees. The main aim of the tutoring sessions was to assist students who experience the need to have additional linguistic and academic support in English (speaking, reading, grammar, etc.) to become more proficient in the language. The sessions thus function as an institutional support service to those students who voluntarily participate in the tutorials. To address the needs of increasing numbers of adult ESL students in the US, community colleges provide easy access to foreign students and immigrants who need to enhance the transition to academic life and American culture through overcoming the language barrier with the curriculum and services provided in these educational settings. The current research site, like most community colleges in large metropolitan areas, is diverse in terms of race, ethnicity, and age due to their 'open door policy.'

The tutoring sessions in the current dataset had pedagogical goals, and thus differ from other small-group learning activities organized for L2 speakers such as 'conversation tables' (see e.g., [Mori & Hayashi, 2006](#)) or 'Enhanced Conversations-for-Learning' (see e.g., [Leyland & Riley, 2021](#)), the purpose of which is typically to offer L2 speakers non-instructed opportunities for spoken interaction with native speakers of the target language. More specifically, there was often an overt goal-setting phase at the beginning of each session during which either the tutee(s) set the agenda through seeking assistance for a particular task (i.e. writing assignments, grammar exercises, comprehension of reading texts) or if the tutee(s) did not hold a particular agenda, the tutor provided



Fig. 1. Tutees with the tutor (middle) in the Learning Resource Center.

suggestions to work on and provided the materials. In this regard, the sessions were not about ‘free-talk’ since the tutor provided service to the tutees who were seeking help for language learning.

Before the videotaping began, the tutees filled out survey forms, which collected information on their gender, first language, length of stay in the US, and prior language learning experience. They came from Bangladesh, Belarus, Ecuador, Senegal, Ukraine and Thailand, and were aged between 18 and 34. Their English proficiency levels varied considerably as is illustrated in the extracts. We did not have information on their actual language competence test results; however, their length of residence in the US ranged between 1.5–6 years at the time of data collection. The tutees received tutoring through weekly scheduled appointments (roughly one hour) with the same tutor, which qualifies them as regular tutorial attendees. At the time of data gathering, the tutor, hired as a full-time language expert, was working in the Learning Resource Center, where the sessions were held, for 1.5 years, teaching ESL writing, conversational practice and exam practices. She was about to complete her master’s degree in TESOL for adults at a US university. The data collection covers a period of 5 weeks during spring 2019. The data include both tutoring dyads and multi-party tutoring interaction, with no more than four students with a single tutor in case of small groups. The sessions took place in the open floor plan tutoring space, i.e. a large seating area with several tables for everyone’s use. In other words, they were private conferences but visible to everyone else in the tutoring area (see Fig. 1). Each participant signed an informed consent form, approved by the institutional review board of the community college. In the extracts, all proper names have been altered to secure the participants’ anonymity. We also blurred the face of those participants who requested it as part of informed consent.

The study employs multimodal CA as the methodological framework (Clift, 2016) for its analysis. A basic tenet of CA is that any detail of interaction can be potentially pertinent to the participants, which thus involves fine-grained analysis of talk and embodied conduct in naturally-occurring situations. CA also aims at understanding the interaction from the participants’ own perspectives with a focus on what they actually do and display. We followed a canonical guideline for doing a CA analysis. First, we watched the videos multiple times to familiarize ourselves with the data and transcribed the whole corpus following the conventions developed by Jefferson (2004, see Appendix). The video recordings were then subjected to a more fine-grained analysis and focal embodied actions relevant to action ascription were transcribed by applying and adapting Mondada’s (2014) conventions. We made several preliminary observations on the transcripts and identified an interesting phenomenon: extended vocabulary explanation sequences where the negotiation of form, meaning and use were integrated. We paid close attention to the sequential con-

texts and participants’ orientations of each case in the data, after which we built a collection of integrated vocabulary explanations, re-transcribed the extracts, and made detailed analytical descriptions. Next, we present the findings through two extended explanation sequences that represent the participants’ recurrent practices of negotiating word knowledge in the data.

#### 4. Data analysis

The analysis focuses on how the participants engage in collaborative work to negotiate the different aspects of word knowledge through integrated vocabulary explanations. For clarity of presentation, however, we have divided the analysis into two sections. The first section centers on the tutor’s multimodal practices of explaining the meaning and use of focal words to the tutees in response to their candidate understandings, i.e. their interpretations of how they understand a particular lexical item (Heritage, 1984). The second section describes how the tutor and the tutees deal with the tutees’ articulation problems that emerge during the meaning-focused segments and that the tutor orients to as trouble sources in need of repair. In doing so, the tutor integrates the teaching of the pronunciation and spelling of the focal words into the explanation sequences. Yet, as the analysis shows, the treatment of form also entails meaning-related work, and vice versa, the treatment of meaning entails form-related work that helps ensure the tutees’ learning of the focal words. In both sections, the analysis demonstrates how the tutees engage in requesting clarification, modeling the tutor’s pronunciation and writing their own notes, thereby rendering the focal words as objects of learning, i.e., as teachables and learnables (Eskildsen & Majlesi, 2018).

##### 4.1. Negotiation of L2 word meaning and use as part of integrated vocabulary explanations

Since agenda-setting is collaboratively negotiated in these tutoring sessions and since the focus is on improving the tutees’ L2 competence, they are given space to steer the interaction and to create an interactional space for extended negotiation of meaning and use of words they have trouble with. The tutees thus often produce candidate understandings of word meanings that they are not sure of, which makes a (dis)confirmation from the tutor relevant. However, the tutor does not simply (dis)confirm them, but elaborates upon and clarifies them in various ways, thereby promoting the tutees’ vocabulary depth through inducing their awareness of different aspects of word knowledge, as is illustrated in Extracts 1–3.

Extract 1 illustrates how a tutee produces a candidate understanding, seeking confirmation through word usage which demon-

## Extract 1. Bizarre.

- 01 T ^yeah it's bizarre^  
tga ^gaze down to her left side^
- 02 (0.4) Ira gaze to notebook  
T glances toward Uma on her right
- 03 Ira ^m.ah hahm #what?  
irage mhead throw upm  
iraga #gaze to T-->  
tga ^gaze in front of her->
- 04 T it's biza::^rre=  
tga ^gaze to Ira->
- 05 Ira =it's like so(ci)=
- 06-> T =^stra::nge^  
tf ^raises eyebrows^
- 07 Ira [m#ohh#  
08-> T [^bizarre #means like <very strange>  
iraf mraises eyebrowsm  
iraga #glances in front of her#  
iraga #gaze to table->  
tga ^gaze shift toward Uma->

(34 lines omitted: T teaches pronunciation, see Extract 5)

- 43 Uma we can use this \$one like fo- er:  
umaga -----\$gaze to T->  
tga to her hair->  
irage --writing->

- 44 ~ (0.8) in person; ~  
umage ~hand gesture to right~

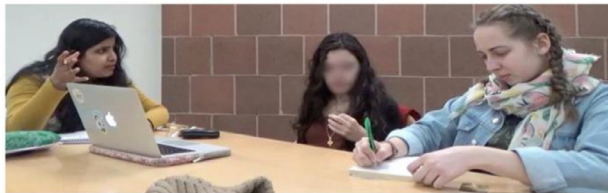


Fig. 1A (from left to right - Uma, Tutor, Ira)

- 45 (3.3) Uma gaze to T, moving hand in the air  
T gaze to her hair, touching it  
then T shifts gaze to Uma

- 46 Uma #\*like this\* \$word;  
umage \*points w/ pen at her notebook\*  
umaga \$gaze to notebook\$gaze to T->  
iraga #gaze to Uma/T

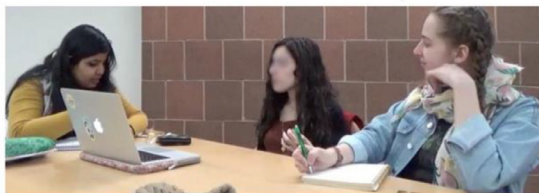


Fig. 1B



Fig. 1C

## Extract 1. Continued

- 47 T .hh >you coul- to describe a person?<
- 48 Uma ^\*mh-huh\*  
 umage \*nods--\*  
 tga ^gaze in front of her->
- 49 T \$yeah you could say like @yeah\$  
 umaga \$gaze to notebook-----\$
- 50-> \$he is a bizarre ^"professor@"\$ (0.4) like^  
 tga ^gaze to Uma-----^  
 tf "raises eyebrows"  
 umaga \$glances between notebook & T \$gaze to Ira->
- 51 (1.4) Uma laughs & sticks out tongue, then covers her face  
 Ira laughs, gaze to Uma  
 T gaze in front of her



Fig. 1D



Fig. 1E

- 52 Ira [#£like weird?£]  
 53-> T [\$my f- my my ] ^my friend\$ is acting quite bi\$za:rre  
 tga ^gaze to Uma->  
 umaga \$glances btw notebook & T \$gaze to notebook \$gaze to T->  
 iraga #gaze to T->
- 54 (0.4)
- 55 Uma [\*ohh\*]  
 56-> T [it's an adjective.  
 umage \*nods\*
- 57 Uma ^\*oh (0.2)\* #°it's an adjective.°  
 umage \*nods----\* \$gaze to notebook->  
 umaga \$gaze to notebook->  
 tga ^gaze in front of her->  
 iraga #gaze to notebook->>
- 58 (2.0) Uma & Ira write in their notebooks
- 59-> T it's not necessarily like a ↓bad thing (0.7)
- 60-> \$it doesn't have to have a ^negative connotation .hh  
 umaga \$gaze to T->>  
 tga ^gaze to Uma->>

strates her awareness of what aspects are relevant to inquire about when encountering a new word. The tutor, in turn, employs different practices in responding to the candidate understanding, including providing synonyms, explaining the part of speech the focal word belongs to and its connotations, and contextualizing its use. The example comes from a multi-party encounter in which the tutees, a Ukrainian female named Ira and a Bangladeshi female named Uma, have been working on a task on transportation prepositions in English. Prior to the exchange, the tutor (T) has told how 'weird' English can become as there is no logical reason for preposition usage. In line 01, she repeats it again, but replaces 'weird' with 'bizarre.' In all the extracts, the turns of particular importance are marked with an arrow to draw the reader's attention to them.

T's repetition in line 01 induces a repair sequence during which she explains 'bizarre' through another synonym. At first, when Ira produces the open-class repair initiator (l. 03), preceded by laughter particles and embodied actions (i.e. head throw up and gaze at T), T orients to it as a hearing problem as she redoes her turn verbatim, yet in a prosodically marked way lengthening the second syllable (l. 04). Note that such a repetition leads to a further repair initiation by Ira (l. 05), and thus shows how the verbatim repetition does not resolve her understanding problem. Ira, however, cuts off her turn as T introduces 'strange' as a synonym to 'bizarre' along with raised eyebrows (l. 06). With this turn, T also addresses the meaning problem Ira has with 'bizarre'. In response, Ira indicates understanding via the news-receipt '↑ohh' (Heritage, 1984) and raised eyebrows. Despite Ira's claim, T produces an explanation



Fig. 2. Tutor, Ira, Uma and Liz.

in the 'X means Y' format (l. 08), thus treating these two lexical items as being synonymous. Instead of probing the meaning further here, the interaction turns into an extended segment of pronunciation teaching of 'bizarre' (see [Extract 5](#)).

Once the pronunciation teaching sequence is completed, the negotiation of the focal word moves to its use, at which point T contextualizes it by providing example sentences ([Stoewer & Musk, 2019](#); [Waring et al., 2013](#)). The negotiation begins when Uma checks her understanding of how to use the word (l. 43–44). Her turn design indexes trouble in formulating her idea in L2: multiple speech perturbations (cut off, hesitation, pause) and gesturing project a word search that she eventually resolves herself (*in person*) (see [Fig. 1A](#)). The declarative turn format invokes the tutor's right to (dis)confirm the candidate understanding and the use of the modal auxiliary 'can' indexes Uma's orientation to the epistemic asymmetry between the participants. Having produced the candidate understanding, Uma gazes at T expecting her to answer (l. 45) but T is fiddling her hair, not orienting to Uma. Since Uma's response pursuit is not fulfilled, she adds an increment (l. 46) to clarify the referent (*bizarre*) (see [Fig. 1B & 1C](#)). When T responds, she first starts to confirm Uma's understanding (*you could*–) but then produces a clarification request (l. 47) that, on the one hand, seems to address Uma's formulation '*in person*' to ensure that T understood what it means (*to describe a person?*) and, on the other hand, to initiate embedded correction on the formulation. After Uma has affirmed T's request both verbally and through nodding (l. 48), T not only confirms Uma's understanding but also elaborates upon it with an example that is produced in an animated voice and the raise of her eyebrows (l. 49–50). The raising of the eyebrows that accompanies the noun '*professor*' helps create a humorous frame (cf. [Pomerantz, 2019](#)) with which the tutees align. Namely, both Uma and Ira laugh, briefly glancing at each other (l. 51). Uma also sticks out her tongue and covers her face, looking at Ira as if sharing a joke related to the example (see [Fig. 1D & E](#)). Under laughter, Ira also produces a candidate understanding in the form of a confirmation check (l. 52), but it is not oriented to by T. Instead, to further contextualize, T gives another example sentence featuring the word (l. 53).

The two usage examples (l. 50 & 53) T provides tap onto Uma's candidate understanding that '*bizarre*' can be used to describe a person. They also help specify the grammatical patterns '*bizarre*' can fit into, i.e. as a descriptive modifier in a noun phrase or as the head of an adjective phrase. When T adds knowledge of the syntactic category of the word (l. 56), she attends to this explicitly. Her addition receives a news-receipt and a quiet verbatim repeat from Uma (l. 57) that manifests that she is processing the information. More importantly, both Uma and Ira orient to their notebooks for note taking (l. 58), treating the new knowledge as learnable that helps them categorize the word. T subsequently brings up the

connotation of the adjective, thus helping students classify words according to the connotations they might have (l. 59–60), in which intonation plays an important part (not shown in transcript).

[Extract 2](#) shows how an extended meaning negotiation sequence emerges when a tutee seeks confirmation to her understanding of two words, that she falsely appears to conflate as being the same ('*emerge*' and '*immersion*'). This misunderstanding surfaces in the context of the participants having discussed the meaning of '*merger*', a third word. The analysis demonstrates how the participants shift between orientation to the meaning of the focal words and their spelling due to the tutee's misunderstanding. It also illustrates the tutor's practices of spelling and writing out the words to establish their difference and the participants' utilization of reciprocal gestures that accompany the verbal explanations to secure intersubjectivity. Prior to the interaction, T and two tutees (Uma and Ira) (see [Fig. 2](#)) have been reviewing a task in which the noun '*merger*' appeared. Shortly after, T checks whether the tutees know its meaning through a vocabulary check ('*you know what merger means?*') and subsequently provides a contextual definition when the tutees have expressed their unfamiliarity with it (see [Duran & Jakonen, 2022](#)). It is at this point in line 01 that Uma produces a candidate understanding of '*emerge*' linking it with '*merger*', most likely due to the words sharing phonological features.

In dealing with the tutee's understanding problem, the tutor employs an animated approach ([Waring et al., 2013](#)), using both talk and depictive gestures to explain the focal words. When Uma produces the candidate understanding of '*emerge*' (l. 01), using the '*it's like X?*' format typical for our data, she simultaneously aligns her hands actively bringing them closer and further away from each other (see [Fig. 2A](#)). She thus partly replicates the gesture T has produced only moments ago when explaining '*merger*', yet creating a mismatch between her speech and gesture ([Seo, 2021](#)). Both the speech-gesture mismatch and the erroneous equating of the words occasion a reiteration of the meaning of '*merge*' from T. The reiteration is performed both verbally and gesturally (l. 02, see [Fig. 2B](#)), which receives an embodied acknowledgement from Uma (l. 03). To secure understanding, Uma produces a designedly incomplete turn-constructural unit ([Koshik, 2002](#)) that prompts T to provide the meaning for '*emerge*'. T responds by accompanying the verbal explanation with a gesture that depicts the '*coming out*' of something (l. 04, see [Fig. 2C & 2D](#)), thus serving to reinforce the meaning of the word ([Belhiah, 2013](#)). Uma's inaudible response, followed by a *sotto voce* 'okay' and the adjustment of her body position (l. 05), is treated by T as signaling continued trouble in understanding. In a further attempt to resolve the problem, T verbally specifies the difference between the two verbs, while moving her hand to the right side to reinforce the difference and prosodically marking it in the word stress (l. 06–07). Uma quietly

## Extract 2. Emergent program.

- 01 Uma ^\$~it's like (.) **\*like** emerger, (0.4) the same?~  
 umaga \$gaze to T-->  
 umaf ~frowns-----  
 umage \*aligns hands moving them closer & apart again\*  
 tga ^gaze toward Uma-->



Fig. 2A (Uma)



Fig. 2B (Tutor)

- 02 T **+<me:rge>** i' means to come together+  
 tge +clenches hands-----+
- 03 Uma \*°uh huh°\* (.) **\*emerge** means;~\*  
 umage \*nods-----\*  
 umage \*brings hands closer to each other\*

- 04 T **emerge+** means ++**to** come out+  
 tge +moves hands close to chest+  
 tge +moves hands away from chest+



Fig. 2C



Fig. 2D

- 05 Uma \$(-) °okay°  
 umab \$leans back in her chair, away from table->
- 06 T **+different.+**  
 tge +rh palm down, moves up-down+
- 07 \$"+merge and++ emerge+" are different words=  
 tf "raises eyebrows-----"  
 tge +rh palm faced down toward Uma+  
 tge +rh palm toward right+  
 umab \$leans toward table->
- 08 Uma =°okay°\$  
 umab \$leans back in her chair, away from table\$  
 umaga \$gaze shift toward the entry->
- 09 (2.6) Uma adjusts sitting position, gaze to the entry  
 T gaze in front of her
- 10 Uma ^~\$then why this~(.) to to like~ emergent program?  
 umab \$leans toward table->  
 umaga ~gaze shift to T~gaze to entry~gaze to T-->  
 tga ^gaze to Uma-->

## Extract 2. Continued

- 11 T no that's "<im|ersion.>"  
tf "raises eyebrows"
- 12 (0.4) Uma leans away from table
- 13 Uma **emergen=**
- 14-> T =>e \*m e r g |e\*< is not "(1.4) i m m e" r s i o n.  
umab \$leans toward table->  
umaf \*nods-----\*  
tf "thinking face"
- 15 Uma §°(---)°  
umab \$leans away from table->
- 16 T one is spelt ~like ^+°can\* I have your pencil?°+\*  
tga ^gaze to Uma's notebook-->  
tge +points toward Uma's pencil+  
umage \*slides notebook over, hands pen\*  
umaga ~gaze to notebook->
- 17-> T +\$one is spelt+ +(4.7)  
tge +takes pencil & notebook+  
tge +writes on notebook->  
umab \$leans toward table, closer to notebook->
- 18-> ~emerge~ means to come out of (3.7) "+~like (3.1)"  
tf "thinking face"  
tge +stops writing->  
umaga ~glance at T~gaze to notebook-----~gaze to T-->
- 19 >try to find an example<° +immersion means when you're (1.2)  
tge +writes on notebook->
- 20-> when^ you're +constantly+^exposed.+  
tga ^gaze to Uma->>  
tge +moves hands in front of face+  
tge +lateral hand gesture over face+



Fig. 2E

claims understanding and leans back (l. 8), which seems to indicate sequence closure.

However, Uma's underlying problem seems not to have been solved yet, which occasions T to utilize spelling and further explanations to clarify the difference. That Uma is still trying to figure out something becomes clear when she first adjusts her position and gazes toward the entry during the 2.6 second silence (l. 09), and then inquires about the word 'emergent' that she links to the concept of 'immersion program' (l. 10), announcements of which are on notice boards at the entrance of the tutoring area (see Fig. 1 in Section 3). T responds with an unmitigated 'no' and specifies what Uma has possibly referred to as 'immersion' (l. 11), with accentuating the word through emphasis and by raising her

eyebrows. Next, Uma attempts to repeat the word 'immersion' but manages to say 'emergen' (l. 13), thus displaying how she heard it. Instead of orienting to the mispronunciation, T focuses on Uma's misunderstanding, as next she spells 'emerge' and projects an upcoming contrast that she puts on hold when thinking of how 'immersion' is spelled (l. 14). When she figures it out, she spells 'immersion'. Uma's response is inaudible (l. 15), but it is treated by T as still not having solved the problem, since T launches an explanation specifying the differences between the words by writing them down (l. 17–19) and repeating the meaning for 'emerge' (l. 18). For the meaning of 'immersion', T verbalizes her search for examples (l. 19), finally explaining it in terms of being constantly exposed and performing a concurrent depictive gesture in front of

her head (l. 20, see Fig. 2E). The sequence continues with T reiterating the explanation of ‘immersion’ using several resources: she prosodically emphasizes the salient words (also Mortensen, 2011), repeats them a couple of times and recycles the depictive gestures to resolve Uma’s understanding problem (not shown here).

Extract 3 differs from Extracts 1 and 2 in that the tutee (Ira, see Fig. 3) actively probes for a precise meaning through multiple candidate understandings, thus displaying some epistemic access to the word meaning and desires to learn it. This in turn influences how the tutor engages in explaining the word meaning through both the analytic and animated approaches (Waring et al., 2013). The extract comes from a reading activity where T has reviewed potentially unknown words from the text. Prior to the interaction, she has checked if the tutees know the meaning of the word ‘solitude’. Since both Uma and Ira have indicated unfamiliarity with the word, T has given an explanation (i.e. ‘being alone’) and also engaged in teaching how it is pronounced (not shown in this extract).

Ira’s first candidate understanding displays uncertainty as she engages in a word search to express her comprehension (l. 01). In her response, T explains ‘introverts’ first (l. 03) and repeats the meaning of ‘solitude’ (l. 04), thus establishing the difference between the words. The gesture that accompanies the noun depicts the act of being alone (see Fig. 3A), which serves to disambiguate its meaning from the adjective (Belhiah, 2013). Ira’s first candidate understanding having now been rejected, she produces another (l. 06), constructed in an alternative question format, which she cuts off before the alternative. Despite Ira’s unfinished turn, in line 07, T rejects the second suggestion and orients to finding another way to express the meaning of ‘solitude’, indicated by T’s multiple sayings of ‘just’ that partly overlap with Ira’s third candidate understanding (l. 08). In the end, T revises her original formulation by explaining the meaning as having time alone (l. 09).

Instead of reacting to T’s revised explanation, Ira repeats her third candidate understanding (l. 11), the wording of which T uptakes (=choose to) in her reiteration of the two previous explanations (l. 12–13) and repetition of the face-covering gesture (l. 13). A 2.5 second silence emerges during which T is waiting for Ira to respond but no uptake is forthcoming as Ira shifts her gaze down biting her lip, thereby signaling a potential problem with T’s repair (Seo & Koshik, 2010). Realizing this, T attends to the computer to find the meaning through online resources (l. 15–16). Having reflected on T’s explanations (l. 17; see Fig. 3B), Ira provides a fourth candidate understanding, combining the explanations previously presented by both T (being alone) and herself (on your own choice) (l. 18). Her turn is try-marked with a rising intonation, which displays that Ira is seeking for confirmation. Yet it seems to mark her candidate as disaffiliative, i.e., it “targets no obvious problem and recycles known information” (Antaki, 2012, p.

544), the interpretation of which is emphasized through the tapping gesture on her notebook. T orients to this disaffiliativeness when she ratifies Ira’s formulation as correct (l. 20). However, the confirmation is mitigated in several ways and projects a disagreement through a ‘but’ clause that mainly restates T’s original explanation with emphasis on the word ‘alone’ and another repetition of the face-covering gesture (l. 21).

To secure a shared understanding, T shifts attention back to the computer, still searching for the word meaning (l. 23–26), and provides an explanation (l. 27–28). Ira produces a simple change of state token and T subsequently formulates a so-prefaced upshot (l. 32), essentially repeating what Ira has proposed already in line 18. Thus, in addition to offering her own explanations accompanied with depictive gestures, T also draws on online resources to resolve the understanding problem and to acknowledge the tutee’s contribution.

Altogether, this section has demonstrated how the tutees’ candidate understandings, generally in the form of yes/no questions or declaratives, manifest in different ways their epistemic asymmetry with regards to the meaning of the focal words. Various lexical devices, such as modal auxiliaries, the qualifier ‘like’ and other hedges, as well as hesitation markers and pauses as signals of word search are employed to display uncertainty viz. the meaning of the words. In view of word knowledge, the candidate understandings seek confirmation through synonyms (Ext. 2 ‘emerge’ and Ext. 3 ‘introvert’) and meaning extensions (Ext. 1 ‘we can use this in person’), which demonstrate the tutees’ awareness of what aspects are relevant to inquire about when encountering unfamiliar words. The tutor in (dis)confirming their understanding builds on this and employs different practices in teaching not only the meaning of the focal words but also their form and usage. In her explanations, she spells and writes out the words (Ext. 2 ‘emerge’ and ‘immersion’), provides synonyms (Ext. 1 ‘bizarre’ -> ‘strange’), and identifies the part of speech the words belong to (Ext. 1 ‘bizarre’ -> an adjective) and provides usage examples (Ext. 2 ‘bizarre professor’). She also adds information about the connotations (Ext. 2 bizarre not being a negative word) and how intonation influences the pragmatic meaning of words when used in context. More importantly, she uses an animated approach to explain the word meanings, often with environmentally coupled gestures (Waring et al. 2013) that help depict the core idea of the word meaning (Ext. 2 ‘merger’ and Ext. 3 ‘being alone’).

#### 4.2. Pronunciation instruction as part of integrated vocabulary explanations

For learners whose L1 and L2 have different sound-letter correspondences and who have not learned to recognize the orthographic form of the target word nor to decode it phonolog-



Fig. 3. Tutor, Uma and Ira.

## Extract 3. Solitude.

01 Ira #like this erm: (0.9) .ts# introverts?^  
 iraga #gaze in front of her---#gaze to T-> (1. 13)  
 tga -gaze down to hair-----^

02 (1.4) T gaze up in front of her

03 T that's more like ^,shy (0.7) ^but  
 tga ^gaze to Ira^gaze up

04-> solitude is like **the** act of ^being alone.+  
 tge +both hand gesture-----+  
 tga in the air-----^gaze to Ira->

05 (0.9)

06 Ira it mea:n «to be betrayed?» (0.2) or to be::=  
 iraf #raises eyebrows--#

07 T =>not to be betrayed ^just just s- [just just]^  
 tga ^gaze up-----^

08 Ira [to choose]

09-> T ^have alone time.  
 tga ^gaze to Ira->

10 (0.9)

11 Ira to choose [to (be)  
 12-> T [^to choose to kind of when you are alone  
 tga ^gaze in front of her-->

13 (0.8) and you are +having ^time alone.+#  
 tga ^gaze to Ira-->  
 tge +hand gesture in front of face+  
 iraga -->#

14 (2.5) Ira gaze down to her right, biting her lip->

15-> T ^I can find an exact definition (0.6)  
 tga ^gaze to laptop-> (1. 21)

16 iraga if that helps °(---) # (one second)°  
 #gaze down to notebook->

17 (4.5) T types on the laptop  
 Ira gaze down to notebook, then does a **thinking face**

18 Ira #like (0.4) \*being alone on your own †choice?#\*  
 iraga #gaze toward T-->  
 irage \*taps notebook with rh fingers at each word\*

19 (1.3) T searches on the laptop

20 T uhmm:: (0.6) it can mean that (0.5) but

21 it's basically just the ^idea +of being #alone^+  
 tga ^gaze to Ira-----^  
 tge +hands gesture in front of face+  
 iraga #gaze down to table->

22 (0.6) T gaze back to laptop

23 T #let me see.  
 iraga #gaze to T's laptop screen-> (1. 32)

24 (4.3) T types on the laptop

25 T (---) °me see°

26 (8.2) T gaze to screen, reads

27 T <when you are alone (1.7) wh- especially  
 28 when it whe- what when this is what you enjoy.»  
 29 (0.5) T gaze to screen ->>

30 Ira ohh.

31 (0.7) T leans back in her chair

32-> T #so it is kind of #like being alone by †choice.  
 iraga #gaze to T-----#gaze down to notebook->>



Fig. 3A



Fig. 3B

ically (Nation, 2001), pronouncing words in the L2 can be challenging. One way of addressing this challenge is to actively learn to pronounce difficult vocabulary items when encountering them (Uchihara et al., 2022). In this section, we investigate the participants' practices of teaching and learning to pronounce those lexical items that tutees have problems with and that the tutor treats as trouble sources during the meaning-focused explanation sequences and initiates repair on them. The analysis describes how the tutor scaffolds the tutees to notice the manner and place of articulation of the target sounds through articulated repetition alongside embodied and material resources. The tutor's repair initiations provide a pronunciation prompt, and thus call for 'doing pronunciation', that the tutees model (also Brouwer, 2004; Kääntä, 2017). The tutees' pronunciation attempts are subsequently assessed by the tutor.

The tutor initiates the pronunciation teaching sequences, either in situations where tutees display misunderstanding of the phonetic realization of vocabulary items (Ext. 4) or where they mispronounce the focal word under discussion (Ext. 5). In both extracts, the tutor orients to the tutee's pronunciation as somehow erroneous and pursues for a version she finds acceptable. To highlight the actual pronunciation of the focal words, and hence to show where the pronunciation problems lie, we use phonetic transcriptions of them, following International Phonetic Alphabet conventions. Extract 4, as part of the extended explanation sequence begun in Extract 2, demonstrates an instance where T employs different practices, such as accentuating the target sound and marking the place of the articulation, to elicit the pronunciation of 'immersion'. As we saw in Extract 2, T established the difference in meaning between 'emerge' and 'immersion' and here the attention is on the difference in their pronunciation. The extract begins with T focusing on the contrast in their pronunciation by demanding a modeling response from Uma (l. 43).

Repeating after T, Uma models the pronunciation of 'emerge' (l. 44). Instead of explicitly assessing it, T utters the word 'immersion' with raised eyebrows and a marked emphasis on the last syllable (l. 45), thereby highlighting in an implicit way the difference in its pronunciation to that of 'emerge.' Uma's modeling of the word (l. 46) occasions an extensive negotiation on its pronunciation, indicating that T finds Uma's version erroneous. T first engages in repair work by emphasizing and slightly exaggerating the manner of articulation of the problematic sound Uma could not produce (/ʒ/ instead of /dʒ/) (l. 47). T also points at her own mouth (see Fig. 4A) and thereby directs Uma's attention to the place of articulation (also Nguyen, 2016), thus using an 'instructional catchment' (Smotrova, 2017). Despite this, Uma still pronounces the word with the /dʒ/ sound (l. 48). Ira also joins in by mouthing 'immersion' to herself (l. 49), practicing its pronunciation without getting involved in the teaching sequence.

T's next repair initiation (l. 50) functions both as a rejection of Uma's candidate pronunciation and as another attempt to proffer the pronunciation to Uma. Uma again models T's pronunciation with the same version *sotto voce* (l. 51). Simultaneously, she brings both hands in front of her, elbows on the table, and makes a gesture that appears to signal 'no more, I cannot do it' (see Fig. 4B). This meaning is further built through the smilingly performed try that also manifests Uma's own awareness of her inability to model T's pronunciation. The repair sequence culminates here, as T produces an unmitigated 'no', reinforced with a head shake and orientation to the notebook (l. 52). In overlap with T's negative assessment, Ira again mouths 'immersion' (l. 53) and Uma brings her hands on her mouth and laughs (l. 54). Uma's laughter here might be functioning as an invitation to co-laugh (Glenn, 2003), and thus treats the pronunciation trouble as nonserious. Note also

that Uma's failure to detect the difference between her way of saying the word (/ɪ'mɜrdʒənd/) and the way T has presented it (/ɪ'mɜʒən/) leads Ira to imitate the word quietly along, not voicing it out loud, and also Uma to mark her difficulty (closing her mouth with hands and laughing) regarding the delivery of the target sound.

Since Uma after several attempts has failed to pronounce 'immersion' according to T's model, as a final resort, T explains the difference between the two sounds, the latter of which causes difficulty for Uma (l. 55 & 59–60). While she explains, T points at the two words she has written in the notebook. When she models the 'g' sound of 'emerge' (l. 55), T moves her right hand quickly away from her mouth, as if to spit the sound out (l. 57, see Fig. 4C). The explanation is followed by another pronunciation model of 'emerge' that Uma repeats without difficulty (l. 58). Then, pointing at 'immersion' in the notebook, T explains how to pronounce the problematic sound in the word (l. 59). The explanation is accompanied by a depictive gesture, i.e. T sweeps smoothly with her left hand above the notebook that helps visualize the smoothness and softness of /ʒ/ compared to the abruptness of /dʒ/. As with 'emerge', Uma models the pronunciation of 'immersion' once more (l. 61) but it is barely audible, so it is difficult to detect whether she manages to model the target sound or not. Simultaneously, T returns the notebook back to Uma and shakes her head, thus marking the negative assessment of Uma's pronunciation and the closure of the pronunciation teaching sequence.

Interestingly, Ira expands the sequence by gazing at Uma and producing different syllables, which include the problematic /ʒ/ sound (l. 62). T aligns with Ira's help through a glance and a pointing gesture toward her, thus acknowledging Ira's production of the target sound. T also repeats the trouble source, prosodically marking the final syllable, and shifts her gaze back to the computer in orientation of final closure. However, in lines 65 and 67, Ira continues producing similar syllables, eventually bursting into laughter. Uma aligns also by laughing and continues in this humorous frame by saying '\$easy busy\$', which Ira repeats in laughing (l. 70). Uma's turn seems to be an ironic comment of the 'doing pronunciation' sequence, while the aligned laughter by both also downplays the seriousness of Uma not being able to learn to pronounce 'immersion' as modeled by T. Through these actions, the tutees display their orientation to managing the tensions between the goals of the tutoring and their performances (Pomerantz, 2019).

Our final example, as an embedded part of Extract 1, demonstrates how T treats Uma's pronunciation of 'bizarre' as erroneous and initiates a repair sequence during which she repeats the problematic sound/syllable and explains its manner of articulation, also by drawing Uma's attention to the difference between the minimal pair 'bizarre-bazaar'.

Recall that in Extract 1, in line 08, T has explained the meaning of 'bizarre' through the synonym 'strange.' In the current case, in line 09, Uma repeats the beginning of T's prior turn with a slight questioning tone, which seeks confirmation from T both of the meaning of the word and whether Uma has heard the right word. That is, her repetition functions here as a candidate hearing (Lilja, 2014). However, instead of producing a (dis)confirmation, T orients to Uma's pronunciation of 'bizarre' by carefully repeating it (l. 10). T's repair initiation is treated by Uma as 'doing pronunciation' as she models it (l. 11). In response, T initiates repair again by producing the first consonant sound (/b(13:italic)Y/(13:italic)/) multiple times, thus marking it as the repairable (l. 12). More pertinently, by raising her chin up and emphasizing the sound by pressing her lips together, T orients to the articulatory aspects of the target sound, and draws Uma's attention to its place and manner of articulation (Fig. 5). After this, T models the whole word again, overlapping with Ira's articulation of the word *sotto voce* (l.

## Extract 4. Immersion.

- 43 T say /I'm3:dʒ/  
tga -gaze to Uma-> (1. 52)  
umaga --down to notebook->
- 44 Uma /I'm3:dʒ/  
45 T an' "</I'm3:ʒən/>"  
tf "raises eyebrows"
- 46 Uma /I'm3:dʒən/ → 
- 47-> T /I'm3:ʒən/ +\$  
tge +points at her mouth+  
umaga \$gaze to T\$
- 48 Uma [\$/I'm3:dʒənt/  
49 Ira [((mouths 'immersion' to herself))  
umaga \$gaze down->
- 50 T \$/I'm3:rʒən/  
umaga \$gaze to T->
- 51 Uma \*£°/I'm3:rʒənd/°£\* → 
- 52 T [^+°no°+  
53 Ira [((mouths 'immersion' to herself))  
54 Uma [\*>ah hah hah hah<\*  
umage \*hands together in front of mouth\*  
tge +head shake+  
tga ^gaze to notebook->
- 55-> T \$+in this one you ^pronounce\$ the +/dʒ/+  
tga ^gaze to Uma->  
tge +points at word in notebook-----rh gesture away from mouth+  
umaga \$gaze to notebook-----\$gaze to T->
- 56 Uma ==yeah\*=  
umage \*nods\*
- 57-> T ==+/I'm3:dʒ/+ → 
- 58 Uma \*/I'm3:dʒ/\*  
59 T [+/I'm3:ʒən/^ it's like++more smooth.+ Fig. 4C  
tge +points at word in notebook++hand gesture+  
umage \*nods-----\*
- 60-> +you don't pronounce+ that +/dʒ/+  
tge +hand gesture in front of face^  
tge +head nod down+
- 61 Uma ^\$+(immersion)+\*  
umage \*takes notebook\*  
umaga \$gaze shift to Ira-->  
tga ^gaze to notebook->  
tge +hands notebook to Uma+
- 62 Ira #ʒ::: £[ʒa:::] [ba::£  
63-> T ^+[z: ]+^  
64-> T ^[</I'm3:ʒən/>  
iraga #gaze to Uma->  
tga ^glances at Ira^gaze to computer-->  
tge +points to Ira+
- 65 Ira ʒi::  
66 (1.3) Ira gaze to table  
Uma gaze to notebook, then to Ira
- 67 Ira #ʒu:: ʒu:: ʒu:: \$eh [heh heh heh] heh#  
68 Uma [hih hih hih]  
iraga #gaze to Uma-----#  
umaga \$gaze to T's computer->>
- 69 Uma #£easy busy£#  
iraga #gaze down her left#
- 70 Ira #£ea(h)sy bu(h)sy£#  
iraga #gaze down her right#

Extract 5. Not like bazaar (embedded sequence from Ext. 1).

- 09 Uma /pɪzɑr/ means<sub>ɛ</sub>=  
 umaga -gaze to T->  
 tga -gaze to Uma->
- 10 T =/bɪ'zɑ:r/  
 11 Uma \$/pɪ'zɑ:r/\$  
 uma \$glances to her right->
- 12-> T [+bə]bə \$bə-bə [/bɪ'zɑ:r/  
 13 Ira [°/bɪzɑ:r/°]  
 tge +chin up-----+  
 iraga #gaze to Uma->  
 umaga \$gaze to T->
- 14 Uma /bɪ['zɑ:r/ ] oh okay.  
 15 Ira [°/bɪzɑ:r/°]#  
 -->#
- 16-> T ^not like /bə'zɑ:r/^  
 tga ^gaze in front of her^
- 17-> ^(.) [+like market+  
 18 Uma [\$\*/bə'zɑ:r/ means\*  
 umaga \$gaze down in front of her->  
 umage \*hand up-down gesture\*  
 tga ^gaze to Uma->  
 tge +hand gesture+
- 19 (0.3)
- 20 T [like +those ] markets+  
 21 Uma [oka' \$bizarre (oh)]  
 tge +hand gesture-----+  
 umaga \$gaze to T-> (l. 36)
- 22 (0.3) Uma gaze to T-->  
 T gaze down in front of her->
- 23 T er: places like (0.7) like in  
 24 (0.4) in ^turkey stuff like  
 tga ^gaze to Uma->
- 25 they- ^+the famous /bə'zɑ:r/+  
 tga ^gaze up in the air->  
 tge +bh hand gesture-----+
- 26 (0.5)
- 27 Uma ohh ba[ɪzɑ:r  
 28 T [^(it's) /bɪ'zɑ:r/^  
 tga ^gaze to Uma-----^
- 29 Uma ^it's /bɪzɑ:r/.=  
 tga ^gaze up in the air->
- 30 T =yeah bazaar it's spelt like  
 31 +er: like+ the market [place]  
 32 Uma [b a ] zed  
 tga +gaze up-+
- 33-> T %b a z a a r  
 irage %Ira writes->
- 34 (0.7) T still gaze up in the air



Fig. 5

## Extract 5. Continued

35->	T	^/bɪ'ʒɑ:r/ with ɪi^ tga ^gaze to Uma-----^
36	(0.6)	T gaze in front of her Uma orients to notebook writing->>
37->	T	b i z a r r e:ɔ
38	(1.5)	T gaze in front of her Ira stops writing, glances at T & laughs
39	Ira	#b i= iraga #gaze at notebook, continues writing-->>
40	T	="z a r" r e:ɔ >I think that's how you spell it< tf "frowns"
41->		is- means like very ^strange (0.3) +abnormal+^ tga ^gaze to Uma-----^ tge +points at Uma+
42	(3.0)	Ira & Uma write notes

13). While pronouncing it, Ira gazes at Uma, whereby she also joins in practicing the pronunciation and modeling to Uma the pronunciation alongside T. In line 14, Uma produces the target-like pronunciation, acknowledging it with 'oh okay'. In overlap, Ira models the pronunciation, again in soft voice and by gazing at Uma.

After the pronunciation of 'bizarre' has been settled, T introduces a minimal pair (l. 16) as a way of exposing the tutees to hearing sound contrasts. T uses an animated explanation (Waring et al., 2013) combining talk and a depictive gesture of spreading hands out to both sides to secure a shared meaning (like market) for the newly-introduced word. Due to Uma's overlapping talk (l. 18), T begins to repeat the meaning of 'bazaar' (l. 20), which is again overlapped with Uma's processing of 'bizarre' (l. 21). Once in the clear, T gives an example in the form of an association to 'the famous bazaar' in Turkey (l. 23–25). The reference is accompanied by a depictive gesture that helps visualize its grandness (spreading hands wide apart from one another), thus serving to disambiguate the meaning (Belhiah, 2013). In response, Uma produces a change of state token and repeats the word 'bazaar' with downward intonation (l. 27) that displays her understanding of the contrast.

The contrast in pronunciation dealt with, T revisits the target word ('bizarre') with salient prosodic delivery (l. 28). With the exaggerated prosody, T might treat Uma as accountable for poor reciprocity. When Uma repeats the teacher's /bɪzɑ:r/ with a pronunciation that matches that of the tutors (l. 29), T responds with an acknowledgement token and then clarifies the spelling of the words by first focusing on 'bazaar' (l. 30–31). While T is still recollecting its spelling, Uma offers the first three letters of the word, thus displaying some familiarity with the orthographic form of the word. At this point, T also spells it (l. 33). Having established the spelling for 'bazaar', T sets up the contrast between the words again, marking 'bi↑zarre' prosodically and mentioning the distinguishing sound 'ɪi' in the minimal pair (l. 35). T subsequently spells 'bizarre' (l. 37) and Uma starts taking notes. That both tutees orient to the word as a learnable becomes visible when Ira also produces an embodied other-initiation of repair (l. 38; Seo & Koshik, 2010) and spells the first two letters of the word (l. 39), completed by T spelling the rest of the word. This way Ira secures her understanding of how 'bizarre' is spelled and is able to write it. The teaching sequence is closed when T repeats the meaning of 'bizarre' by first saying the original synonym (strange) and introducing a new one (abnormal).

Overall, this section has shown how the participants subject the focal words to extended negotiation and pronunciation practice, whereby the tutees demonstrate their orientation to the tutor's pronunciation models as learnables. In teaching the pronunciation of the focal words, the tutor employs a range of verbal, embodied and material resources. A recurrently used 'instructional catchment' (Smotrova, 2017) involves pointing gestures that direct the tutees' attention to the place of articulation of the difficult sounds (Ext. 4 & 5; also Ngyuen, 2016). Alongside with or in lieu of pointing, the tutor also emphasizes the problematic sounds/syllables by repeating them (Ext. 5) or accentuating them (Ext. 4), thus addressing the manner of pronunciation, e.g., in the case of voiced vs. voiceless sounds (Ext. 5 /b/ vs. /p/). When other resources fail, the tutor provides explicit explanations of key differences between the problematic sounds, often with environmentally coupled gestures that depict the sound quality (Ext. 4 /ɜ/ vs. /dʒ/). Interestingly, the tutor also utilizes meaning-related viewpoints, such as making a contrast through minimal pairs and their spelling (Ext. 5), to teach the pronunciation of the target items.

## 5. Conclusion and implications

This study set out to explore how a tutor and tutees build word knowledge through integrated vocabulary explanations, i.e., extended explanation sequences during which participants not only jointly negotiate the meaning and use of lexical items but also their pronunciation and spelling in an integrated manner, in a hitherto under-researched ESL tutorial setting. Using multimodal conversation analysis, the study has approached the teaching and learning of word knowledge as a context-sensitive activity that helps zoom in on those aspects of word knowledge the participants orient to as relevant for their L2 learning. By drawing on the findings from three L2 research areas, i.e., tutorials, vocabulary explanations and pronunciation instruction, the study has offered new insights on integrated vocabulary explanations. The research questions the study has answered are how the participants collaboratively negotiate the form, meaning and use of vocabulary items, and what kinds of learning opportunities integrated vocabulary explanations offer for the L2 tutees. In what follows, we will first summarize the ways in which the participants accomplish the negotiation of word knowledge, and we will conclude with the implications of the study in relation to the kinds of L2 learning op-

portunities integrated vocabulary explanations afford for both ESL tutorials and L2 classroom interaction.

The findings show that integrated vocabulary explanations arise from instances when the tutees produce multimodally designed candidate understandings of words they are unfamiliar with, thereby orienting to the tutor's epistemic right to (dis)confirm them (Antaki, 2012). In response to the candidate understandings, the tutor spends a great deal of instructional time, first of all, helping the tutees identify morphological and semantic relationships of the focal words, and thus promotes their awareness of how words extend their meanings and how contextual variations emerge in word usage. In this regard, our findings of the tutor's practices both echo and build on previous research on L2 vocabulary explanations (e.g., Mortensen, 2011; Sert, 2017; Stoewer & Musk, 2019; Tai & Brandt, 2018; Waring et al., 2013). For example, the tutor provides synonyms and information about connotations, identifies word classes, and contextualizes the use of the words by giving example sentences that tap onto the tutees' word choices in their candidate understandings (Exts. 2 & 3), thus showing sensitivity to their partial understanding and developing her explanations based on them. The tutor also draws on gestures in reinforcing (Exts. 1 & 3) or disambiguating (Ext. 2) the meaning of the focal words (also Belhiah, 2013; Koç & Ergül, 2023; Waring et al., 2013). The tutor's practices of spelling and writing the words on paper to provide also a visual input to the tutees contribute towards constructing a 'last resort' when intersubjectivity is not secured (Ex. 2). When the tutees call into question the tutor's explanations, they influence the way she resorts to external evidence, an online source, to resolve the trouble source (Ext. 3). The tutees thus also display epistemic access to word meanings through disaffiliative candidate understandings (Antaki, 2012), whereby the negotiation of the acceptability of the tutor's response becomes plainly visible in interaction.

While clarifying and elaborating on the meaning and use of the focal words, the tutor also draws the tutees' attention to their form, when she orients to the tutees' pronunciation of them as somehow erroneous. By bringing the tutees' attention to issues of form in the midst of meaning-focused interaction, the tutor signals that pronunciation and spelling are an essential part of doing vocabulary work (Jakonen, 2014; Morton & Jakonen, 2016). Our study thus sheds light on the multiplicity of locally-occurring instances of vocabulary learning, in which both the tutor and the tutees orient to 'doing pronunciation teaching' through repair and correction practices such as providing pronunciation models and repeating the words along. Similar practices of modeling and repeating have been reported by Brouwer (2004) and Käätä (2017) in other L2 contexts. The tutor also highlights the place and manner of where and how specific sounds are produced (also Nguyen, 2016) by using articulatory gestures and writing the words down. The use of the embodied and material resources as instructional tools act as a visual highlighter of the prosodic and phonetic structure of the language in pronunciation instruction (Smotrova, 2017), serving as essential resources to secure an acceptable pronunciation from the tutees.

Through demonstration of how her mouth and lips are properly positioned with gestures, the tutor articulates the problematic sounds/syllables and also upgrades gestures as an additional attempt to correct the tutees' pronunciation that deviates from L1 users of US English (also Nguyen, 2016; Smotrova, 2017). This deviation is something the tutor, as a native speaker of North American English, orients to and assesses. Thus, it is her evaluation of the articulatory proximity to her native speaker model that is visible in the data and that induces the pronunciation instruction sequences in the midst of meaning-focused work. Due to her role as the tutorial service provider, she seems to adhere to the idea that certain standards are essential for helping learners appropri-

ately communicate in the language they learn (see Heritage et al., 2015 for an extensive research on assessment practices for English language learners). In this sense, the ability to use the standard accent may be useful especially when it is widely accepted; however, it is important to remember that standard dialects are relative in nature just as native speakers are. Yet, correcting pronunciation is crucial when intersubjectivity is potentially at stake, for example, due to the tutee's misunderstanding (Ext. 2 'emergent' vs. 'immersion'). In such instances, the tutor also treats the issue as a recognition problem, thereby establishing familiarity regarding the meaning of the learnable. The tutees actively engage in seeking clarifications from the tutor and modeling after her pronunciation prompts. They also write notes at crucial moments, thus marking the information as relevant for the development of their L2 competence (Svinhufvud, 2015). All these actions demonstrate that 'correct' pronunciation (Gardner, 2005) and spelling are treated as relevant practices of learning L2 vocabulary.

All in all, we argue that understanding the nature of integrated vocabulary explanations has practical implications. Given the dual nature of tutoring as informal talk between peers and pedagogical interaction between a tutor and (a) tutee(s) as well as the tutees' key role in setting the goals for the tutoring sessions, we can observe the unique side of the current research setting: the tutorial context lends itself well to individualization and immediacy, thereby allowing opportunities for tutors to be more responsive to their tutees' needs and concerns than teachers in classroom settings (see also Duran & Jakonen, 2022). Language learners in ESL tutorials may have more opportunities to practice ways of doing argumentation in the target language than in language classrooms and, thus have more ways of accessing the conversational floor, which also allows the tutor to have the flexibility to carry out the pedagogical goals in the best direction to suit their needs, and thus provides individual attention to different tutees, whether it be a student in need of more vocabulary exercises or a student requesting for more speaking practices in L2. Moreover, identification of the interactional practices that characterize integrated vocabulary explanations is certainly beneficial for current and future teachers who would like to incorporate those practices into their classroom practices. To that end, our findings provide implications for tutor education and tutoring center administration. The fine-grained analysis of 'tutorial business' as it occurs in interaction allows us to see those real-time experiences, which may inform us to make the best use of tutees' and tutors' time and the centers' resources.

Insights on teaching and learning word knowledge from ESL tutorials can also help mainstream classroom teachers effectively engage with multilingual and multicultural student populations. As our study reveals, it is not sufficient to explain only the semantic meaning of words: language learners benefit from a more integrated approach to explaining the unfamiliar vocabulary they encounter, for example in texts or other people's speech, as the knowledge of a word also includes its orthographic, phonological, syntactic, collocational and pragmatic characteristics. This is why we argue that teachers and teacher trainees should allocate time to address the different aspects of word knowledge during classroom interaction. A practical method would be to remember to answer (some of) the following questions whenever an instance of vocabulary explanation emerges: what is the meaning of the word?, what are its synonyms or antonyms?, how is it spelled and pronounced?, how can it be used in a sentence and what is its word class?, and what kinds of collocations, connotations and associations does it have? Not only would this be relevant for teachers to bear in mind, but students could also be taught to think of these questions when they encounter new words, both during whole-class and peer group interaction, and even during one-on-one tutoring. This would develop their agency and provide more learning op-

portunities for them. To conclude, language teachers/tutors should encourage learners to be aware of their perception and production abilities by offering learners multiple opportunities to practice the language in both contextualized (vocabulary explanations in our data) and decontextualized (pronunciation instruction in our data) manners.

#### Declaration of Competing Interest

None.

#### Data availability

The authors do not have permission to share data.

#### Acknowledgements

We would like to thank our research participants. We are indebted to the editor and anonymous reviewers for the valuable feedback they offered on this manuscript. Any remaining errors are ours.

#### Appendix. Transcription conventions

#### Appendix. Transcription conventions

(1.8)	Numbers enclosed in parentheses indicate a pause. The number represents the number of seconds of duration of the pause, to one decimal place. A pause of less than 0.2 s is marked by (.)
[ ]	Brackets around portions of utterances show that those portions overlap with a portion of another speaker's utterance.
=	An equal sign is used to show that there is no time lapse between the portions connected by the equal signs. This is used where a second speaker begins their utterance just at the moment when the first speaker finishes.
::	A colon after a vowel or a word is used to show that the sound is extended. The number of colons shows the length of the extension.
(hm, hh)	These are onomatopoeic representations of the audible exhalation of air
.hh	This indicates an audible inhalation of air, for example, as a gasp. The more h's, the longer the in-breath.
?	A question mark indicates that there is slightly rising intonation.
.	A period indicates that there is slightly falling intonation.
,	A comma indicates a continuation of tone.
-	A dash indicates an abrupt cut off, where the speaker stopped speaking suddenly.
↑↓	Up or down arrows are used to indicate that there is sharply rising or falling intonation. The arrow is placed just before the syllable in which the change in intonation occurs.
Under	Underlines indicate speaker emphasis on the underlined portion of the word.
CAPS	Capital letters indicate that the speaker spoke the capitalized portion of the utterance at a higher volume than the speaker's normal volume.
◦	This indicates an utterance that is much softer than the normal speech of the speaker. This symbol will appear at the beginning and at the end of the utterance in question.
><, <>	'Greater than' and 'less than' signs indicate that the talk they surround was noticeably faster, or slower than the surrounding talk.
(would)	When a word appears in parentheses, it indicates that the transcriber has guessed as to what was said, because it was indecipherable on the tape. If the transcriber was unable to guess what was said, nothing appears within the parentheses.
\$C'mon\$	Dollar signs are used to indicate a smiley or jokey voice.

## References

- Antaki, C. (2012). Affiliative and disaffiliative candidate understandings. *Discourse Studies*, 14, 531–547.
- Belhiah, H. (2009). Tutoring as an embodied activity: How speech, gaze and body orientation are coordinated to conduct ESL tutorial business. *Journal of Pragmatics*, 41, 829–841.
- Belhiah, H. (2013). Using the hands to choreograph instruction: On the functional role of gesture in definition talk. *The Modern Language Journal*, 97, 417–434.
- Brouwer, C. (2004). Doing pronunciation: A specific type of repair sequence. In R. Gardner, & J. Wagner (Eds.), *Second language conversations* (pp. 93–113). Continuum.
- Brown, A. (2014). *Pronunciation and phonetics: A practical guide for English language teachers*. Routledge.
- Clift, R. K. (2016). *Conversation analysis*. Cambridge University Press.
- Duran, D., & Jakonen, T. (2022). Mobilizing context: Vocabulary checks in ESL tutoring sessions. *System*, 107. <https://doi.org/10.1016/j.system.2022.102816>.
- Elgort, I., & Nation, I. S. P. (2010). Vocabulary learning in a second language: Familiar answers to new questions. In P. Seedhouse, S. Walsh, & C. Jenks (Eds.), *Conceptualising “learning” in applied linguistics* (pp. 89–104). Palgrave Macmillan.
- Eskildsen, S. W., & Majlesi, A. R. (2018). Learnables and Teachables in Second Language Talk: Advancing a Social Reconceptualization of Central SLA Tenets. *The Modern Language Journal*, 102, 3–10.
- Eskildsen, S. W., & Wagner, J. (2013). Recurring and shared gestures in the L2 classroom: Resources for teaching and learning. *European Journal of Applied Linguistics*, 1(1), 139–161.
- Gardner, H. (2005). A comparison of a mother and a therapist working on child speech. In K. Richards, & P. Seedhouse (Eds.), *Applying conversation analysis* (pp. 56–72). Palgrave Macmillan.
- Glenn, P. J. (2003). *Laughter in interaction*. Cambridge University Press.
- Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. In J. M. Atkinson, & J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (pp. 299–345). Cambridge University Press.
- Heritage, M., Walqui, A., & Linquanti, R. (2015). *English language learners and the new standards: Developing language, content knowledge, and analytical practices in the classroom*. Harvard Education Press.
- Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 13–31). John Benjamins.
- Jakonen, T. (2014). Knowing matters: How students address lack of knowledge in bilingual classroom interaction. *Jyväskylän tutkimus 235*. University of Jyväskylä.
- Kasper, G., & Kim, Y. (2015). Conversation-for-learning: Institutional talk beyond the classroom. In N. Markee (Ed.), *Handbook of classroom discourse and interaction* (pp. 390–408). Wiley-Blackwell.
- Kasper, G., & Wagner, J. (2011). A conversation-analytic approach to second language acquisition. In D. Atkinson (Ed.), *Alternative approaches to second language acquisition* (pp. 117–142). Routledge.
- Kim, S., & Cho, S. (2017). How a tutor uses gesture for scaffolding: A case study on L2 tutee's writing. *Discourse Processes*, 54(2), 105–123.
- Koç, T., & Ergül, H. (2023). The use of material resources in vocabulary explanations: A conversation analytic inquiry. *Text and Talk*. <https://doi.org/10.1515/text-2021-0118>.
- Koshik, I. (2002). Designedly incomplete utterances: A pedagogical practice for eliciting knowledge displays in error correction sequences. *ROSLI*, 35, 277–309.
- Kääntä, L. (2017). In search of proper pronunciation: Students' practices of soliciting help during read-aloud. In M. Kuronen, P. Lintunen, & T. Nieminen (Eds.), *Näkökulmia toisen kielen puheeseen - Insights into second language speech* (pp. 61–81). AFinLA-e. Soveltavan kielitieteen tutkimuksia. 2018 /n:o 10.
- Kääntä, L., & Kasper, G. (2018). Clarification requests as a method of pursuing understanding in CLIL physics lectures. *Classroom Discourse*, 9(3), 205–226.
- Kääntä, L., Kasper, G., & Piirainen-Marsh, A. (2018). Explaining Hooke's Law: Definitional practices in the CLIL physics classroom. *Applied Linguistics*, 39(5), 694–717.
- Leyland, C. (2018). Resistance as a resource for achieving consensus: Adjusting advice following competency-based resistance in L2 writing tutorials at a British university. *Classroom Discourse*, 9(3), 267–287.
- Leyland, C. (2020). Academic writing tutorials for international students: Deferring to an expert and follow-up advice. *Language and Education*, 34(3), 212–230.
- Leyland, C. (2021). The interactional construction of the academic reader in writing tutorials for international students: An advice-giving resource. *Linguistics and Education*, 61, 1–13.
- Leyland, C., & Riley, J. (2021). Enhanced English conversations-for-learning: Constructing and using notes for deferred correction sequences. *Linguistics and Education*, 66, 32–43.
- Lilja, N. (2014). Partial repetitions as other-initiations of repair in second language talk: Re-establishing understanding and doing learning. *Journal of Pragmatics*, 71, 98–116.
- Long, J. R. (2013). *Language labs: Their impact on english language learners*. doctoral dissertation. Capella University.
- Majlesi, A. R. (2014). Learnables in action: The embodied achievement of opportunities for teaching and learning in Swedish as a second language classrooms. PhD dissertation. Linköping University, Sweden.
- Miller, G. A. (1999). On knowing a word. *Annual Review of Psychology*, 50, 1–19.
- Mondada, L. (2014). Conventions for multimodal transcription. Available at [https://franz.unibas.ch/fileadmin/franz/user\\_upload/redaktion/Mondada\\_conv\\_multimodality.pdf](https://franz.unibas.ch/fileadmin/franz/user_upload/redaktion/Mondada_conv_multimodality.pdf)
- Mori, J. (2004). Pursuit of understanding: Conversation analytic account of a small-group activity in a Japanese language classroom. In R. Gardner, & J. Wagner (Eds.), *Second language conversations* (pp. 157–177). Continuum.
- Mori, J., & Hayashi, M. (2006). The achievement of intersubjectivity through embodied completion: A study of interactions between first and second language speakers. *Applied Linguistics*, 27, 195–219.
- Mortensen, K. (2011). Doing word explanation in interaction. In G. Pallotti, & J. Wagner (Eds.), *L2 learning as social practice: Conversation analytic perspectives* (pp. 135–162). Manoa: National Foreign Language Resource Center.
- Morton, T. (2015). Vocabulary explanations in CLIL classrooms: A conversation analysis perspective. *The Language Learning Journal*, 43(3), 256–270.
- Morton, T., & Jakonen, T. (2016). Integration of language and content through languaging in CLIL; Classroom interaction: A conversation analysis perspective. In T. Nikula, E. Dafouz, P. Moore, & U. Smit (Eds.), *Conceptualising integration in clil and multilingual education* (pp. 171–188). Multilingual Matters.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.
- Nguyen, M. (2016). A micro-analysis of embodiment and speech in the pronunciation instruction of one ESL teacher. *Issues in Applied Linguistics*, 20(1), 111–134.
- Park, J. (2007). Co-construction of nonnative speaker identity in cross-cultural interaction. *Applied Linguistics*, 28(3), 339–360.
- Park, I. (2019). The use of embodied self-directed talk in teaching writing. In J. K. Hall, & S. D. Looney (Eds.), *The embodied work of teaching* (pp. 160–180). Multilingual Matters.
- Pomerantz, A. (2019). Negotiating the terms of engagement: Humor as a resource for managing interactional trouble in after-school tutoring encounters. *Linguistics and Education*, 49, 22–30.
- Ro, E. (2021). Using PowerPoint slides as a resource for coordinating understanding during presentation consultations at an L2 speaking center. *Applied Linguistics Review*. <https://doi.org/10.1515/applirev-2020-0052>.
- Seo, M. (2021). Multimodally enhanced opportunities for language learning: Gestures used in word search sequences in ESL tutoring. *Journal of Language Teaching and Research*, 12(1), 44–56.
- Seo, M., & Koshik, I. (2010). A conversation analytic study of gestures that engender repair in ESL conversational tutoring. *Journal of Pragmatics*, 42, 2219–2239.
- Sert, O. (2017). Creating opportunities for L2 learning in a prediction activity. *System*, 70, 14–25.
- Smotrova, T. (2017). Making pronunciation visible: Gesture in teaching pronunciation. *TESOL Quarterly*, 51(1), 59–89.
- Smotrova, T., & Lantolf, J. P. (2013). The function of gesture in lexically focused L2 instructional conversations. *The Modern Language Journal*, 97, 397–416.
- Svinhufvud, K. (2015). Participation in the master's thesis seminar: Exploring the lack of discussion. *Learning, Culture and Social Interaction*, 5, 66–83.
- Stoewer, K., & Musk, N. (2019). Impromptu vocabulary work in English mother tongue instruction. *Classroom Discourse*, 10(2), 123–150.
- Szczepek Reed, B. (2012). A conversation analytic perspective on teaching English pronunciation: The case of speech rhythm. *International Journal of Applied Linguistics*, 22(1), 67–87.
- Tai, K. W. H., & Brandt, A. (2018). Creating an imaginary context: Teacher's use of embodied enactments in addressing learner initiatives in a beginner-level adult ESOL classroom. *Classroom Discourse*, 9(3), 244–266.
- Tai, K. W. H., & Khabbazzashi, N. (2019). Vocabulary explanations in beginning-level adult ESOL classroom interactions: A conversation analysis perspective. *Linguistics and Education*, 52, 61–77.
- Vick, N., Robles-Piña, R. A., Martirosyan, N. M., & Kite, V. (2015). The effectiveness of tutoring on developmental English grades. *The Community College Enterprise*, 21, 11–26.
- Uchiyama, T., Eguchi, M., & Clenton, J. (2022). The contribution of guessing from context and dictionary use to receptive and productive vocabulary knowledge: A structural equation modeling approach. *Language Teaching Research*. <https://doi.org/10.1177/13621688221122138>.
- Waring, H. Z., Creider, S. C., & Box, C. D. (2013). Explaining vocabulary in the second language classroom: A conversation analytic account. *Learning, Culture and Social Interaction*, 2, 249–264.