

The Mediation and Organisation of Gestures in Vocabulary Instructions: A Microgenetic Analysis of Interactions in a Beginning-level Adult ESOL Classroom

Journal:	<i>Language & Education</i>
Manuscript ID	LE-2710.R1
Manuscript Type:	Paper
Keywords:	Microgenetic analysis, Second language development, Conversation analysis, ESOL classroom interaction, Vocabulary explanations
Abstract:	<p>There is limited research on second language (L2) vocabulary teaching and learning which provides fine-grained descriptions of how vocabulary explanations (VE) are interactionally managed in beginning-level L2 classrooms where learners have a limited L2 repertoire, and how the VEs could contribute to the learners' conceptual understanding of the meaning(s) of the target vocabulary items (VIs). To address these research gaps, we used a corpus of classroom video-data from a beginning-level adult ESOL classroom in the United States and applied Conversation Analysis to examine how the class teacher employs various gestural and linguistic resources to construct L2 VEs. We also conducted a 4-month microgenetic analysis to document qualitative changes in learners' understanding of the meaning of specific L2 VIs which were previously explained by the teacher. Findings revealed that the learners' use of gestures allows for an externalization of thinking processes providing visible output for inspection by the teacher and peers. These findings can inform educators' understanding about L2 vocabulary development as a gradual process of controlling the right gestural and linguistic resources for appropriate communicative purposes.</p>

SCHOLARONE™
Manuscripts

The Mediation and Organisation of Gestures in Vocabulary Instructions: A Microgenetic Analysis of Interactions in a Beginning-level Adult ESOL Classroom

There is limited research on second language (L2) vocabulary teaching and learning which provides fine-grained descriptions of how vocabulary explanations (VE) are interactionally managed in beginning-level L2 classrooms where learners have a limited L2 repertoire, and how the VEs could contribute to the learners' conceptual understanding of the meaning(s) of the target vocabulary items (VIs). To address these research gaps, we used a corpus of classroom video-data from a beginning-level adult ESOL classroom in the United States and applied Conversation Analysis to examine how the class teacher employs various gestural and linguistic resources to construct L2 VEs. We also conducted a 4-month microgenetic analysis to document qualitative changes in learners' understanding of the meaning of specific L2 VIs which were previously explained by the teacher. Findings revealed that the learners' use of gestures allows for an externalization of thinking processes providing visible output for inspection by the teacher and peers. These findings can inform educators' understanding about L2 vocabulary development as a gradual process of controlling the right gestural and linguistic resources for appropriate communicative purposes.

Keywords: microgenetic analysis, second language development, conversation analysis, ESOL classroom interaction, vocabulary explanations

Introduction

Previous research on second language (L2) vocabulary teaching and learning (e.g. Schmitt, 2000) has illustrated that learning a vocabulary item (VI) involves knowing its form, meaning and use. The majority of these studies focused on *what* needed to be achieved in vocabulary instruction including strategies for guessing meaning (Nation, 1990) and using first language (L1) in explaining vocabulary (Tian and Macaro, 2012).

However, there is a lack of research which illustrates *how* vocabulary explanation (VE) is accomplished in real-life L2 classroom interaction (Waring et al., 2013). Furthermore, most studies exploring VEs in L2 classrooms predominantly analysed the verbal part of such VEs without taking into account their non-verbal accompaniment (e.g. Morton, 2015).

An alternative analytic approach of microgenetic analysis has been taken by second language acquisition (SLA) researchers working within the framework of Sociocultural Theory (SCT). This analytical framework aims at capturing the moment-to-moment qualitative changes of learners' linguistic improvements over a shorter period of time. For example, van Compernelle (2010) and Matsumoto and Dobs (2017) employed Conversation Analysis (CA) in their micro-analysis of classroom interactions to trace learners' L2 grammatical development while Smotrova and Lantolf (2013) focused on tracking learners' incipient understandings of vocabulary. However, the time frame in these studies was short, typically involving only a lesson. Furthermore, the majority of the studies investigating the nature of VEs in L2 classrooms (e.g. Waring et al., 2013) failed to illustrate how learners demonstrate their understandings of the meanings of specific VIs over time since they conducted one-off analyses of classroom talk rather than performing continuous observations.

To address this research gap, this study utilised CA to first examine how a teacher employs gestures to construct L2 VEs in beginning-level adult English to Speakers of Other Languages (ESOL) classrooms, followed by a microgenetic analysis with multiple classroom observations over an expanded time-frame of four months to document the qualitative changes in learners' understandings of the meanings of the L2 VIs. To the best of our knowledge, no research to date has studied how beginning-level

ESOL learners display understandings of the meanings of VIs which constitutes our rationale for targeting this learner group.

In the following sections, we first discuss some of the key concepts used in this study before presenting the details of our research.

Learning and development in sociocultural theory

This study adopted SCT (Vygotsky, 1962) as its theoretical framework and conceptualises learning as a social and situated process which is embedded in interaction. Vygotsky suggested that higher forms of cognition are first mediated by the 'expert' or others who are more competent, but as learners participate in culturally organised activity, they gain control over their mental processes by appropriating the symbolic artefacts, including gestures and speech, of the experts (i.e. making something one's own) and controlling them voluntarily. This process of moving from external (inter-psychological) to internal (intra-psychological) regulation is viewed as internalisation. Evidence for L2 development can be learners' demonstration of the knowledge that they have taken from the interactions, and also their ability to use the L2 to articulate their own thoughts appropriately and spontaneously in a new interactional context as a consequence of the process of mediation. The notion of internalisation occurs through appropriation, which is not merely copying or replicating culturally constructed ways of thinking and acting, but one that involves 'transforming these ways of thinking and acting to make them one's own' (Lantolf and Thorne, 2006: 176).

Microgenetic Analysis

Microgenesis focuses on instances of learning as they occur 'in flight' during interpsychological activity over a short period of time (Lantolf and Thorne, 2006). As

Pekarek-Doehler and Lauzon (2015) argued, learning and development have typically been assessed through comparing learning outcomes between experimental and control groups, most often entailing research designs based on pre- and post-tests to observe the products of change. Nevertheless, little attention has been paid to explore how learning processes emerge in and through the minute details of naturally-occurring interactions. Results of such fine-grained analyses potentially offer a comprehensive picture of the process of learning and development in the moment-to-moment unfolding of interaction, and as embedded within jointly managed and locally accomplished courses of action.

It is important to acknowledge that microgenetic analysis is not without its limitations since different variables, including the nature of L2 tasks, learners' motivation for L2 learning and their proficiency can affect their L2 use. It is difficult for researchers to make inferences about learner-internal processes and developmental trajectories by simply observing external interactive processes. Moreover, one can argue that it is impossible for microgenetic analysis to fully examine the transitions that learners have undergone, as L2 learning can occur outside formal settings. Therefore, findings from microgenetic analysis are illustrative of learners' L2 development in specific sociocultural activities and not necessarily generalizable to other learning contexts.

Some earlier microgenetic studies (e.g. Storch, 2002) which adopted a process-product approach to examine the role of interaction in L2 development, tended to conduct discourse analysis which involved using coding schemes to conduct simple linguistic analysis of talk. This simplification of the nature of social interactions prevented researchers from understanding how the L2 developmental process is visible through social interactions. There is a growing number of microgenetic studies (e.g. Ohta, 2000; van Compernelle, 2010; Matsumoto and Dobs, 2017) that employed CA to

conduct micro-analysis of the classroom or peer interactions to trace the process of learning which leads up to the appropriation, by learners, of previously socially-elaborated features, including VIs, within a new interactional context.

CA as the methodological tool for conducting microgenetic analysis

CA is a sociological methodology which adopts a participant-relevant perspective to explore how participants achieve different social actions through interactions without pre-theorising the relevance and importance of language-in-use, which includes semiotic resources such as gesture (Sidnell, 2010).

The combination of CA and SCT has been a key issue in CA research. A number of scholars (e.g. Hauser, 2011) have pointed to problems inherent in combining CA with an external theory arguing that analyses may be led by a pre-determined theory rather than by observations in interactions (see Hauser, 2011 for further discussion). The complementarity of CA and SCT has been extensively discussed elsewhere (e.g. Mondada and Pekarek Doehler, 2004). In short, scholars have argued that CA focuses on the way participants systematically organise their social actions through interactions, while SCT focuses on the way human activity and mental development are mediated by socially-constructed cultural artefacts. Since CA does not provide a theory of learning, CA's focus on talk-for-action can offer the empirical basis for making claims about mediation, learning and development (van Compernelle, 2010). However, CA is not 'merely an analytic tool in the service of [SCT]' (Mondada and Pekarek Doehler, 2004: 504) and SCT is not merely 'an explanatory theory of learning in the service of CA' (van Compernelle, 2010: 69). Rather, CA's analytic approach allows researchers to explore SCT notions from the participants' perspective. They are therefore complementary enabling researchers to understand the intricacies of human activity and

psychological development. Adopting CA as a methodological tool can potentially allow researchers to trace the process of learning which leads up to the appropriation of a previously socially elaborated feature, including a vocabulary or grammatical item, within a new interactional context.

Several studies (e.g. Ohta, 2000; van Compernelle, 2010) have successfully demonstrated the usefulness of CA in conducting close empirical analyses of a given interaction to identify objects of learning and track their development over a short period of time. Nevertheless, the time frame of the microgenetic development was short: about seven minutes in a French language proficiency interview in van Compernelle (2010) and one Japanese L2 lesson in Ohta (2000). It remains unclear whether the learners could recall the target language over a longer time frame. To date, there is a lack of CA research that explores L2 learning and development over a more expanded time frame (e.g. a few weeks or months).

The role of gestures in L2 learning and development

Gestures are generally viewed as physical movements that co-occur with speech. Smotrova and Lantolf (2013) argued that both gesture and speech form a unit that is necessary to be analysed as a whole in order to understand the role of gestures in enhancing speaking and thinking. As a result, gestures have an important role as a mediational tool in L2 learning and development, particularly in relation to VEs (e.g. Smotrova and Lantolf, 2013), grammatical forms (e.g. Hudson, 2011) and pronunciations (e.g. Author and Poon, 2016) that are not familiar to learners.

McCafferty (2002; 2004) examined the communicative and cognitive functions of L2 learners' use of gestures when speaking in an L2 outside the classroom. McCafferty (2002) illustrated that gestures enhanced comprehension in social interactions between

native and non-native English speakers. McCafferty (2004) also found that several types of gestures including representation gestures (e.g. iconic and deictic gestures) and beats were employed by a learner to elicit the correct L2 VI from the researcher and illustrate the meaning of the VI that the learner could not verbalise.

A recent study by Matsumoto and Dobs (2017) demonstrated how teachers employed deictic and metaphoric gestures as resources to make abstract concepts related to English tense and aspect concrete and visible to learners. Most importantly, when examining the interaction which occurred a week after the lesson, it was noticeable that the learners appropriated the teachers' deictic gesture to demonstrate their understanding of present tense, which served as evidence of their microgenetic development. However, the authors only analysed one excerpt to explain the effects of gestures on learners' L2 development.

A study by Smotrova and Lantolf (2013) explored the mediational function of a teacher's use of gestures in clarifying Russian learners' understanding of English VEs. The analyses indicated that the teacher employed a variety of gestures (metaphoric and ironic) to visualise the contextual meaning of the words. It was shown that the learners appropriated and modified the teacher's gestures indicating their improved understanding of the new concepts. However, the time frame of the microgenetic development was limited to one lesson only.

Focusing on the case of Carlo – an adult Mexican Spanish-speaking learner of English – Eskildsen and Wagner (2015) conducted a longitudinal analysis of how Carlo's use of gestures developed over time in an ESOL classroom. The analysis of his use of pronouns ('under' and 'across') indicated that Carlo appropriated the teacher's gestures to demonstrate his understanding as illustrated in his spontaneous use of the

pronouns accompanied with the same or similar gestures as when he first encountered the pronouns in class.

The studies discussed above indicated that studying both learner and teacher gestures in the L2 classroom is a significant field of enquiry which requires much more attention from SLA researchers. More evidence from the analysis of speech and gestures together is needed to demonstrate the potential for CA in tracking learners' developmental processes.

The role of gestures in explaining vocabulary

One of the few studies that investigated the role of teacher gestures in L2 VE was Lazaraton (2004). Based on the fine-grained analysis of an ESOL classroom interaction in the US, Lazaraton concluded that gestures enhanced the teacher's input and made it more comprehensible for L2 learners. Nevertheless, this conclusion may be problematic, as the research did not explore the ways in which learners responded to the teacher's gestures. Moreover, the learners' non-verbal behaviours were not captured in the transcripts, which prevented a full understanding of the effects of the teacher's use of gestures on L2 learning.

A recent study by Author and Brandt (2018) examined teacher's use of embodied enactments in contingently explaining vocabulary to learners in an adult beginning-level ESOL classroom. Embodied enactment refers to the participants' employment of embodied resources, including use of gestures/body movement and/or verbal resources, to represent an aspect of hypothetical events. The authors argued that the notion of embodied enactment differs from the term 'embodied explanations' (Sert, 2017) where enactments are conceptualised as a supplement to the teacher's verbal responses instead of 'a distinct form of interactional and embodied conduct' (Author and Brandt, 2018: 6).

The teacher in Author and Brandt (2018) offered a verbal and physical representation of an imagined outside-of-the-classroom context, which helped learners understand how the specific VIs could be employed in that specific context.

To date, there is a lack of CA classroom interaction research which provides detailed descriptions of how VEs are interactionally managed, and how the VEs could contribute to learners' conceptual understandings of the meaning of target VIs. This study aimed to address these research gaps by investigating how a beginning-level ESOL teacher draws on various gestural and linguistic resources to construct L2 VEs and how learners display their conceptual understandings of the meanings of the VIs. While both gestural and linguistic resources play an important role in mediating L2 learning and development, the focus of this paper is predominantly on gestures; this is to reflect our study's context – a beginning-level classroom – where linguistic resources are limited and there is heavier reliance on gestures to achieve communicative goals.

Methodology

The video-data for this study were drawn from the Multimedia Adult English Learner Corpus (MAELC). The data were collected at Portland Community College. This corpus was compiled to allow researchers to conduct longitudinal studies of adult ESOL learners' SLA processes (Reder, 2005). The full corpus includes over 4000 hours of classroom interactions, recorded employing six video-cameras in the classrooms.

The segments of data selected for this study were collected from a beginning-level ESOL classroom from January to April 2002 covering the full teaching term at the end of which most students moved to an upper-level intermediate class. There were two lessons per week each lasting two hours. For this study, there were 30 lessons in total and we observed all of them. The ESOL teacher was an experienced teacher who had

studied German and Spanish at a US university. There were twenty-one adult learners of English enrolled in the class: two from Romania, nine from Latin American countries, one from Russia, three from Africa, six from China and one from Korea.

In line with the SCT perspective which perceives L2 learning and development as mediated in and through interaction (e.g. Lantolf and Thorne, 2006), this study combines the analytic methods of CA with SCT in order to trace the qualitative changes in learners' microgenetic development. Since this study aims to investigate learners' participation and teacher's pedagogical practices in real time, this study is classified as an applied CA study which 'tells us how to look, and what we must do in order to show how the features of institutions, like education, are produced in situ, in real time' (Heap, 1997: 223). Although scholars (e.g. Hauser, 2011) caution against using exogenous theories to inform CA analysis, the insights from SCT will only be brought into the discussion section when relevant. The analysis of the classroom data is completed within a CA framework. As Heap (1997) argued, applied CA studies can discuss the data with regard to exogenous theories, before and after conducting the CA analysis. Nevertheless, the CA analysis itself is to be proceeded as any other CA studies, with line-by-line analysis.

The first stage of analysis involved taking a stance of 'unmotivated looking' (Mori, 2004) as the guiding principle when reviewing the video-recordings from MAELC. This required us to watch multiple classroom-videos without any particular interest or research focus before conducting the exploratory analysis. In this process, we identified three noticeable features: firstly, the teacher devoted a significant amount of time in explaining unfamiliar VIs to her learners¹, not least because of the limited linguistic resources shared between the participants. The learners, on the other hand, displayed some ability to understand the teacher's English although the English they produced

was somehow limited. Secondly, and as a result of the learners' limited English abilities, the teacher's VEs drew heavily on embodied resources, and this was considered worthy of further analyses. Thirdly, some learners demonstrated their understandings of the VIs – that were previously explained by the teacher – in subsequent lessons. As Sidnell (2010) argues, once an interesting phenomenon has been identified, researchers can build up a collection of similar occurrences so that differences and similarities between each occurrence could be examined. This motivated us to review more video-recordings in order to locate further instances that entailed a transfer of knowledge from an earlier lesson to subsequent individual performances.

The second stage of analysis entailed the transcription of all excerpts that involved VE sequences as well as excerpts that illustrated the learners' transfer of knowledge. Data were transcribed using Jeffersonian CA conventions (Jefferson, 2004, see Appendix for the transcription conventions). Several symbols were used to represent the non-verbal actions (see Appendix). It is common for CA researchers to conduct multimodal analysis by including non-verbal conduct, and screen-shots of relevant actions captured in the video-recordings in order to focus on various semiotic resources employed by the participants including gestures and body movements (Mondada, 2018). After transcribing the data, we carried out line-by-line analyses of the sequences which entailed teacher's VEs and learners' transfer of knowledge by focusing on both the sequential organisation of talk and the different multimodal resources enacted by the classroom participants. We constructed the descriptions of the interactional organisations of VEs and learners' transfer of knowledge based on the full range of the extracts. The collections consisted of nineteen extracts on teacher's VEs and six extracts on learners' transfer of knowledge. This led to the identification of three features of the

VE sequences and three features of learners' transfer of knowledge. In this paper, we will only present two main features of the VE sequences and three features of learners' transfer of knowledge.

When undergoing the examination of the sequential organisation of VEs and interactional features of learners' transfer of knowledge, we began with 'finding patterns and explicating their logic' (ten Have, 2007: p.120). The CA's interactional mechanisms (i.e. turn-taking, adjacency pairs, preference organisation and repair) were applied to carry out the analysis.

For reporting purposes, we can only present illustrative extracts; this can raise concerns regarding the extent to which selected extracts are an adequate representation of all analysed data. It is therefore important to examine all extracts for similar and/or deviant instances. (ten have, 1990). In this study, the chosen extracts are typical VE sequences and characterise instances of learners' transfer of L2 knowledge that were found in the whole collection. Atypical VE sequences were not identified in the data. It is also worth noting that the goal of CA analysis is to find the 'devices', or 'the technology of conversation' in the speakers' situated interaction, instead of justifying the best possible representative extracts (ten Have, 1990). Therefore, as long as the selected extracts can address the research questions to reveal the relevant 'orderliness' with their representative nature, this can be said, to a large extent, that the representativeness is sufficient, or the research findings can be reliable.

Analysis

This section provides an analysis of how the ESOL teacher explains new VIs to her students followed by a demonstration of the learner's internalisation of the learnt VIs through the analysis of a teacher's speech and gesture.

Case One: Today, Tomorrow and Yesterday

Nature of VE: Embodied explanation

This extract was taken from the lesson on 7-1-2002 and was between teacher (T) and student 3 (S3). It illustrates the use of ‘embodied explanation’ which is defined as ‘explanations that include deployment of visual behaviours, like the use of hand gestures’ (Sert, 2017: 15). In this extract, T employs deictic and metaphor gestures to teach the pronouns for expressing time.

Extract 1

For Peer Review

01 T: okay↓
 +S3 eye gaze on T

02 (0.4)

03 T: yesterday (0.3) pasado
 ((tr. yesterday))
 +T moving her right hand backward

04 (0.3)

05 S3: yeah

06 (0.3)

07 T: past

08 (0.3)

09 S3: toda↑y?
 +S3 both hands pointing to the table

10 (0.2)

11 T: present
 +T pointing on the ground

12 (0.3)

13 S3: hm: present=

14 T: =um hm

15 (0.4)

16 T: tomorrow
 +T moving her right hand forward

17 (0.4)

18 T: future
 +S3 moves her hands forward

19 (0.2)

20 S3: y °después° de tomorrow?
 ((tr. and after tomorrow))
 +S3 moving her arms and moving them further forward

21 (0.4)

22 T: future
 +T lifting up her right hand and moving her right hand further forward

23 (0.7)

24 S3: future?

25 (0.4)

26 T: future (0.2) futuro=

((tr. future))

27 S3: =future
 28 (0.5)
 29 T: future=
 30 S3: =ah:
 31 (0.9)

Prior to this extract, learners were asked by T to employ the adverbials of time (i.e. ‘yesterday’, ‘today’ and ‘tomorrow’) to complete a sentence. Although student 3 (S3) was selected by T to complete the sentence, S3 seized the opportunity to ask T to clarify the meanings of these adverbials of time.

In line 3, T first utters ‘yesterday’ and code-switches to Spanish by saying ‘pasado’ to offer the equivalent of the noun ‘yesterday’. This leads to S3’s acknowledgement (‘yeah’, line 5) of T’s explanation possibly because S3 is a native Spanish speaker and she is able to understand T’s use of Spanish. Concurrently, T gestures backwards to suggest that the noun ‘yesterday’ refers to the past. S3 then initiates a follow-up question and asks T regarding the correct adverb for describing ‘today’. T responds to S3’s question by saying ‘present’ and pointing to the ground in line 11. T’s use of deictic gesture can be seen as an imitation of S3’s deictic gesture in line 9 to illustrate the present time frame.

After S3 acknowledges T’s response in line 13, T initiates a turn to offer a short explanation of ‘tomorrow’. She first moves her hand forward in line 16 and then utters ‘future’ in line 18 which implies that the meaning of ‘tomorrow’ refers to the future. However, in line 20, S3 asks another follow-up question regarding the word for describing ‘after tomorrow’. Note that S3’s employment of metaphoric gestures is referring to two different time frames. S3’s first metaphoric gesture in line 18 (i.e. moving her hands forward) is referring to the future time frame but the second one in

line 20 (i.e. moving her arms further forward) is referring to ‘the future of the future’ (Gutierrez, 1995). Hence, these metaphoric gestures are different in terms of the spatial extent. T offers an explanation to S3 in line 21 by uttering ‘future’ as well as moving her right hand further forward which is an imitation of S3’s second metaphoric gesture. This explains to S3 that the word ‘future’ itself implies the meaning of ‘after tomorrow’. Although this explanation does not lead to immediate uptake from S3, as indicated in her utterance of ‘future?’ with a rising intonation which indicates her uncertainty of T’s explanation, T offers the Spanish equivalent of ‘future’ (‘futuro’) to S3 in line 26 and this eventually leads to a display of understanding from S3, as shown by her use of change-of-state token ‘ah:’ (Heritage, 1984) in line 30.

As illustrated, T sometimes imitates S3’s gestures to complement her own explanations. By doing so, T treats S3’s gestures as an appropriate interactional resource for T to explain the target words to other learners. Nevertheless, using embodied explanations may not always work in explaining particular VIs. For example, T eventually needs to draw on her L2, Spanish, to explain the meaning of the ‘future of the future’ to S3. This serves as a good example of how T draws on her available gestural and linguistic resources to construct her VEs.

Microgenetic L2 Development: Appropriation of Teacher’s Metaphoric Gestures

The following extract illustrates how S3 display her understandings to T through appropriating T’s metaphoric gestures in a new interactional context.

Extract 2

01 T: so I am november (1.8) I am november

+T stands next to the sign of November

+T pointing to herself

02 (1.0)

03 S15: you are november

04 (0.2)

05 T: before

+T waves her right hand backward over her shoulder, palm facing downward

06 (0.2)

07 S15: before

08 (1.7)

09 T: october (0.5) after

+T remains standing next to the November sign

+T moves her right hand toward space in front at chest level,
palm facing upward

10 (1.2)

11 S3: °december°

12 (0.3)

+T pointing at the sign of December on the whiteboard

13 T: december

14 (1.9)

15 T: december

16 (0.7)

17 S3: it's (0.3) it's (0.2) como (0.3) yesterday tomorrow
((tr. like))

+S3 waves her hands backward
over her shoulder, palms facing downward

+S3 moves her
hands toward space in front at
chest level, palms facing upward

18 (0.8)

- 19 T: little bit the same
 +T holds her thumb and index finger a little way apart
- 20 (0.3)
- 21 T: [yeah] little bit the same
 +T holds her thumb and index finger a little way apart
- 22 S3: [hahaha]
- 23 (1.0)
- 24 T: um hm
- 25 (2.8)
- 26 T: before today ↑
 +T waves her right hand backward over her shoulder, palm facing downward
- 27 (0.7)
- 28 S3: ah ha
- 29 (0.3)
- 30 T: yesterday
- 31 (0.7)
- 32 S3: [yesterday]
- 33 S9: [yesterday]
- 34 (0.2)
- 35 T: yes (0.2) before today is yesterday (0.6) after today
 +T waves her right hand backward over her shoulder, palm facing downward
- +T moves her right hand toward space in front at chest level, palm facing upward
- 36 (1.1)
- 37 S3: tomorrow
- 38 (0.3)
- 39 T: tomorrow (0.7) tomorrow (0.7) ah ha

Extract 2 is taken from the lesson on 11-2-2002 approximately a month after the lesson on 7-1-2002 when T explained the meanings of ‘yesterday’, ‘tomorrow’ and ‘future’ to S3 (Extract 1). Prior to the extract, student 15 (S15) self-initiated a question regarding

the possibility to claim that ‘October is before September’. T rejected S15’s response and explained the meaning of ‘before’ to S15.

As Extract 2 begins, T first utters ‘so I am November’ and stands next to the November sign. T then utters ‘before’ and gestures backward to draw learners’ attention regarding the month which comes before November (line 5). Here, it becomes clear that T’s utterance (‘before’, line 5) is considered as a designedly-incomplete utterance (DIU) (Koshik, 2002) since T does not immediately provide a response to the learners. By doing so, T is providing thinking time for learners and inviting them to complete the utterance for T. Although S15 repeats ‘before’ voluntarily in line 7, T does not treat S15 as a preferred response. As no one offers a preferred answer during a long 1.7-second pause (line 8), T utters ‘october’ in line 65 to complete the DIU. After a 0.5-second pause (line 9), T remains standing next to the November sign and initiates another DIU by saying ‘after’ to invite learners to identify the month which comes after November. T also offers a hint for the learners by gesturing forward when she utters the DIU. In line 11, S3 quietly utters ‘december’ to respond to T’s DIU. Shortly afterwards, T points at the December sign (line 12) and utters ‘december’ in lines 13 and 15 to confirm S3’s response to the DIU.

After T offers positive feedback to S3’s response, S3 self-initiates a turn to display her understanding regarding the similarities between ‘yesterday’ and ‘before’, and ‘tomorrow’ and ‘after’. S3 first code-switches to Spanish when she utters ‘it’s (0.2) como (like)’ in line 17 to indicate her intentions to make a comparison. Afterwards, S3 utters ‘yesterday’ and simultaneously gestures backward to illustrate the past time frame. S3 then enunciates ‘tomorrow’ and concurrently gestures forward to demonstrate the future time frame (line 17). The metaphoric gestures that S3 employs share similarities with T’s use of gestures in explaining ‘yesterday’ and ‘tomorrow’ in Extract

1. T acknowledges S3's self-initiated response by explaining that S3's comparison between the two VIs is a 'little bit the same' (lines 19 and 21). T then elaborates on her commentary on S3's response in line 26. She initiates a DIU ('before today') and gestures forward to indicate the past time frame. Similar to lines 5-8, T's DIU does not lead to any response from the learners to complete the DIU in line 27 and T eventually completes the DIU by enunciating 'yesterday' in line 30. T then initiates another DIU 'after today' in line 35. Note that T also offers a hint for the learners by gesturing forward (line 35). S3 eventually responds to T's DIU in line 37 by saying 'tomorrow'. T subsequently provides positive feedback to S3 by repeating enunciating 'tomorrow' twice and uttering acknowledgement tokens 'ah ha' (line 39) to confirm S3's response.

It could be argued that S3 notices T's use of metaphoric gestures for explaining the meanings of 'before' (gesturing backward, line 5) and 'after' (gesturing forward, line 9), and appropriates these gestures to exemplify her understanding of the meanings of 'yesterday' and 'tomorrow' in line 17. Hence, S3 may not necessarily appropriate the metaphoric gestures that were visually enacted by T on 7-1-2002 (Extract 1). Nevertheless, S3 draws on the vocabulary knowledge that she learnt on 7-1-2002 and establishes the similarities between the meanings of the VIs ('yesterday' and 'tomorrow') and the target VIs ('before' and 'after') in a new interactional context. This allows T to understand S3's current states of conceptual understanding.

Case 2: Excuse me

Nature of VE: Embodied enactment as explanation

Extract 3 was taken from the lesson on 1-4-2002. It illustrates how T employs embodied enactment to create a hypothetical context to explain the meaning of 'excuse me' to her learners.

Extract 3

01 T: now (0.5) there's (0.2) excu↓se me? (0.3) >excu↓se me?<

+T pointing at the phrase 'excuse me' with high intonation on the whiteboard on the left

+T pointing at the phrase 'excuse me' with high intonation on the whiteboard on the left

02 (0.5)

03 T: and (0.8) excu ↑ se me ↓

+T pointing at the phrase 'excuse me' with low intonation on the whiteboard on the right

04 (1.4)

05 T: up (0.8) and (.) down

+T moving her right hand from low to high position

+T moving her right hand from high to low position

06 (.)

07 S10: this is no good? (0.2) this is good?

+S10 pointing at the phrase 'excuse me' with low intonation on the left blackboard

+S10 pointing at the phrase excuse me with high intonation on the right blackboard

08 (0.2)

09 T: different (0.7) different (0.3) different

+T spreads out her hands in opposite directions #1



#1

10 (0.4)

11 S10: ah

- 12 (0.2)
- 13 T: different (0.2) example (0.4) °please stand up°
 +T points at S13, making beckoning and inviting motion
- 14 (0.8)
 +S13 stands up
- 15 T: °please stand here° (0.8) °okay° (0.8) excu↑se me↓
 +T stretches out her arms and points to the ground
 +S13 walks towards T
 +T moves towards S13
- 16 (1.4)
 +T touches S13's shoulder
 +S13 moves to the right to offer space for T to walk through
- 17 S?: oh
- 18 (1.0)
- 19 S3: excuse me
- 20 (0.3)
- 21 SS: hahaha
- 22 (0.2)

Prior to this extract, T was teaching the meaning of ‘excuse me’ (i.e. asking someone to repeat their utterance) to her learners. In this extract, T aims to offer additional explanations regarding the meanings of two different intonations of ‘excuse me’. In lines 1-5, T explains to the learners that there are two ways of pronouncing ‘excuse me’: ‘excudse me?’ and ‘excuse me↓’. This leads to a follow-up question initiated by student 10 (S10) asking which pronunciation is considered as preferable (line 7). T responds to S10's question by uttering ‘different’ three times and employing iconic gestures (spreading out her hands in opposite directions, figure 1) to reinforce the difference between the two pronunciations (line 9).

Although S10 initiates a change-of-state token ‘ah’ in line 11 which possibly displays her understanding of T's explanation, T initiates a new turn in line 13 and asks student 13 (S13) to stand up which projects that T will offer an additional explanation to S10's question through the use of an example (line 13). T makes a request to S13 to ‘°please

stand up^o quietly, making a ‘standing up’ motion with her hands. S13 stands up and follows T’s request by moving closer to T (line 15), although at this stage S13 has not yet been informed of the reason for her to stand up. T establishes a hypothetical scenario by walking towards S13 in line 15 to indicate her walking direction, which signals to the class that T’s forthcoming action will be performative. T then enacts a hypothetical context by uttering ‘excuse me[↓]’ (line 15), with raising intonation on the first word ‘excuse’ and falling intonation on the second word ‘me[↓]’, and physically touching S13’s shoulder (line 16), to represent the embodied enactment itself. In line 16, S13 follows T’s request by moving to the right to offer space for T to walk through, which is a demonstration of S13’s understanding of T’s previous action. As shown, T leads the construction of the embodied enactment from lines 15-18 by acting as the pedestrian who initiates request and S13 as the passive pedestrian who moves aside.

In summary, Extract 3 illustrates how T enacts a hypothetical context through verbal and multimodal resources to facilitate learners’ understandings of ‘excuse me’. By offering verbal and physical representations of the imaginary context, T potentially allows learners to understand how the target vocabulary can be employed in real-life situations.

Microgenetic L2 Development: Appropriation of Teacher’s Embodied Enactments

Extract 4 is the subsequent part of the interaction in Extract 3 (approximately two minutes after Extract 3), which occurred on 1-4-2002. It should be noted that on 7-1-2002, T employed embodied enactments to explain how S3 could employ ‘sorry’ and ‘excuse me’ in situations when S3 wished to make her way through a crowd (Author and Brandt, 2018), which was approximately four months after the occurrence of this extract.

Extract 4

109 S10: (ah l) understand

110 (.)

111 T: hahaha

112 (0.5)

113 S13: excuse me

114 (0.2)

115 T: ah:=

116 S3: =in (0.3) in (0.2) in bus

+S3 pointing at T

117 (1.2)

118 T: in the bus

119 (0.2)

120 S3: [in in] the bus

121 T: [yeah]

+S3 raising her right hand

122 (0.2)

123 T: excu ↑ se me ↓

+T walks forward and moves her right hand to the side

124 (0.3)

125 S3: excu ↑ se me ↓

126 (0.2)

127 T: my voice goes down

+T puts her left hand near her mouth

+T moves her right hand from high to low position #2



#2

128 (0.9)

129 S3: l um ma :

+S3 holds her hands up in parallel at shoulder level and move them inward, palms half-cupped, facing each other #3



#3

130 (0.2)

131 T: many people

+T spreads out her arms outward and moves her fingers up and down continuously

#4



#4



#5

132 (0.2)

133 S3: ya (0.3) [many people] (0.5) er

134 T: [many people]

+S3 moves her arms to the sides #5

135 (0.2)

+S3 stands up from her chair

136 T: [excu ↑ se me ↓]

+T walks towards S11 and moves her hands to the opposite sides #6



#6



#7

137 S3: [excu ↑ se me ↓] (0.2) [excu ↑ se me ↓]
 +S3 turns towards S11, slowly walks towards S11 and moves her body to and fro
 #7

138 T: [excu ↑ se me ↓]

139 (0.3)

140 S11: ah=

141 S3: =[excu ↑ se me ↓] (0.4) excu ↑ se me ↓

142 S13: [ah]

143 (0.2)

144 T: uh ha (0.3) uh ha (0.3) um hm (0.2) um hm
 +S3 sits down

145 (0.4)

146 S3: yeah in the bus

Prior to this extract, T constructed embodied enactments with S13 to convey the meaning of ‘excuse me’ in terms of asking people to move aside (Extract 3). After this, T and S3 co-constructed another embodied enactment to reinforce the meaning of ‘excuse me’ for requesting someone to move aside (lines 86-101, described in Author, 2018).

As Extract 4 begins, T initiates an elongated ‘ah:’ marker in line 114 which potentially signals her plan to initiate a turn to move on to the next conversational topic. However, S3 immediately interrupts the talk and initiates a new hypothetical context by saying ‘in (0.3) in (0.2) in bus’ in line 116. By doing so, S3 is relating the phrase ‘excuse me’ with the context that she and T employed in constructing the embodied enactments on 7-1-2002. T repairs S3’s initiative by adding a preposition ‘the’ (‘in the bus’) in line 118. Note that typically-speaking it is more common to say ‘on the bus’ rather than ‘in the bus’ in everyday English. In line 123, T utters ‘excu↑se me↓’, accompanied with her body movements by walking forward and moving her right hand to the side to represent an act of pushing. By doing so, T potentially aims to take the opportunity to reinforce the target phrase that was taught earlier (i.e. Extract 3). T takes

another turn in line 127 to provide an explanation of the intonation of ‘excu↑se me↓’ by stating ‘my voice goes down’, accompanied by performing iconic gestures (moving her right hand from high to low position, figure 11) to indicate the change of intonations.

In line 129, S3 produces unintelligible utterances (I um ma). Although S3 fails to search for appropriate verbal expressions to articulate her thoughts, T is able to comprehend S3’s verbal responses, possibly because T pays attention to S3’s use of metaphoric gesture (holding up her hands in parallel and moves them inward) in line 129. Subsequently, T repairs S3’s utterance by uttering ‘many people’ in line 131, accompanied by T’s metaphoric gesture as T spreads out her arms and moves her fingers up and down continuously in line 131 (figure 13). By doing so, both T and S3 are co-establishing the imaginary context prior to the enactment. In line 135, S3 first stands up from her chair and then turns to student 11 (S11) and walks towards him. Simultaneously, S3 slightly moves her body to and fro (line 137, figure 16). Here, S3 is enacting a person who is making her way on the bus and S11 spontaneously becomes a passenger who prevents S3 from moving along.

Concurrently, T also constructs her own embodied enactment while S3 is enacting. In line 137, T utters ‘excu↑se me↓’ simultaneously with S3 (line 136), walks towards S11 and moves her hands to the opposite sides to demonstrate a request for a person to move aside (figure 15). By doing so, T is reiterating the meaning of ‘excu↑se me↓’ to her learners through embodied enactment again. Note that both S11 and S13 initiate ‘ah’, a change-of-state token, in lines 140 and 142 respectively. This potentially demonstrates S11’s and S13’s claim of understanding of T’s and S3’s embodied enactments. T acknowledges S3’s embodied enactment through uttering ‘uh ha’ and ‘um hm’ repeatedly in line 144 and S3 closes the sequence by saying ‘yeah in the bus’

which emphasises the hypothetical context that S3 draws on constructing her embodied enactment.

As shown, by re-introducing this context four months later, S3 remembers the hypothetical context employed by T and appropriates it in the current lesson to demonstrate her understanding of the meaning of ‘excuse me’. Although T’s embodied enactments on 7-1-2002 are similar to S3’s embodied enactment in this extract, T typically used her hand to touch on the learner’s shoulder (e.g. Extract 3, lines 15-16) or moved her hands to the opposite sides to signal the need for the person to move aside (e.g. Extract 4, line 136). Nevertheless, examining S3’s embodied enactment in this extract reveals how S3 creatively adapts T’s use of gestural resources slightly (i.e. as S3 is moving her body to and fro) to represent her action of walking through the bus. This displays her current knowledge of the meaning of ‘excuse me’.

Microgenetic L2 Development: Providing a Description of a Past Event

This extract was taken from the lesson on 15-4-2002. On 7-1-2002, T explained the meaning of ‘excuse me’ to her learners (described in Author and Brandt, 2018). Note that T explained the meaning of ‘excuse me’ again on 1-4-2002 (Extract 4), which was two weeks after the lesson containing this extract.

Extract 5

- 01 (5.0)
 +S11 and S17 are walking on the same direction
 +S11 hits S17's shoulder
 +S3 glances at S11
- 02 S3: excuse me
 +T glances at S3
 +S3 points at S11
 +S3 points at S17
- 03 (1.1)
- 04 S11: >我就這樣了講<
 ((wǒ jiù zhè yàng liǎo jiǎng))
 ((tr. I am going to stay seated))
 +S3 points at S11 repeatedly
 +S11 talks to S13
- 05 (0.2)
- 06 S3: excuse me (0.3) excuse me
 +S3 points at S17
 +S3 points at S17 again
- 07 (0.2)
- 08 T: [excuse me]
- 09 S3: [pheu:]
 +S3 nudges S11 with her elbow
- 10 (0.4)
- 11 S11: oh
- 12 (0.2)
- 13 S3: (ellos)
 ((tr. they))
 +S3 first points at S11 and then points at S17
- 14 (1.0)
- 15 S3: (ellos) (0.2) hu:
 ((tr. they))
 +S3 nudges S11 with her elbow
- 16 (.)
- 17 S11: sorry
 +S11 glances at S17
- 18 (1.0)
- 19 T: good (0.2) sorry (0.2) [hahaha]
- 20 S11: [hahaha]=

21 S3: =yeah excuse me

22 (0.2)

Prior to this extract, the learners worked on a vocabulary task individually. They were then asked by T to write their answers on the whiteboard. After that, learners returned to their seats.

As Extract 5 begins, both S11 and student 17 (S17) are walking together and S11 hits S17's shoulder accidentally (line 1). S3 witnesses the situation and glances at S11 during the pause. S3 then utters 'excuse me' and simultaneously uses deictic gesture (first pointing at S11, then pointing at S17) to remind S11 to say 'excuse me' to S17, who accidentally gets hit by S11. S11 takes up the turn in line 4 and explains in Chinese that he will stay seated. Although S3 does not speak Chinese, S3 first points at S11 continuously in line 4, and then repeats 'excuse me' twice and repeatedly points at S17 in line 6 to emphasise the need for S11 to apologise to S17.

In line 8, T utters 'excuse me' voluntarily although at this point she has not been informed what has happened in the classroom. Concurrently, S3 produces a noise 'phew' in line 9 to imitate the sound of pushing, and instantaneously nudges S11 with her elbow to illustrate the act of hitting a person. Although S11 utters a change-of-state token 'oh' which potentially displays his understanding of S3's action, S3 continues to explain the past event to T in line 13 by saying 'ellos (they)' in Spanish as well as pointing at S11 and S17 to illustrate that both S11 and S17 are involved in this incident. Nevertheless, there is a significant 1.0-second pause in line 14 which indicates that S3 may expect a response from T. S3 then repeats 'ellos (they)' and produces a noise 'hu:' while simultaneously nudging S11 with her elbow again in line 15. This self-repair-initiated turn reinforces the fact that both S11 and S17 accidentally hit each other. This

leads to S11's display of an understanding, indicated by S11's glancing at S17 and his utterance 'sorry' to S17. T provides positive feedback to S11's response by saying 'good' and repeating S11's response 'sorry' in line 19 to confirm the appropriateness of S11's response. S3 acknowledges T's positive feedback by uttering an acknowledgement token 'yeah' and repeating the phrase 'excuse me' in line 21.

In summary, by identifying the need for S11 to utter 'excuse me' to S17, S3 displays her understanding of the meaning of 'excuse me' (i.e. requesting someone to move aside). Based on the evidence from Extract 4 and this extract, it illustrates that S3 can produce the phrase 'excuse me' in her own spontaneous speech without receiving any assistance from T. This displays her gains in gradual control in using this phrase contingently and accurately.

Discussion and Conclusion

In this paper we identified two interactional features of VE sequences in a beginning-level adult ESOL classroom: embodied explanations (Extract 1) and embodied enactments as explanations (Extract 3). Our analyses showed that T draws heavily on gestural resources in explaining VIs; a finding which closely aligns with Lazaraton (2004) and Smotrova and Lantolf (2013) where gestural resources were shown to allow teachers to visualise the contextual meaning of the VIs thus making the VE more comprehensible for learners who have limited L2 repertoire. The analysis of embodied explanations in Extract 1 demonstrated how T's use of deictic and metaphoric gestures helped visualise the adverbs for expressing time. Furthermore, the analysis of T's embodied enactments in Extract 3 showed that gestural resources (e.g. physical actions) are not only used to offer visual illustrations of the meanings of VIs but that they can be employed to create hypothetical contexts that facilitate learners' deeper understanding

of VIs; a finding which is in line with Author and Brandt (2018). In representing the action of bumping into S13 physically and verbally, T circumvents the possibility of students misunderstanding the abstract explanations of 'excuse me' thus allowing students to observe how the phrase 'excuse me' can be employed in a real-life context. These findings can provide insights for the teaching and explaining of VIs in L2 beginner classrooms.

The analysis also illustrated how learners' understandings of the meanings of the target VIs can be evidenced through their appropriations of T's metaphoric gestures (Extract 2), embodied enactments (Extract 4), and their re-using of taught target phrases to describe a past event (Extract 5). These findings support Matsumoto and Dobs's (2017) claim and demonstrate that the students' appropriations of T's metaphoric gestures and embodied enactments are employed to externalise their internal thinking processes, particularly in instances where the linguistic mode of explanation is not as rich as the gestural one, thus rendering them visible to the teachers and peers for inspection. Moreover, S3's description of a past event is mediated by her use of various gestural and linguistic resources (e.g. deictic gestures, use of L1, and sound of pushing) in order to accurately narrate the incident to T. Furthermore, the appropriation of T's embodied enactment as evidence of L2 microgenetic development further operationalises the construct of 'embodied enactment' (Author and Brandt, 2018). This study demonstrated that employing embodied enactments does not only allow the teacher and learners to bridge the gap between classroom interaction and L2 use outside the classroom, as demonstrated in Extract 3. Rather, as illustrated in Extract 4, embodied enactments can also be used as interactional resources for learners to visualise their conceptual understandings in progress and provide valuable diagnostic

information for the teacher and facilitate the evaluation of learners' current knowledge states.

As mentioned earlier, L2 development has typically been studied through employing traditional quantitative methods, including pre- and post-tests, to solely assess the learners' outcomes of performance. Reflecting the Vygotskian internalization processes (Lantolf and Thorne, 2006), CA's focus on sequential details of L2 illustrates how L2 development is a gradual process which requires learners to get acquainted with the target VIs and employ relevant verbal and multimodal resources contingently in a range of different but relevant situations to display their conceptual understandings of the meanings of target VIs.

The analysis has revealed that CA allows researchers to observe *some*, albeit incomplete, evidence of L2 development. Throughout the analysis, S3 is the learner with the most evident display of understandings through interactions. Therefore, it is unclear whether other learners in the classroom have benefited from T's VEs. In addition, it is possible that there are other factors including personality, motivation, personal history affecting learners' L2 development. Furthermore, learners' use of language can be restricted by the nature of classroom activities which do not necessarily provide opportunities for displaying understandings of the meanings of particular VIs. Hence, only tracing learner's spoken discourse over a period of time may not offer a full picture of learners' L2 developmental trajectories; learners' cognitive development may not always be publicly observable during interactions. This limitation is of course not only restricted to this study; it is virtually impossible for SLA studies to fully examine the transition that the learners undergo, as L2 development can occur outside formal classroom settings and can also occur internally within learners with no explicit external

manifestation. Therefore, it is reasonable to claim that using CA can reveal *some* aspects of learners' L2 development in a specific sociocultural activity.

Implications

Methodologically, this study has helped advance the field by demonstrating how the use of CA to conduct multimodal analysis can shed light on the role of gestures in the explanation of VIs. Second, the study shows how using CA to trace L2 development over time can help paint a richer picture of students' learning processes. It captures how gestures mediate the emergence of vocabulary knowledge over a lengthy period of time, as a result of moving from the inter-mental plane (i.e. interaction through using various gestural and linguistic resources) to the intra-mental place (i.e. knowledge growth and appropriation). CA allows us to depict how the achievement of conceptual understanding can be an embodied activity, indicating that gestural resources can work as a 'window' to understand the current state of a learner's knowledge in the learning process. This study emphasises the need to examine both verbal and gestural elements together when considering students' developmental processes of L2 vocabulary knowledge.

This study has shown that L2 learners pay attention to the teacher's use of gestural resources and can appropriate them to display their understanding of L2 knowledge. A possible suggestion therefore is to enhance teachers' awareness of the positive effects of their use of gestural resources to facilitate learners' L2 learning and development.

It is worth noting that learners' cultural backgrounds can affect their willingness to employ embodied resources in the classrooms (e.g. So, 2012) and it is important for teachers to acknowledge the socio-cultural background of the learners as a factor in affecting their use of gestural resources. Nevertheless, studying both verbal and gestural

elements together is necessary when considering L2 learners' developmental processes of L2 vocabulary knowledge, as limited linguistic resources may translate into learners' heavier reliance on gestures as means of communication. A consideration of both elements together can provide teachers with a better picture of learners' knowledge states which they can accordingly draw on to modify their instructions and provide appropriate scaffolding for learners.

Notes

1. Conducting a timed analysis was beyond the scope of this paper and it is impossible to report exact percentages of time dedicated to VI explanations. This is based on our judgment from observations of various hours of data.

References

Author and B.L.M. Poon. 2016.

Author, 2018.

Author and A. Brandt. 2018.

Eskildsen, S.W. and J. Wagner. 2015. 'Embodied L2 Construction Learning'. *Language Learning* 65 (2): 268-297.

Gutiérrez, M. 1995. 'On the future of the future tense in the Spanish of the southwest'. In *Spanish in four continents: Studies in language contact and bilingualism*, ed. C. Silva-Corvalán, 214-223. Washington: Georgetown University Press.

Hauser, E. 2011. 'On the danger of exogenous theory in CA-for-SLA: A response to Hellermann and Cole (2009).' *Applied Linguistics* 32 (3): 348-352.

Heap, J.L. 1997. 'Conversation analysis methods in researching language and education'. In *Encyclopedia of language and education: Vol. 8. Research methods in language and education*, ed. N.H. Hornberger and D. Corson, 217-225. Boston, MA: Kluwer Academic Publishers.

Heritage, J. 1984. *Garfinkel and ethnomethodology*. UK: Blackwell Publishers.

Hudson, A. 2011. *Teacher gesture in a post-secondary English as a second language classroom: A sociocultural approach*. PhD thesis, University of Nevada.

Jefferson, G. 2004. 'Glossary of transcript symbols with an introduction'. In *Conversation Analysis: Studies from the first generation*, ed. G. Lerner, 14-31. Philadelphia: John Benjamins.

Koshik, I., 2002. 'Designedly incomplete utterances: A pedagogical practice for eliciting knowledge displays in error correction sequences.' *Research on Language and Social Interaction* 35 (3): 277-309.

Lantolf, J. P., and S. Thorne. 2006. *Sociocultural theory and the genesis of second language development*. UK: Oxford University Press.

Lazaraton, A. 2004. 'Gestures and speech in the vocabulary explanation of one ESL teacher: A microanalytic inquiry.' *Language Learning* 54 (1): 79-117.

Matsumoto, Y. and A. Dobs. 2017. 'Pedagogical gestures as interactional resources for teaching and learning tense and aspect in the ESL grammar classroom.' *Language Learning* 67 (1): 7-42.

McCafferty, S.G. 2002. 'Gesture and creating zones of proximal development for second language learning.' *Modern Language Journal*, 86 (2): 192–203.

McCafferty, S.G. 2004. 'Space for cognition: Gesture and second language learning.' *International Journal of Applied Linguistics* 14 (1): 148–165.

Mori, J. 2004. 'Negotiating sequential boundaries and learning opportunities: A case from a Japanese language classroom.' *Modern Language Journal* 88 (4): 536-550.

Mondada, L., and S. Pekarek Doehler. 2004. 'Second language acquisition as situated practice: Task accomplishment in the French second language classroom'. *Modern Language Journal* 88 (4) 501–518.

Mondada, L. 2018. 'Multiple temporalities of language and body in interaction: Challenges for transcribing multimodality.' *Research on Language and Social Interaction* 51 (1): 85–106.

Morton, T. 2015. 'Vocabulary explanations in CLIL classrooms: A conversation analysis perspective.' *Language Learning Journal* 43 (3): 256-270.

Nation, I.S.P. 1990. *Teaching and learning vocabulary*. New York: Newbury House.

Ohta, A. 2000. 'Rethinking interaction in SLA: Developmentally appropriate assistance in the zone of proximal development and the acquisition of L2 grammar'. In *Sociocultural theory and second language learning*, ed. J.P. Lantolf, 51–78. UK: Oxford University Press.

Pekarek Doehler, S., and V. Fasel Lauzon. 2015. 'Documenting change across time: longitudinal and cross-sectional CA studies of classroom interaction'. In *Handbook of classroom interaction*, ed. N. Markee, 409-424. New Jersey: Wiley-Blackwell.

Reder, S. 2005. "The 'Lab School'". *Focus on basics: Connecting research and practice* 8: 1-7.

Schmitt, N. 2000. *Vocabulary in language teaching*. UK: Cambridge University Press.

Sert, O. 2017. 'Creating opportunities for L2 learning in a prediction activity.' *System* 70: 14-25.

Sidnell, J. 2010. *Conversation analysis: An introduction*. West Sussex, UK: Wiley Blackwell.

Smotrova, T., and J. P. Lantolf. 2013. 'The function of gesture in lexically focused L2 instructional conversations.' *Modern Language Journal* 97 (2): 397–416.

Storch, N. 2002. 'Patterns of interaction in ESL pair work.' *Language Learning* 52 (1): 119-158.

So, W.C. 2012. 'Cross-cultural transfer in gesture frequency in Chinese–English bilinguals.' *Language and Cognitive Processes* 25 (10): 1335-1353.

ten Have, P. 1990. 'Methodological issues in conversation analysis'. *Bulletin of Sociological Methodology* 27 (1): 23-51.

ten Have, P. 2007. *Doing Conversation Analysis (2nd Edition)*. London: Sage.

Tian, L. and E. Macaro. 2012. 'Comparing the effect of teacher codeswitching with English-only explanations on the vocabulary acquisition of Chinese university students: A lexical focus-on-form study.' *Language Teaching Research* 16 (3): 361 – 385.

van Compernelle, R.A. 2010. 'Incidental microgenetic development in second-language teacher-learner talk-in-interaction.' *Classroom Discourse* 1 (1): 66-81.

Vygotsky, L.S. 1962. *Thought and language*. Cambridge, MA: MIT Press.

Waring, H.Z., C.C. Creider and C.D. Box. 2013. 'Explaining vocabulary in the second language classroom: a conversation analytic account.' *Learning, Culture, and Social Interaction* 2 (4): 249-64.

Word Count: 8225

Appendix: CA Transcription Conventions (Adapted from Jefferson, 2004)

Sequential and Timing Elements of the Interaction

[Beginning point of simultaneous speaking (of two or more people)
]	End point of simultaneous speaking
=	Talk by two speakers which is contiguous (i.e. not overlapping, but with no hearable pause in between)
	OR
	continuation of the same turn by the same speaker even though the turn is separated in the transcript
(0.2)	The time (in tenths of a second) between utterances
(.)	A micro-pause (one tenth of a second or less)

Paralinguistic Elements of Interaction

wo:rd	Sound extension of a word (more colons = longer stretches)
word.	Fall in tone (not necessarily the end of a sentence)
word,	Continuing intonation (not necessarily between clauses)
wor-	An abrupt stop in articulation
word?	Rising inflection (not necessarily a question)
<u>word</u>	(underline) Emphasised word, part of word or sound
word↑	Rising intonation
word↓	Falling intonation
°word°	Talk that is quieter than surrounding talk
hh	Audible out-breaths
.hh	Audible in-breaths
w(hh)ord	Laughter within a word

>word<	Talk that is spoken faster than surrounding talk
<word>	Talk that is spoken slower than surrounding talk
\$word\$	Talk uttered in a 'smile voice'

Other Conventions

(word)	Approximations of what is heard
((comment))	Analyst's notes
#	Indicating the exact locations of the figures in the transcripts
+	Marks the onset of a non-verbal action (e.g. shift of gaze, pointing)
XX	Inaudible utterances

For Peer Review

Author's Response to Reviewer Comments (Anonymous)

Reviewer 1's Comments	Authors' Response to the Comments
<p>This is a very interesting paper which examines how teachers mediate ESOL students' learning of vocabulary through talk and gestures. The research combines CA with SCT, which breaks new ground conceptually and methodologically. The microgenetic analysis of how teachers' instructions are socially and interactionally organised over a lengthy period of time is interesting and original and I applaud such a good attempt to advance the field. However, there are still a number of important issues that need to be addressed properly before the paper can be published. The main problem, in my view, is that the authors aim to achieve too many goals in one paper, so the strengths become weaknesses. In the following, I outline some suggestions for revision, some of which entail substantial restructuring and rewriting, but I hope these constructive comments are helpful for improving the overall quality of the paper.</p>	<p>We thank Reviewer 1 for all their insightful comments. We have taken most of the comments into account and have made changes accordingly which we believe has greatly improved the overall quality of our paper.</p>
<p>(1) The overall conceptualisation of the paper is sound which seeks to understand, from a sociocultural and discursive perspective, how teachers organise their instruction and how different organisations of talk and gestures contribute to ESOL students' vocabulary knowledge development. The literature review focuses on two areas. The first part reviews SCT's</p>	<p>Following the reviewer's comment, we decided to change the paper title as follows:</p> <p>“The Mediation and Organisation of Gestures in Vocabulary Instructions: A Microgenetic Analysis of Interactions in a Beginning-level Adult ESOL Classroom”</p>

<p>contributions to SLA and learning more broadly; the second part focuses on the role of gestures in SLA and learning. Based on what you attempt to do in the analysis, I wonder whether it might be useful to reposition your research as a sociocultural study of gestures in SLA. As it currently stands, the research straddles between a sociocultural study of talk and a sociocultural study of gestures. The literature on oracy/dialogic teaching and that of gesture are very different although both could be used to frame the research. Given the analysis that you have conducted, I would strongly recommend that you retitle the paper as ‘the mediation and organisation of gestures in teachers’ vocabulary instructions in the beginning-level adult ESOL classroom: A micro-genetic analysis’ or something along the lines.</p>	<p>We agree with the reviewer that the paper should be repositioned to focus more strongly on teacher’s and students’ use of gestures and body movements in the classrooms. However, we also believe that talk and gestures cannot be separated in the analysis because both linguistic and gestural resources play a role in shaping the talk. Particularly, in introducing the concept of embodied enactment (Author and Brandt, 2018) the authors assert that linguistic and semiotic resources cannot be separated for analysis. The conceptualisation of embodied enactment goes beyond the idea that verbal explanations are aided by gestural conduct. Teachers create imaginary contexts through using bodily resources and linguistic resources.</p> <p>Therefore, throughout the paper, we refer to “gestural and linguistic resources” when necessary. We also explain why we only focus on analysing gestures in our paper. This is clearly stated on page 9: “While both gestural and linguistic resources play an important role in mediating L2 learning and development, the focus of this paper is predominantly on gestures; this is to reflect our study’s context – a beginning-level classroom – where linguistic resources are limited and there is heavier reliance on gestures to achieve communicative goals.”</p>
<p>(2) In light of the new conceptualisation, the literature review needs to be rewritten to include more empirical studies</p>	<p>We agree with the reviewer’s comment and have expanded the literature on sociocultural studies of gestures in SLA. This can</p>

on gestures. You can largely keep what you have written, but the section needs to be expanded and the research needs to be situated more clearly within the established field of gesture (see early works on talk and gesture in SLA and SCT).

be found under the sub-section (The role of gestures in L2 learning and development, pages 6-7).

“McCafferty (2002; 2004) examined the communicative and cognitive functions of L2 learners’ use of gestures when speaking in an L2 outside the classroom. McCafferty (2002) illustrated that gestures enhanced comprehension in social interactions between native and non-native English speakers. McCafferty (2004) also found that several types of gestures including representation gestures (e.g. iconic and deictic gestures) and beats were employed by a learner to elicit the correct L2 VI from the researcher and illustrate the meaning of the VI that the learner could not verbalise.

A recent study by Matsumoto and Dobs (2017) demonstrated how teachers employed deictic and metaphoric gestures as resources to make abstract concepts related to English tense and aspect concrete and visible to learners. Most importantly, when examining the interaction which occurred a week after the lesson, it was noticeable that the learners appropriated the teachers’ deictic gesture to demonstrate their understanding of present tense, which served as evidence of their microgenetic development. However, the authors only analysed one excerpt to explain the effects of gestures on learners’ L2 development.

	<p>A study by Smotrova and Lantolf (2013) explored the mediational function of a teacher's use of gestures in clarifying Russian learners' understanding of English VEs. The analyses indicated that the teacher employed a variety of gestures (metaphoric and ironic) to visualise the contextual meaning of the words. It was shown that the learners appropriated and modified the teacher's gestures indicating their improved understanding of the new concepts. However, the time frame of the microgenetic development was limited to one lesson only.</p> <p>Focusing on the case of Carlo – an adult Mexican Spanish-speaking learner of English – Eskildsen and Wagner (2015) conducted a longitudinal analysis of how Carlo's use of gestures developed over time in an ESOL classroom. The analysis of his use of pronouns ('under' and 'across') indicated that Carlo appropriated the teacher's gestures to demonstrate his understanding as illustrated in his spontaneous use of the pronouns accompanied with the same or similar gestures as when he first encountered the pronouns in class.</p> <p>The studies discussed above indicated that studying both learner and teacher gestures in the L2 classroom is a significant field of enquiry which requires much more attention from SLA researchers. More evidence from the analysis of speech and gestures together is needed to demonstrate the potential for CA</p>
--	--

	in tracking learners' developmental processes.”
<p>(3) Given the data are video-based, I wonder whether it might be useful to draw on multimodal analysis which might provide you with more current ideas of how to analyse semiotic resources including talk and gestures. Indeed, CA is a well-established methodology which has been used to research different professional and educational settings. I also appreciate that there is a good attempt in SLA recently to use CA to examine sociocultural issues. However, although I applaud this new development, I also wonder to what extent there might be epistemological tensions between CA and SCT. CA is a bottom-up approach to sociology, while SCT, by nature, is a socio-cognitive theory of development. They are not necessarily compatible conceptually and methodologically, which raises a fundamental question - what is the ultimate goal that the analysis aims to achieve. I appreciate that SCT and micro-genetic analysis usually involve analysis of discourse, but the focus is usually on emergence of thinking and cognitive development (cf. the focus of CA is on how social institutions are talked into being). The overlapping area seems to be social cognition and discursive construction of social cognition, but I doubt this is what your research is about.</p>	<p>1. Regarding whether it is useful to draw on multimodal analysis in our study, we argue that CA is a methodology which allows for the analysis of participants' verbal and non-verbal utterance in talk. Thus, using video-based data allows us to conduct multimodal conversation analysis in our study. We argue that this study follows a Conversation Analytic methodology by also using a multimodal approach (Mondada, 2018) to the data. We examined detailed transcriptions of the sequences by focusing on both the sequential organisation of the interaction and different multimodal resources enacted by the participants such as gaze, gestures, body movements and use of classroom artefacts.</p> <p>In order to make this clearer for the reader, we have added the following explanations in the Methodology section (page 11):</p> <p>“Several symbols were used to represent the non-verbal actions (see Appendix). It is common for CA researchers to conduct multimodal analysis by including non-verbal conduct, and screen-shots of relevant actions captured in the video-recordings in order to focus on various semiotic resources employed by the participants including gestures and body movements (Mondada, 2018). After transcribing the data, we carried out line-by-line analyses of the sequences which entailed teacher's VEs and</p>

	<p>learners' transfer of knowledge by focusing on both the sequential organisation of talk and the different multimodal resources enacted by the classroom participants.”</p> <p>2. Regarding the issue of combining CA with SCT, we agree with the reviewer's comment about the epistemological tensions between CA and SCT. In our original version, we briefly touched on this tension under the sub-section (CA as the methodological tool for conducting microgenetic analysis). We want to emphasise that CA and SCT offer complementary elements in relation to our understanding of everyday activities and of cognitive processes. CA focuses on the way participants achieve social actions (e.g. learning) and SCT emphasises the sociocultural dimensions of activities and cognitive development.</p> <p>Thus, in answering the question: “What is the ultimate goal that the analysis aim to achieve”, the CA analysis aims to provide empirical basis for making claims regarding learners' internalisation of target vocabulary knowledge. When conducting CA analysis, the insights from SCT will only be brought into the discussion section when relevant. The analysis of the classroom data is completed within a CA framework. In other words, the CA analysis itself has to be proceeded as any other CA studies, with line-by-line analysis.</p>
--	--

	<p>Therefore, in the sub-section (CA as the methodological tool for conducting microgenetic analysis), we further explained the complementarity of CA with SCT on pages 5-6:</p> <p>“The combination of CA and SCT has been a key issue in CA research. A number of scholars (e.g. Hauser, 2011) have pointed to problems inherent in combining CA with an external theory arguing that analyses may be led by a pre-determined theory rather than by observations in interactions (see Hauser, 2011 for further discussion). The complementarity of CA and SCT has been extensively discussed elsewhere (e.g. Mondada and Pekarek Doehler, 2004). In short, scholars have argued that CA focuses on the way participants systematically organise their social actions through interactions, while SCT focuses on the way human activity and mental development are mediated by socially-constructed cultural artefacts. Since CA does not provide a theory of learning, CA’s focus on talk-for-action can offer the empirical basis for making claims about mediation, learning and development (van Compernelle, 2010). However, CA is not ‘merely an analytic tool in the service of [SCT]’ (Mondada and Pekarek Doehler, 2004: 504) and SCT is not merely ‘an explanatory theory of learning in the service of CA’ (van Compernelle, 2010: 69). Rather, CA’s analytic approach allows researchers to explore SCT notions from the participants’</p>
--	---

perspective. They are therefore complementary enabling researchers to understand the intricacies of human activity and psychological development. Adopting CA as a methodological tool can potentially allow researchers to trace the process of learning which leads up to the appropriation of a previously socially elaborated feature, including a vocabulary or grammatical item, within a new interactional context.”

We also explained how we integrated CA with SCT in the methodology section on page 10:

“In line with the SCT perspective which perceives L2 learning and development as mediated in and through interaction (e.g. Lantolf and Thorne, 2006), this study combines the analytic methods of CA with SCT in order to trace the qualitative changes in learners’ microgenetic development. Since this study aims to investigate learners’ participation and teacher’s pedagogical practices in real time, this study is classified as an applied CA study which ‘tells us how to look, and what we must do in order to show how the features of institutions, like education, are produced in situ, in real time’ (Heap, 1997: 223). Although scholars (e.g. Hauser, 2011) caution against using exogenous theories to inform CA analysis, the insights from SCT will only be brought into the discussion section when relevant. The analysis of the classroom data is completed within

	<p>a CA framework. As Heap (1997) argued, applied CA studies can discuss the data with regard to exogenous theories, before and after conducting the CA analysis. Nevertheless, the CA analysis itself is to be proceeded as any other CA studies, with line-by-line analysis.”</p>
<p>(4) Clarification for the CA approach is therefore needed. As it currently stands, the discussion is too encompassing. Is it pure CA, applied CA or ethnomethodology? Would it be more useful to draw upon sociocultural discourse analysis instead, which is more compatible with SCT anyway? If Sociocultural Discourse Analysis or SDA is used, then the focus of analysis should be knowledge development, and in this case, the development of vocabulary knowledge.</p>	<p>We have now expanded our methodology section stating that our study is classified as an applied CA study. This can be seen on page 10:</p> <p>“Since this study aims to investigate learners’ participation and teacher’s pedagogical practices in real time, this study is classified as an applied CA study which ‘tells us how to look, and what we must do in order to show how the features of institutions, like education, are produced in situ, in real time’ (Heap, 1997: 223).”</p> <p>We also thank the reviewer’s suggestion of considering using Sociocultural Discourse Analysis (SDA) in our study. SDA is based on a sociocultural perspective on the nature and functions of language, thinking and social interaction (Mercer, 2005). However, SDA does not only refer to one particular method, such as the qualitative, interpretative procedure but to the methodology as whole (which involves several methods, both qualitative and quantitative). It differs from linguistic discourse</p>

	<p>analysis (e.g. CA) in being less focused on language itself and more on its functions for the pursuit of joint intellectual activity.</p> <p>The goal of using CA in this study is to explore how teaching, learning and cognitive development are accomplished in the classroom through multimodal practices which are embedded, publicly displayed in actual interaction. We would therefore argue that CA provides the most appropriate methodological framework for us to conduct detailed analysis of participants' use of linguistic and gestural resources in understanding and internalising the target language.</p>
<p>(5) The methodology section is relatively weak, lacking a lot of details of the research process. More information about the corpus and the methodology (data collection, analysis and presentation) is needed. The section needs to be substantially expanded, addressing three key questions: (1) how are the three main discourse features derived from the corpus and in what way are they related to the overall focus of the paper? (2) How representative are the excerpts selected, given a large corpus covering a period of four months? (presumably a lot of data, in terms of time, number of episodes and units of analysis?) (3) What analytical tools (linguistic CA devices) are actually employed in the analysis (e.g. turn-taking, sequence of organisation, repair? What else?) and how do they work together to accomplish the microgenesis of learning? A sentence which</p>	<p>We have now expanded our methodology section to address the concerns raised by the reviewer.</p> <p>1. Addressing the first question (“how are the three main discourse features derived from the corpus and in what way are they related to the overall focus of the paper?”), we have now added the following explanation on page 10 to explain what unmotivated looking means and how this notion allows us to come up with the three discourse features:</p> <p>“The first stage of analysis involved taking a stance of ‘unmotivated looking’ (Mori, 2004) as the guiding principle when reviewing the video-recordings from MAELC. This required us to watch multiple classroom-videos without any</p>

says that ‘the analysis follows the CA conventions (Jefferson, 2004)’ is not sufficient.

particular interest or research focus before conducting the exploratory analysis. In this process, we identified three noticeable features: firstly, the teacher devoted a significant amount of time in explaining unfamiliar VIs to her learners¹, not least because of the limited linguistic resources shared between the participants. The learners, on the other hand, displayed some ability to understand the teacher’s English although the English they produced was somehow limited. Secondly, and as a result of the learners’ limited English abilities, the teacher’s VEs drew heavily on embodied resources, and this was considered worthy of further analyses. Thirdly, some learners demonstrated their understandings of the VIs – that were previously explained by the teacher – in subsequent lessons. As Sidnell (2010) argues, once an interesting phenomenon has been identified, researchers can build up a collection of similar occurrences so that differences and similarities between each occurrence could be examined. This motivated us to review more video-recordings in order to locate further instances that entailed a transfer of knowledge from an earlier lesson to subsequent individual performances.”

2. Addressing the second question (“How representative are the excerpts selected, given a large corpus covering a period of four months? (presumably a lot of data, in terms of time, number of episodes and units of analysis?”), we have added the following

	<p>explanation on page 12:</p> <p>“For reporting purposes, we can only present illustrative extracts; this can raise concerns regarding the extent to which selected extracts are an adequate representation of all analysed data. It is therefore important to examine all extracts for similar and/or deviant instances. (ten have, 1990). In this study, the chosen extracts are typical VE sequences and characterise instances of learners’ transfer of L2 knowledge that were found in the whole collection. Atypical VE sequences were not identified in the data. It is also worth noting that the goal of CA analysis is to find the ‘devices’, or ‘the technology of conversation’ in the speakers’ situated interaction, instead of justifying the best possible representative extracts (ten Have, 1990). Therefore, as long as the selected extracts can address the research questions to reveal the relevant ‘orderliness’ with their representative nature, this can be said, to a large extent, that the representativeness is sufficient, or the research findings can be reliable.”</p> <p>3. Addressing the third question (“What analytical tools (linguistic CA devices) are actually employed in the analysis (e.g. turn-taking, sequence of organisation, repair? What else?) and how do they work together to accomplish the microgenesis of learning?”), we have added the following explanation on page</p>
--	--

	<p>12:</p> <p>“When undergoing the examination of the sequential organisation of VEs and interactional features of learners’ transfer of knowledge, we began with ‘finding patterns and explicating their logic’ (ten Have, 2007: p.120). The CA’s interactional mechanisms (i.e. turn-taking, adjacency pairs, preference organisation and repair) were applied to carry out the analysis.”</p> <p>However, the second-part of the question: “how do they work together to accomplish the microgenesis of learning?” is precisely what we had set out to do throughout the analysis section and therefore it is not possible to direct the reviewer to a particular section but emphasise that this has been our overall approach.</p>
<p>(6) The finding section needs to be reshaped in light of the above. More specifically, I would suggest that you restructure the findings in two sections. The first section focuses on the types of gestures and the process of gestural mediation. The discussion needs to address three aspects: what types of gestures are observed? What do different gestures aim to mediate? How is mediation discursively accomplished? The second section focuses on the temporal dimension of teaching and learning by</p>	<p>We completely agree with the reviewer’s comment and have now restructured the findings section to make it more coherent. We have divided it into two cases: Case One: Today, Tomorrow and Yesterday and Case Two: Excuse me.</p> <p>In case one, we have presented Extract 1 which illustrates the nature of embodied explanation. In this section, we stated that “T employs deictic and metaphor gestures to teach the pronouns</p>

explicitly scrutinising vocabulary knowledge. By tracing how different knowledge points are taken up, appropriated and further developed, you can then make a bigger claim regarding sociocultural theorisation of SLA. What remains to be done in the finding chapter is to show the trajectory of development of vocabulary knowledge. Practically, what you need to do is to identify one or two examples of vocabulary knowledge which are discussed over the four months and then show how the learning/appropriation is facilitated by gestures (and talk) over this period. This is where the originality of your research lies; as you indicate, this is the major contribution of the paper, which has yet been achieved successfully in my opinion.

for expressing time.” (p.13). Then we presented one extract (Extract 2) on page 16 which demonstrated microgenetic L2 development (Appropriation of Teacher’s Metaphoric Gestures).

Similarly, in case two, we first presented Extract 3 (page 20) to explain how embodied enactment is interactionally organised and constructed and how the phrase ‘excuse me’ is explained through using embodied enactment. We then presented two extracts (Extracts 4 and 5) to illustrate the trajectory of development of vocabulary knowledge (i.e. how students develop their conceptual understanding of ‘excuse me’).

We didn’t have a sub-section which specifically focuses on the types of gestures and the process of gestural mediation because we integrated the analysis of the nature of vocabulary explanations under case 1 and case 2. We hope that such restructuring will be more coherent in the sense that it reads like a narrative so that the readers do not need to go back and forth to refer to the extracts.

Reviewer 1 also suggested addressing these three aspects in the discussion section: “what types of gestures are observed? What do different gestures aim to mediate? How is mediation

discursively accomplished?” In addressing the first and second questions, we have provided an explanation on page 31:

“In this paper we identified two interactional features of VE sequences in a beginning-level adult ESOL classroom: embodied explanations (Extract 1) and embodied enactments as explanations (Extract 3). Our analyses showed that T draws heavily on gestural resources in explaining VIs; a finding which closely aligns with Lazaraton (2004) and Smotrova and Lantolf (2013) where gestural resources were shown to allow teachers to visualise the contextual meaning of the VIs thus making the VE more comprehensible for learners who have limited L2 repertoire. The analysis of embodied explanations in Extract 1 demonstrated **how T’s use of deictic and metaphoric gestures helped visualise the adverbs for expressing time.** Furthermore, the analysis of T’s embodied enactments in Extract 3 showed that **gestural resources (e.g. physical actions) are not only used to offer visual illustrations of the meanings of VIs but that they can be employed to create hypothetical contexts** that facilitate learners’ deeper understanding of VIs; a finding which is in line with Author and Brandt (2018). In representing the action of bumping into S13 physically and verbally, **T circumvents the possibility of students misunderstanding the abstract explanations of ‘excuse me’** thus allowing students to observe how the phrase ‘excuse me’ can be employed in a real-life context. ”

We also explained the types of gestures used by the students in demonstrating their conceptual understandings and the aim of using these gestures on page 32:

“The analysis also illustrated how learners’ understandings of the meanings of the target VIs can be evidenced through their appropriations of T’s metaphoric gestures (Extract 2), embodied enactments (Extract 4), and their re-using of taught target phrases to describe a past event (Extract 5).

These findings support Matsumoto and Dobs’s (2017) claim and demonstrate that **the students’ appropriations of T’s metaphoric gestures and embodied enactments are employed to externalise their internal thinking processes**, particularly in instances where the linguistic mode of explanation is not as rich as the gestural one, thus rendering them visible to the teachers and peers for inspection. Moreover, S3’s description of a past event is **mediated by her use of various gestural and linguistic resources (e.g. deictic gestures, use of L1, and sound of pushing) in order to accurately narrate the incident to T.**”

Regarding the third question, we believed that the analysis

	<p>section serves the purpose of demonstrating how the mediation of knowledge development is displayed in the classroom interactions. As this is not a specific suggestion, we were unable to address it directly.</p>
<p>(7) The conclusion mainly focuses on pedagogical implications; there is scope to highlight the theoretical and intellectual contributions of the project by stating explicitly that conceptually and methodologically, the research has advanced the field: it makes a good attempt to understand how gestures are organised in teachers' instructions and how they mediate the emergence of vocabulary knowledge over a lengthy period of time, as a result of moving from the inter-mental plane (interaction through talking and gesturing) to the intra-mental plane (knowledge growth and appropriation). As I suggested above, the paper would be usefully retitled as 'the mediation and organisation of gestures in vocabulary teaching and learning in the ESOL classroom' or something along the lines.</p>	<p>We agree with the reviewer's comment and have included an explanation of our methodological contribution to the field on page 34:</p> <p>"Methodologically, this study has helped advance the field by demonstrating how the use of CA to conduct multimodal analysis can shed light on the role of gestures in the explanation of VIs. Second, the study shows how using CA to trace L2 development over time can help paint a richer picture of students' learning processes. It captures how gestures mediate the emergence of vocabulary knowledge over a lengthy period of time, as a result of moving from the inter-mental plane (i.e. interaction through using various gestural and linguistic resources) to the intra-mental place (i.e. knowledge growth and appropriation). CA allows us to depict how the achievement of conceptual understanding can be an embodied activity, indicating that gestural resources can work as a 'window' to understand the current state of a learner's knowledge in the learning process. This study emphasises the need to examine both verbal and gestural elements together when considering</p>

	students' developmental processes of L2 vocabulary knowledge.”
(8) In terms of writing, the ideas are in general well-articulated, which makes the paper a very interesting read. Great effort and well done overall!	We would like to thank the reviewer for their positive and constructive feedback.

Reviewer 2's Comments	Authors' Response to the Comments
I suggest that the authors include a page that explains the meaning of Conversation Analysis transcription symbols in order to accommodate a wider readership. Educators, and not just linguists, would be interested in this article.	We would like to thank Reviewer 2 for their helpful comments. We now have included the CA transcription conventions under the Appendix section on page 40 in order to accommodate a wider readership.
On page 9, reviewer 2 made a comment next to the sentence “teacher devoted a lot of her time in explaining unfamiliar VIs”. The reviewer highlighted “a lot” and stated that: “perhaps use a more specific description”.	<p>We acknowledge the reviewer's comment and we have changed that into the following:</p> <p>“the teacher devoted a significant amount of time in explaining unfamiliar VIs to her learners” (p.10).</p> <p>Since we didn't conduct timed analysis to count the exact amount of time that the teacher spent on explaining vocabulary items to her learners, thus we are unable to offer the specific number of hours in the methodology section. We therefore added a footnote under the 'Notes' section to justify this:</p> <p>“1. Conducting a timed analysis was beyond the scope of this</p>

	<p>paper and it is impossible to report exact percentages of time dedicated to VI explanations. This is based on our judgment from observations of various hours of data.” (p.35)</p>
--	---

For Peer Review