

Open Research Online

The Open University's repository of research publications and other research outputs

Gesture in multimodal language learner interaction via videoconferencing on mobile devices

Thesis

How to cite:

Lee, Helen (2020). Gesture in multimodal language learner interaction via videoconferencing on mobile devices. PhD thesis. The Open University.

For guidance on citations see [FAQs](#).

© 2020 Helen Lee

Version: Version of Record

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's [data policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

(Redacted Version)

**Gesture in multimodal language learner interaction via
videoconferencing on mobile devices**

Helen Lee

BA (Hons) (English Literature and Music, Hope University, Liverpool, UK)

MA (TESOL and ICT, University of Brighton, UK)

Submitted for the degree of Doctor of Philosophy in the field of Language Learning and Technology. Faculty of Wellbeing, Education and Language Studies, The Open University.

Abstract

This thesis focuses on how adult English language learners exploit and experience gesture while communicating with one another via mobile technologies. Mobiles create opportunities for multimodal language learning beyond the classroom (Kukulka-Hulme et al., 2017), however, modes such as gesture are mediated and transformed by technology in complex ways (Hampel & Stickler, 2012). In a small-scale qualitative study, learners from a range of nationalities who were studying on language programmes in the UK were connected in dyads via Skype videoconferencing (VC) in order to complete information gap tasks using tablets, 2-in-1 devices, and smartphones. These communicative tasks had been intentionally designed around a diversity of informal ‘settings’ (Benson, 2011) which included cafés, museums, and historical buildings. Following the tasks, participants took part in stimulated recall interviews in order to reflect on their multimodal forms of communication.

This exploratory, qualitative study examines gesture from a theoretical perspective which links the mode to spoken language (Kendon, 2004; McNeill, 1992; Norris, 2004) and positions gesture within the wider framework of the negotiation of meaning (Varonis & Gass, 1985). As the role of speech-associated gestures within language learning via technology has not been widely researched, an interdisciplinary methodology had to be designed to analyse the video recorded data from the learners’ tasks. This is based on transcription procedures from gesture-speech analysis (McNeill, 1992; McNeill & Duncan, 2000). As gesture in this study is understood as being closely aligned to speech, a multimodal unit of analysis was combined with the Varonis and Gass (1985) framework of the negotiation of meaning. The multimodal method allowed for the categorisation and analysis of gesture to investigate how learners may co-orchestrate the two modes in relationship to their deployment of mobile technologies from beyond the classroom. The participants were asked to reflect on their interactions from multimodal perspectives and interview data were triangulated with the task performances. Theoretical and pedagogical conclusions are drawn as to the manner in which learners exploit gesture as an integral part of the negotiation of meaning.

ACKNOWLEDGEMENTS

My sincerest thanks go out to those who made this thesis possible. I am eternally grateful to my supervisors Professors Regine Hampel and Agnes Kukulska-Hulme. This thesis would not have been possible without their inspiration, encouragement, advice and support. I am also very grateful to The Open University for their ongoing financial, administrative and academic support. My thanks in particular go to Anne Forward, Dr Tim Lewis and Dr Lindsay O'Dell. Thanks also to Paul Slater and Simon Wilkinson from the MA TESOL programme. A special mention should be given to my colleague and good friend Rachel Vallins and most of all to the participants for their time and co-operation. I would also like to thank my family and partner David for their support.

Table of Contents

Chapter 1	1
1.1 Introduction.....	1
1.2 Learning with mobiles from beyond the classroom.....	2
1.3 The impetus for the research.....	3
1.4 Why examine gesture?.....	4
1.5 The research context and focus.....	5
1.6 The outline of the thesis.....	7
Chapter 2 Literature review	10
2.1 Introduction.....	10
2.2 An overview of the history of the study of gesture.....	11
2.3 Gesture and speech relationships	13
2.3.1 Kendon's continuum.....	16
2.3.2 Gesture and speech as utterance	18
2.3.3 McNeill's theory of hand and mind.....	19
2.4 Gesture types.....	22
2.5 Aspects of SLA.....	27
2.6 Gesture: multimodal language teaching and learning.....	30
2.6.1 Incorporating gesture into the teaching of language.....	33
2.6.2 Gesture as communication and meaning	35
2.6.3 Gesture and cognition	40
2.6.4 Gesture and the learning environment	41
2.7 The negotiation of meaning and the Varonis and Gass framework.....	44
2.8 Research in CALL	51
2.8.1 Studies in audio and videoconferencing tools	52
2.9 Language learning with mobile devices	58
2.9.1 Mobiles and SLA	60
2.9.2 Language learning opportunities from beyond the classroom.....	62
2.9.3 Multimodal literacies	63
2.9.4 Context-sensitive paradigms of learning	65
2.10 Conclusion to the literature review	68
Chapter 3 Methodology	71
3.1 Introduction.....	71
3.2 Philosophical perspectives on qualitative research.....	72
3.3 The roles of the researcher: ethical consideration around use of mobiles ...	76
3.4 Multimodality as research methodology.....	80
3.4.1 Approach to the research	82
3.4.2 The coding of the multimodal negotiation of meaning.....	85
3.5 Introduction to the data collection and analysis.....	87
3.5.1 The pilot study	89

3.5.2	A collaborative task designed around learners' use of VC on mobiles	91
3.5.3	Other possible methodological approaches	93
3.6	The research study	96
3.6.1	Participants and settings.....	96
3.6.2	Introduction to data collection instruments and procedures	100
3.6.3	The VC sessions on mobiles	101
3.6.4	Stimulated recall.....	106
3.7	The data analysis procedures	108
3.7.1	Data management procedures.....	113
3.7.2	A methodological toolkit	114
3.7.3	Gesture types.....	115
3.7.4	The Gesture Phrase	117
3.7.5	A gesture-speech unit of analysis	120
3.8	The multimodal analysis of the videoconferencing tasks	122
3.8.1	The use of ELAN	122
3.8.2	Challenges to the existing framework of the negotiation of meaning	127
3.8.3	The final multimodal transcription phase	129
3.9	The analysis of the stimulated recall data	131
3.10	Triangulation of data sources	134
3.11	Choices regarding presentation of data	136
3.12	Conclusion to the methodology	140
Chapter 4	The Findings	141
4.1	Introduction.....	141
4.2	Dyad one: Fay and Andrea	144
4.3	Dyad two: Nadia and Simone	162
4.4	Dyad three: Paul and Lily	166
4.5	Dyad four: Angela and Bobbi.....	187
4.6	The content analysis findings from the interviews	200
4.6.1	Making meaning	200
4.6.2	Gesture and language use.....	201
4.6.3	Affordances and constraints of technology and setting	203
4.6.4	Intersubjectivity	204
4.6.5	Peer collaboration	205
4.7	Conclusion to presentation of findings	205
Chapter 5	Discussion.....	207
5.1	Research Question 1	207
5.2	Research Question 2	214
5.3	The role of gesture	217
Chapter 6	Conclusions	219
6.1	Introduction.....	219
6.2	The pedagogical implications	220

6.3	Methodological implications	227
6.4	Contribution of the study to research.....	232
6.5	Limitations of the study	234
6.6	Future research directions	237
	References	241
APPENDIX 1	Letter about the research and meeting	258
APPENDIX 2	Participant consent form	259
APPENDIX 3	Task design for Skype VC on mobile devices beyond the classroom 260	
APPENDIX 4	Interview guidance sheet.....	263
APPENDIX 5	Transcription key for gesture and speech via Skype VC.....	264

List of Tables

Table 1	Kendon's continuum.....	16
Table 2	The individual gesture phrases within the overall gesture phrase.....	21
Table 3	Gestures of interest within the present study.....	23
Table 4	An overview of the data collection and analysis procedures.....	88
Table 5	Participants and settings.....	97
Table 6	Use of technology employed in the first data collection phase.....	99
Table 7	Use of technology employed in the second data collection phase.....	100
Table 8	Possibilities for capturing gestures when learners use Skype VC.....	105
Table 9	The nature and scope of data collected.....	110
Table 10	A multimodal pedagogic framework based on gesture and speech.....	226
Table 11	Future research directions in multimodality and SLA via technology.....	239

List of Figures

Figure 1	The gesture phases in operation in relationship to their co-occurring language.	21
Figure 2	The coding categories for speech turns.....	47
Figure 3	Relationship & overlap within the 3 fields.....	59
Figure 4	Artefacts for task design via mobiles to stimulate the negotiation of meaning.....	92
Figure 5	The three movement phases of an online iconic gesture via a tablet.....	118
Figure 6	Andrea forms a deictic gesture with the device and camera within setting.....	120
Figure 7	A gesture-speech analytic unit for prosthetic deictic via Skype VC.....	121
Figure 8	A section of multimodal transcription from ELAN.....	125
Figure 9	A multimodal trigger for negotiation based on location.....	128
Figure 10	Example analysis of interview from Atlas.ti8.....	133
Figure 11	Visualising linkages between quotations and codes.....	134
Figure 12	The building of online rapport.....	145
Figure 13	This is the place!.....	146
Figure 14	An accurate prosthetic gesture formed with a tablet.....	151
Figure 15	Fay searches for vocabulary through gesture and speech.....	153
Figure 16	Fay gestures the animal's antlers.....	154
Figure 17	Andrea gesturally mirrors the concept of antlers.....	154
Figure 18	Fay gestures the head of the animal.....	156
Figure 19	Nadia and Simone assume imaginary roles.....	163
Figure 20	Lily practises pointing with the device before the task.....	167
Figure 21	Paul triggers negotiation via a sequence of gestures and speech.....	177
Figure 22	Paul's simultaneous orchestration of gesture and speech.....	179
Figure 23	Lily exploits a deictic gesture as an indicator: "the face?".....	180

Figure 24	A random image on Skype versus a deliberate prosthetic deictic gesture.....	181
Figure 25	A series of attempts to form an accurate prosthetic deictic gesture stroke.....	193
Figure 26	Angela attempts to form an accurate prosthetic gesture with a tablet device.	193
Figure 27	Learners exploit deictic gestures in different ways.....	195
Figure 28	Task-based cycles within MALL.....	224

List of Excerpts

Excerpt 1	Learners negotiate for meaning based on setting and artefacts	127
Excerpt 2	An example of a visual comprehension check.....	128
Excerpt 3	Negotiation of meaning on mobiles from stately home and garden.....	148
Excerpt 4	Interview	158
Excerpt 5	Negotiation of meaning on mobiles from café and museum.....	164
Excerpt 6	Negotiation of meaning on mobiles from two café locations.....	169
Excerpt 7	Interview	184
Excerpt 8	Interview	186
Excerpt 9	Negotiation of meaning on mobiles from restaurant and gallery	189
Excerpt 10	Interview	197

Glossary of abbreviated terms

Aud CMC	Audio Computer-Mediated Communication
BAAL	British Association of Applied Linguistics
BERA	British Educational Research Association
CA	Conversational Analysis
CALL	Computer-Assisted Language Learning
CMC	Computer-Mediated Communication
CLT	Communicative Language Teaching
CS	Communication Strategy
DELTA	Diploma in Teaching English to Speakers of Other Languages
EFL	English as a Foreign Language
GP	Growth Point
IATEFL	International Association of Teaching English as a Foreign Language
ICT	Information Communication Technologies
IH	Interaction Hypothesis
L1	First language
L2	Second language
LOCH	Language learning Outside the Classroom with Handhelds
MALL	Mobile-assisted language learning
ML	Mobile learning
M-learning	Mobile learning
NS	Native speaker
NNS	Non-native speaker
OCMC	Oral Computer-Mediated-Communication
SLA	Second Language Acquisition
SCMC	Synchronous Computer-Mediated Communication
TBLL	Task-based Language Learning
TEFL	Teaching English as a Foreign Language
TESOL	Teaching English to Speakers of Other Languages
UG	Universal Grammar
VC	Videoconferencing
Vid CMC	Video Computer-Mediated Communication
ZPD	Zone of Proximal Development

Chapter 1

1.1 Introduction

The evolution in communication technologies has ensured that language teachers and learners now have access to a wide array of ubiquitous tools which range from language learning apps to video-based platforms. This has led to new frontiers of language learning but also raises issues regarding pedagogical direction in terms of how best to support learners' multimodal forms of communication in ways which enable them to acquire a second language in the twenty-first century. The field has also opened up new avenues for research. Kress and Van Leeuwen (2001) define multimodality as the exploitation of modes in the design of "a semiotic product or event" with modes combined "to reinforce each other [...], fulfil complementary roles [...] or be hierarchically ordered" (p. 20). It has been argued for some time now that communication plays a key role in language learning with concepts such as interaction and multimodality used to describe processes within second language acquisition (SLA) enabled by technology use (Hampel & Stickler, 2012). Both face to face and online learning exchanges encompass forms of expression which are not restricted to language but which entail learners' co-orchestration of modes in simultaneity. The deployment of modes may therefore include language but also the exploitation of image, gesture, gaze, proxemics; and even interaction with the everyday objects which surround us (Norris, 2004). As a result, modes have been defined as a set of cultural resources which are organised by principles and continuously shaped by the social interaction of people (Jewitt, 2011).

The modes of interest within the present study are those of gesture and speech. The rationale behind the chosen multimodal focus is that gesture has been shown to represent an important yet neglected resource in the case of second language learning and teaching. Perceptions of second language (L2) learners' use of gesture often include the notion that gesture acts as a compensatory mode in that its deployment is restricted to action-based mimes which are designed to bypass language. Conversely, whilst L2 learners do exploit gestures in order to deal with a range of communicative problems, their gestures do not inevitably replace speech but frequently accompany it (Gullberg, 1998). Moreover, gestures which are combined with speech (gesticulations) have been found to be exploited across numerous different languages and cultures studied (McNeill, 1992). The interconnections between the modes of gesture and speech have been previously examined

within the field of SLA in scenarios which encompass experimental settings (Gullberg, 1998); face-to-face classrooms (McCafferty & Stam, 2008); and under naturalistic conditions (McCafferty & Ahmed, 2000). However, there has been less attention paid to the manner in which online language learners exploit gesture and speech in relationship to processes within SLA when they deploy mobile devices to interact across a range of different settings from beyond the classroom.

Skype videoconferencing (VC) potentially affords a valuable medium to enable language learners to practise skills such as speaking and listening within a wider variety of context-rich settings than is currently possible within the physical classroom. For example, effective language teachers, and the latest coursebooks, continually attempt to replicate versions of the real world (that is, the environment outside the classroom where the second language is spoken routinely) in ways which are designed to situate and authenticate spoken dialogues according to communicative teaching and learning paradigms. In addition, teachers spend considerable time and effort in order to bring stimulating affordances from the world into their institutions via the use of 'realia' (that is material objects brought into the classroom). This approach is designed to ideally contextualise and elicit language use based around learners' interaction with everyday objects and authentic communicative settings.

1.2 Learning with mobiles from beyond the classroom

The exploitation of mobile technologies in language institutions, as an affordance to enable and support collaborative speaking tasks, is frequently overlooked. Perceptions based around learners' use of mobiles may entail the argument that devices create unwanted distraction, and, as a result, have little to do with the serious business of language learning. There still remains a disconnect between the types of multimodal communication which learners experience as they move through the social spaces of their everyday lives and the daily pedagogic practice of institutions and educators. However, an increasing interest in language learning with mobiles in relationship to English language teaching practices was evidenced by fifteen presentations on the subject at the IATEFL (International Association for Teaching of English as a Foreign Language) conference in 2015. Language teachers working within a wide diversity of contexts today are seemingly afforded rich opportunities to harness mobile devices as a means to support their learners to achieve a range of communicative goals beyond the classroom.

Whilst Skype was launched in 2003, its use across mobile devices represents a relatively recent social phenomenon which has remained underexplored within the field of technology-mediated language teaching and learning. This communicative tool is freely accessible on a diversity of smaller devices which could encompass its use on tablets, 2-in-1 devices (features of a tablet and laptop), and smartphones. Mobile devices and the use of Skype can be accessed ‘on the move,’ with the potential for teachers to enable groups of learners to interconnect virtually as a way to exploit an array of everyday settings within task-based paradigms. Furthermore, implementing tasks from communicative settings situated beyond the language classroom may more accurately reflect language learners’ existing ‘life-worlds’ (Pachler et al., 2010) in that their learning experiences and activities are situated within transport hubs, homes, cafés, restaurants, outdoor spaces, and places of historical and cultural interest. These types of settings should be recognised and harnessed by teachers as legitimate sites for learning with reconsideration as to how forms of strategic task design and theories within SLA might be exploited to support interactions from a wider variety of social spaces. As Benson (2011) remarks, ‘settings’ beyond the language classroom represent more than mere geographical location: they offer “an arrangement for learning” in specific places with learners engaging in “social or pedagogical relationships” and interacting with “material or virtual resources” (p. 13). Therefore, settings which lie beyond the language classroom would appear to offer an array of potential affordances which could include interaction with authentic artefacts and use of technology as teachers encourage learners to exploit these in an autonomous manner. Van Lier (2004) debates the issue of ‘affordance’ within the context of language learning and defines the concept in terms of what is immediately available for a person to do something with. From an ecological and semiotic perspective on language use, he notes that affordances for learners include their use of prosodic features, gesture, and facial expression but also their engagement with cultural artefacts.

1.3 The impetus for the research

In a previous Master’s degree in TESOL (Teaching English to Speakers of Other Languages) and ICT (Information Communication Technology), I first exploited the theory and framework of the negotiation of meaning (Varonis & Gass, 1985) (also known as the Varonis and Gass model) to elucidate how three cross-cultural learner dyads interacted via a series of speaking tasks across audioconferencing and videoconferencing tools. In this earlier study, the learners had accessed Skype VC and audioconferencing on desktops and

tablets from their institution. As a researcher, I implemented the Varonis and Gass (1985) coding model which acted as a framework with which to code and examine their interactive language use. In terms of my previous MA study, I gradually realised that my detailed transcription of the learners' use of language in isolation had completely failed to capture the multimodal manner in which they had actually negotiated for meaning. My data demonstrated that the dyads had exploited rich combinations of gesture and speech, to a greater or lesser degree, during the negotiation of meaning, however, at the time I had not understood what they had been aiming to achieve.

Following this earlier MA research study, I became intrigued as to what role gestures had played for the dyads and began to consider the subject in greater depth; and started asking a number of questions which laid the ground for the present study:

- How had the learners' gestures related to their language use?
- How had their gestures, in combination with their use of technology, enabled or constrained their meaning-based communication within the negotiation of meaning?
- In what specific manner could learners' gestures be pedagogically harnessed to support processes considered potentially beneficial to language acquisition and to raise awareness of multimodal forms of communication?
- How had the meanings which learners conveyed via gesture related to their interaction with the wider world of meanings around them?

1.4 Why examine gesture?

According to Kendon (2004), gesture is defined as "a label for actions that have the features of manifest deliberate expressiveness" which must be recognised by interlocutors as communicatively relevant and based on aspects of context (pp. 15-16). Kendon's definition is useful in that it separates learners' seemingly conscious and meaning-based hand movements from their wider embodied behaviours or incidental movements which lack communicative intent. The specific term 'gesticulation' is employed by McNeill (1992) to describe speech-associated gestures formed with the hands and arms; and to distinguish them from those gestures which occur without speech.

Unlike gestures which may replace speech such as mimetic pantomimes and emblems, gesticulations are unconventionalised in form and occur more spontaneously within people's unfolding dialogues. McNeill (1992) explains that if you watch someone speaking

in almost any language, in the majority of situations, you will witness the compulsion to move the hands and arms in conjunction with speech. It has been theorised that the modes of gesture and speech are therefore closely interconnected in that they are planned and processed together and can simultaneously fulfil communicative and cognitive roles for speakers (Gullberg, 2010; McCafferty & Stam, 2008; McNeill, 1992, 2000).

Gullberg (2010) notes that gestures have received surprisingly little attention within the paradigm of SLA. However, whilst still emerging as a field, there is an increasing body of knowledge which is based on research around the role of gesture as part of SLA within face-to-face teaching and learning contexts. Findings indicate considerable benefits for language learners which demonstrate that gestures, in conjunction with speech, act as an L2 communication strategy (Gullberg, 1998); support forms of self-regulation during tasks (Platt & Brooks, 2008); contribute to the comprehension of language for listeners (Dahl & Ludvigsen, 2014); and can improve the recall of vocabulary items (Allen, 1995).

In consideration as to the role of gesture today, there has been recent discussion based on the complex relationships which exist between the human body and the mediating effects of digital technologies (Chanier & Lamy, 2017; Farr et al., 2012; Hampel & Stickler, 2012; Jones, 2015; Kern, 2014; Lamy & Flewitt, 2012; Satar, 2015). These observations raise further issues regarding current research objectives, for example, they indicate a perceived need to shed light as to how mobile technologies may mediate and potentially transform the mode of gesture for language learners. Smith (2005) discusses SLA noting that technology-mediated communication is not a single homogenous category but instead demonstrates different facets in terms of features such as temporality, modality, and spatiality (p. 35). For example, gesturing across mobile devices via Skype VC may encompass features such as the mobility of the learner, the portability of their device; and video-based exchanges which involve people's simultaneous engagement with hybrid combinations of physical and virtual worlds.

1.5 The research context and focus

During my time working as a language teacher, I came to realise the importance of fostering speaking skills within a range of interactive scenarios. The requirement for language teachers to create an array of communicative opportunities for learners is widely recognised within private language schools and university curricula in the UK. For instance, it has long been acknowledged that, whilst considered important, the passive

knowledge of grammatical rules and structures is insufficient in itself to develop key skills such as communicative competence (Canale & Swain, 1980; Hymes, 1972). As a consequence, communicative language teaching (CLT) and task-based language learning (TBLL) remain dominant approaches within TEFL (see Ellis, 2003; Skehan, 1998). TBLL also features extensively in the literature on teaching and learning with technology; and this is increasingly associated with the creation of peer-based opportunities for collaboration and communicative interaction (Hampel, 2003; Hampel & Stickler, 2012; Kukulska-Hulme et al., 2017; Lee et al., 2019; Pegrum, 2014; Thomas & Reinders, 2012).

Communicative forms of teaching and learning consistently highlight the manner in which language is employed in the ‘real world’ (Howatt, 2004), with the argument posited that conversational interaction is an essential condition for second language acquisition to occur (Hatch, 1978). The SLA theory of the negotiation of meaning (Long, 1996; Pica, 1994; Varonis & Gass, 1985) suggests that the ongoing interactional modifications which occur when interlocutors experience difficulty in understanding messages are valuable for the acquisition of language. However, as Pegrum (2014) suggests within MALL, social approaches to learning can also be compatible with interactionist ideas as communicative teaching and learning paradigms require extension in new directions in relationship to learners’ deployment of mobile technologies (p. 93). The problem remains that the theory of the negotiation of meaning (Varonis & Gass, 1985) is defined as a mono-modal theoretical concept and framework which fails to recognise that “[l]anguage use is fundamentally multimodal” (Seyfeddinipur & Gullberg, 2014, p. 1). Furthermore, the theory fails to adequately address the critical importance of context in terms of how this might impact the manner in which learners negotiate for meaning.

This exploratory, qualitative study aimed to elucidate the role of gesture in conjunction with spoken turns within the negotiation of meaning when achieved on mobile devices from settings beyond the classroom. The research design entailed that learners would be completing their speaking tasks from a diversity of informal locations. The cross-cultural dyads taking part in the current study were actively encouraged to explore and share an array of settings as they were interconnected from geographically-separated locations through Skype VC accessed on mobile devices. The informal settings were consciously located in proximity to their language institution. The artefact-based affordances designed to support language learners’ interactions were already situated in these settings and were pre-identified and addressed through a context-sensitive approach to task design by myself

as the researcher. The task design was exploratory but also underpinned by the theory of the negotiation of meaning and enabled through the concept of the information gap task (see Pica et al., 1993). This approach implied that learners would ideally be prompted by interlocutors to ask for and respond to missing sources of information in order to negotiate for meaning around a series of material objects which were situated in the world around them. In this particular research context, it was also considered that learners might gesture with their hands or point with an embedded camera and mobile device as a way to possibly share an aspect of the world around them.

This research study is primarily concerned with elucidating the roles and relationships evidenced within learners' co-orchestration of gesture and speech, in terms of their potential modal interplay, whilst taking account of the influence of technology and material setting on this. As a result, I adopt the stance that gesture is an integral part of the discourse which takes place in the world and that learners may bring together gesture and speech as distinct but related meaning-making resources.

1.6 The outline of the thesis

As a result of the previously discussed areas of interest in the introduction to this chapter, chapter 2 constitutes a literature review which begins with an overview of the history of gesture studies. This is because the scope of this exploratory study is acknowledged as interdisciplinary in that it draws on theories and methods from wider fields than language learning; due to the fact that these have had a significant influence on the study of L2 gesture within classrooms and other learning contexts (see McCafferty & Stam, 2008 for an overview). Following the section on the history of gesture studies, there is discussion as to the potential interrelationships which exist between gesture and speech with some established theories presented. Next, there is a presentation of gesture types which have been of interest to previous researchers working in the field of gesture studies, including the field of SLA, and which pertain to the current study. In the subsequent section, the area of SLA is introduced more generally with a brief overview of some established theories. This is followed by a discussion as to how previous findings have demonstrated that gesture plays a role in multimodal forms of language teaching and learning from both communicative and cognitive perspectives. The subsequent section outlines the theory and framework of the negotiation of meaning and draws on the literature from both face-to-face and computer-assisted language learning (CALL) contexts, highlighting gaps. Finally,

principles and findings from the area of MALL are presented and discussed in relationship to the specific focus of the present study.

In chapter 3, relationships between the chosen philosophical and theoretical frameworks behind this research are explained. I outline the task-based rationale, the chosen methodological framework, the research design; and the procedure behind the implementation of spoken tasks beyond the classroom via mobile technologies. This section is followed by a detailed description of the participants and the research procedures. I introduce some theoretical perspectives on research methods used to examine people's multimodal forms of communication with an inclusion of methods designed to analyse gesture in relationship to speech. Following this, I debate the advantages and limitations of the chosen data collection instruments. Next, I explain my own methodological procedure for multimodal data collection and analysis, including the electronic means and qualitative research tools which better enabled this. In the subsequent section, the framework of negotiation of meaning (Varonis & Gass, 1985) is reconsidered within an extended multimodal paradigm. Finally, I present the analytic unit designed to analyse gesture-speech relationships (McNeill, 1992; McNeill & Duncan, 2000) and explain how this was integrated with the original negotiation of meaning framework as a way to examine gestures via mobile devices from settings beyond the classroom.

In chapter 4, I present the multimodal transcripts of excerpts from the tasks and a fine-grained, interpretative analysis of these. The negotiation of meaning is initially presented through a series of microgenetic multimodal transcripts with extensive discussion of gesture types, the possible interplay of gesture-speech combinations for learners, how gestures may help or hinder learners; and elucidation of the influence of the technology and material setting on language learners' multimodal negotiation of meaning from beyond the classroom. The chapter illustrates the learners' orchestration of gesture and speech from the researcher's interpretative stance whilst then presenting these same excerpts in relationship to a learner-centred perspective. The interviews are initially presented through verbatim excerpts from the stimulated recall interviews with the learners. Following this, the interview data is subjected to further examination to establish broader themes via a content analysis approach. As a consequence, several sources of data collection methods and levels of analysis are drawn on in order to attempt to address the research objectives from different vantage points and to understand where these may converge but also diverge.

In chapter 5, I revisit the research questions in the light of the findings and discuss these. In chapter, 6, I draw conclusions surrounding the pedagogical and methodological implications of studying language learners' gestures in relationship to their speech within the negotiation of meaning across mobile devices. In this final chapter, I present a multimodal pedagogic framework; designed for teachers who may be interested in supporting language learners to better understand their multimodal communication. I suggest a series of potentially valuable areas to focus on, based on raising awareness of the interplay between gesture and speech, within online videoconferencing environments. I also posit ways to potentially broaden the scope of research within the field of SLA through adopting a perspective which conceptualises the processes involved as multimodal. Negotiation is viewed through a lens which examines gesture and spoken language operating together whilst taking account of the influence of technology, social interaction; and material setting on these modes. Finally, suggestions are made as to directions for further research and how the field might evolve in the future.

The literature review is covered in the following section. It encompasses the scope of theories, studies and findings which forged a path towards enabling the methods and tools required in order to conduct the current research.

Chapter 2 Literature review

2.1 Introduction

In this chapter, I first provide a brief overview of the history of the study of gesture. Following this, I discuss several theoretical models which suggest that gesture and speech are closely linked. In the subsequent section, some established theories of language acquisition are presented and discussed (though this is in no way comprehensive given the innumerable theories in this field and the restrictions of length). Next, the role of gesture in SLA is critiqued with a focus on how gesture has been shown to contribute to different aspects involved in the multimodal teaching and learning of a language in face-to-face contexts. In the next section, the theory and framework of the negotiation of meaning (Varonis & Gass, 1985) is outlined. This is followed by a critique of some relevant studies in CALL which have previously exploited this methodological framework in the context of the use of desktop computers in order to draw a range of conclusions about learners' interactions. Finally, the area of MALL is discussed in relationship to SLA and the enabling of opportunities to learn beyond the classroom in ways which emphasise the importance of collaborative and contextualised approaches to L2 communication.

Gesture has represented a focus of interest within such diverse scholarly fields as rhetorical speaking, anthropology, aphasiology, neuroscience, performance art, behavioural psychology, child studies, ethnic studies, multimodality, interactional linguistics, SLA within teaching and learning contexts, and artificial intelligence. It is perhaps relevant to note that the concept of the orchestration of multimodal resources within human communication, and speech in particular, is not a new phenomenon, and did not begin with the revolution in communication technologies. It was therefore considered impossible to carry out a detailed study on gesture within the negotiation of meaning in a technology-mediated environment without first understanding something of the history, rationale, and taxonomies behind the study of gesture and speech within face-to-face contexts. Whilst previous studies cannot indicate information about the mediating impact of mobile technologies on gesture itself, as they have not investigated SLA via learners' use of these devices, they are important in that they encompass theoretical arguments which explicate perspectives on the interrelationship between gesture and speech in ways which significantly influenced the design of the present study. As a result, they act as a valuable starting point and guide for the researcher or teacher interested in the field of the study of

gesture within SLA. In the following section, there is an overview of the field of gesture studies in relationship to the current study.

2.2 An overview of the history of the study of gesture

Scholars have contested whether language evolved from an earlier system of gesture, developed before language, or whether there was a parallel development of gesture and vocal language (Corballis, 2002; Kendon, 2011). Vygotsky (1978) links gesture to children's symbolic play, for example, they have been found to exploit gesture in order to indicate the meaning of objects as playthings. Goldin-Meadow (2014) concluded that children's gestures can work alongside language, affording an additional representational format within learning processes. This use of multimodal resources within communication appears to begin at a very early age. Lancaster (2012) explains how by the age of two, children have already been 'grappling' with an entire world of semiosis, for instance, they will have experienced a variety of different ways of "displaying and representing meanings and information" (p. 326). Children have been found to use deictic language and gesture to "instantiate meaning" as they try to explain to others what they are attempting to express via their markings, whilst simultaneously pointing and talking: "that's x" (p. 316).

In terms of adult learners, the use of gesture was debated by scholars as early as Roman times, for example, a teacher of rhetoric named Quintilian (writing in the first century AD) cited gesture as a powerful communicative resource within his *Institutio oratoria* (cited in Kendon, 2004). He depicts gesture (*gestus*) in relationship to its co-occurring discourse (*vox*) and, as a result, theorised that the hands operate in conjunction with wider aspects of multimodal interaction such as facial expression and a person's use of the head. However, the hands are awarded special attention due to their capacity for expressivity within the performance of the individual speaker.

As for the hands, without which all action [i.e. Delivery] would be crippled and enfeebled, it is scarcely possible to describe the variety of their motions, since they are almost as expressive as words. For other portions of the body merely help the speaker, whereas the hands may be almost said to speak (Book XI, III. 85-87, cited in Kendon, 2004, p. 18).

One of the earliest attempts to classify gestures in relationship to their semiotic meaning occurred through the work of William Wundt. Wundt (1921/ 1973) was concerned with how specific gestural movements could be linked to meaning: gestures are deemed to form

a language in their own right and, as a result, he does not discuss gesture in relationship to the mode of speech. Efron (1941/1972) moved the focus of gesture studies towards achieving greater understanding as to gesture's relationship to speech and he regarded use of the hands as an integral component of communication. For example, he studied how people from different cultures assimilated their gestures in ways which resembled those of the native inhabitants of their adopted land. His findings demonstrate that it is not just language which can be acquired, for example, it was shown that Italian participants reduced the size and frequency of their gestures and that Eastern Europeans began to produce wider-type gestures when both cultural groups had spent time in the US.

Eckman and Friesen (1969) categorise gestures which accompany speech as illustrators in ways which suggest a co-expressivity of meaning operating between these two modes. They concluded that gestures were directly linked to speech as they reflected and supported what a speaker was simultaneously expressing in their language use. During this period, the analytic focus among researchers also shifted towards increasingly concrete definitions of gesture which went beyond *kinesics* (Birdwhistell, 1970). As a result, movements around 'self-grooming' or inadvertent behaviours remain absent from later taxonomies of gesture such as Kendon (1972, 2004) and McNeill (1992, 2000).

Contemporary studies in the field of multimodal and embodied forms of communication demonstrate how hand gestures emerge as part of talk-in-interaction and, as a result, they contribute to frameworks of co-operative action and the establishment of mutual understanding. This aspect of gesture has been explored through the deployment of methods such as conversation analysis (CA) (see Goodwin, 2000, 2003, 2007, 2014; Mondada, 2014; Streeck et al., 2011). It becomes clear from these studies that gestures cannot easily be disconnected from examples of their unfolding language use nor from the distinct contexts in which the gestures are shaped by the participants themselves. Goodwin (2014) argues that communicative acts are always situated in that they reflect the environments in which they take place. As a consequence, gestures will operate in direct relationship to the participants' "perspectives upon the material, real-world setting in which they interact" (p. 201).

LeBaron and Streeck (2000) argue that gestures should not be exclusively studied in relationship to language use in that the hands are regularly engaged in practical forms of action: "[h]ands are entangled in the world that they reach: touching objects, grasping tools, wielding instruments, managing matter" (p. 120). As a result, gestures are situated

within a wider ecology, consisting of material artefacts, which influences the manner in which people's meanings are formed through the hands within social situations.

Whilst not directly linked to speech, there have also been recent innovative re-conceptualizations of the mode of gesture in relationship to participants' engagement with forms of technology which are exploited out in the world or "in the wild" (Brown, 2010). For example, Jaworksi and Thurlow (2011) examine '*Gesture and movement in tourist spaces*' as they argue that the construction of deictic pointing gestures is not restricted to the human hand and finger but can be "embodied technologically – or prosthetically" through people's use of cameras and video cameras (p. 260). This notion implies that the act of pointing can also be potentially executed and replicated in new ways through language learners' engagement with a range of digital technologies. As a result, the use of technology ensures that people have access to an entire repertoire of representational and communicational modes in ways which prompt a reconsideration of what it means to learn (Jewitt, 2011). For example, there have been arguments that the new media age has resulted in a shift from 'telling' someone about the world to 'showing' them instead (Kress, 2003, p. 140). Kress is concerned with the multimodal nature of reading in the new media age, however, in the case of gesturing with digital communication technologies, it is possible that learners may be enabled to simultaneously talk as they show a particular aspect of the world to an interlocutor.

In the following section, established and comprehensive theories around the study of gesture and speech are outlined. These models are relevant to the current research study and thesis because they are regularly taken account of and deployed in the field of gesture studies within SLA.

2.3 Gesture and speech relationships

It is important to note that there have been several theories and models proposed around the degree of closeness which may exist between gesture and speech. In consideration of the potential relationships forged between gesture and speech, studies on aphasic speakers demonstrate how speech and gesture can both break down (McNeill, 1992) but that gesturing can equally support individuals to communicate more successfully in relationship to this debilitating condition (Goodwin, 2003). Whilst it is difficult to know which of the following cited theories can most accurately explain the extent of the relationship which exists between the modes of gesture and speech, they indicate indisputable

interconnections which are frequently ignored within the examination of SLA across teaching and learning scenarios. For example, gestures have been previously dismissed as ‘paralinguistic’ and, as a consequence, considered beyond the focus of areas such as applied linguistics (Kendon, 2004). Yet it has been argued that “gesture, speech and language are increasingly seen as linked in production, comprehension and development” in that they form an integrated system which operates together (Gullberg, 2010, p. 78). Gestures have thus been shown to support language learners in a variety of important ways within a paradigm of SLA. These include their development of skills such as speaking, comprehension, vocabulary learning, repair work, and forms of intrapersonal problem-solving (McCafferty & Stam, 2008). The study of gesture with speech may also afford a resource for researchers and teachers which could enable them to examine how individuals and groups of language learners construct, negotiate, and perceive meanings from multimodal perspectives in relationship to existing theories within SLA. In order to explain why the present study is concerned with the elucidation of gesture-speech relationships for language learners, it is first necessary to discuss some of the work within scientific approaches which posit that there is an interface, or degree of connection, which exists between these modes.

Proponents of the *Speech-precedes-gesture* model (see Butterworth & Hadar, 1989) assign speech a more dominant role than gesture; arguing that speech is the principal mode of expression with gesture integrated in a more limited way by speakers. Butterworth and Hadar (1989) put forward the *Lexical retrieval hypothesis* which suggests that gesture can support speakers to locate ‘sought-after lexical items’ via ‘holding’ the conceptual properties in their mind while their lexical search continues (cited in McCafferty, 2004, p. 151). The *Information packaging hypothesis* (Kita, 2000) suggests that language and gesture support forms of thought in multimodal ways. This model posits that speech and gesture are two independent processes which collaborate in ways where “gesturing helps the speaker organize information in a way suitable for linguistic expression” (p. 180). The *Sketch model* (see De Ruiter, 2000) assumes that speech and gesture are driven by communicative intent and that they come together to form a coherent multimodal utterance. Kendon (2004) is also principally concerned with the role of gesture as a multimodal communicative resource as he explains that speakers orchestrate the *gesture-speech ensemble* as they both plan and process the gestural and the verbal component together (pp. 134-135). Finally, McNeill (1992, 2000) adopts a Vygotskian perspective on

gesture, arguing that speech and gesture originate from the same underlying mental process, as he puts forward the theory that the components of thought, speech, and gesture dialectically intertwine and influence one another. He views gestures from dual perspectives in that they are communicative in nature but simultaneously afford ‘a window’ onto the speaker’s thought processes. His theoretical construct of the *Growth point* (GP) is a widely accepted theory within the field of gesture studies within SLA (Dahl & Ludwigsen, 2014; McCafferty, 2004; Negueruela & Lantolf, 2008) and has been previously exploited to examine the role of gesture within L2 comprehension (Dahl & Ludwigsen, 2014); the construction of meaning within L2 spoken narratives (Negueruela & Lantolf, 2008); and as a way to attempt to understand the ongoing interrelationships between gesture and speech as they operate within L2 conversational interaction (McCafferty, 2002).

Özyrüek (2014) comprehensively argues for a multimodal perspective on language use and illustrates that there are overlaps in the areas of the brain involved in the semantic processing of both co-speech gestures (gesticulations) and speech when used in isolation. He notes that gestures form an integrated part of spoken communication in terms of meaning:

Spoken languages are traditionally characterized as auditory-vocal languages [...] [T]hey are essentially multimodal in nature and also exploit the visual-gestural modality for communicative purposes [...] They are processed semantically during comprehension and as an integrated part of the speaker’s communicative message.

(p. 8)

To examine and elucidate how gesture and speech may form an integrated system, Kelly et al. (2010) conducted a laboratory study where listeners were prompted to assimilate sources of information gleaned from both gesture and speech into an integrated and holistic representation of semantic meaning. It was found that if the gestures used contained congruent information, for example, speech ‘chop’ and a gesture indicating ‘chop,’ then the meanings were perceived faster and more accurately than if the gestures contained examples of incongruent information to the speech channel. From the results, it was concluded that gesture and speech form an integrated system in the comprehension of language. Moreover, it has been shown that words or phrases are memorized more effectively if participants perform the accompanying gesture as they learn; rather than

simply hearing or reading isolated words (Engelkamp & Krumnacker, 1980). Another method used to examine the role of gesture in relationship to speech involved an examination of participants' ability to retrieve and recall fifty target words from their definitions via placing participants into groupings which either enabled or restricted hand gestures (Frick-Hornbury & Guttentag, 1998). Participants whose hand movements were restricted were found to retrieve fewer words than those who were freely able to exploit gestures: “[t]his finding suggests that gestures may do more than facilitate message comprehension; gesture production may serve an important function for the speaker as well as for the listener” (p. 53). Since the 1970s, Paivio has argued for a theory of dual coding where the brain itself is perceived as operating from multimodal perspectives. The theory considers that aspects of learning and memory are enabled via combinations of verbal and non-verbal (visual) codes which operate in order for people to form critical interconnections, resulting in two mental representations of a particular concept (Paivio, 2014).

2.3.1 Kendon’s continuum

To investigate the relationships between gesture and speech, it is necessary to comprehend the different components of gesture in terms of their synchronicity with speech (Kendon, 1980; McNeill, 1992, 2000). To illustrate the various degrees of relationship to speech, gestures can be placed on a continuum in terms of how ‘language-like’ and conventionalised they appear. Kendon’s continuum (an ordering of gestures first described by Kendon and further developed and dedicated in honour of him by McNeill) is next illustrated in Table 1.

Table 1: (Kendon’s continuum, after McNeill, 2000, p. 2)

Gesticulation →	Emblems →	Pantomime →	Sign language
Obligatory presence of speech	Optional presence of speech	Obligatory absence of speech	Obligatory absence of speech

The arrows in Table 1 illustrate the gradual progression of the different categories of gesture with *gesticulation* and *sign language* positioned at the extreme ends of the continuum. At the *sign language* end of the continuum in Table 1, gestures fulfil the role of becoming the language itself with no requirement for speech; as used and understood by deaf communities. These are undoubtedly languages in their own right and, as a result,

they demonstrate complex morphological and syntactic features (Stokoe, 1980). At the other extreme of the continuum, there are the *gesticulations/speech-associated* gestures (Kendon, 2004; McNeill, 1992) which form the focus of the current study. These types of gesture have been of particular interest to researchers working within teaching and learning contexts within SLA (McCafferty & Stam, 2008). This is perhaps due to the increasing prevalence of communicative, rather than behaviourist, approaches to language learning where gestures are seen to act as a multimodal resource within L2 spoken language tasks and instructional dialogues between teachers and learners. *Gesticulations* are spontaneous and, as a consequence, they do not demonstrate the same conventions or rules of gesture as demonstrated within the use of *emblems*, *pantomime (mime)* or *sign language* (see continuum). An *emblem* gesture represents a fixed, rather than a spontaneously formed, type of sign and is constructed, for example, when a speaker brings their thumb and forefinger together. In doing this, the speaker is constructing a sign which is immediately recognisable and which functions with or without the simultaneous requirement for speech. A ubiquitous example of a quotable *emblem* gesture is the ‘OK’ sign in that speakers all around the world will probably recognise it; despite the gesture provoking both positive and negative cultural responses. *Emblem* and *pantomime* gestures can also be unique to specific cultures, for example, Neapolitans have been found to exploit an entire gestural vocabulary which enables them to engage in conversations across distances and within noisy environments, without the requirement to speak (Kendon, 2004, p. 353).

Gesticulations in L1 (first language) have been shown to occur across every language studied within McNeill’s gesture laboratory which was based at *The University of Chicago* (African languages, English, French, German, Japanese, Korean, Mandarin, Russian, Spanish, Turkish). In this experimental setting, gestures were elicited via the deployment of a narrative speaking task based on a simple cartoon stimulus. It was found that there were similarities across languages regarding the deployment of *gesticulations* when used by speakers in order to depict the same event (see McNeill, 2006).

However, one distinct disadvantage of the exploitation of a narrative task within L2 research on gesture, is that gesture is controlled and constrained by forms of experimental design which cannot account for the complexity of interactional dynamics normally evidenced within conversations based around an L2 collaborative speaking task. Gullberg (1998) also argues that speakers are not generally aware of their *gesticulations* and, as a result, they may have no memory of their shape or occurrence within spontaneous forms of

spoken language use (p. 38). From this perspective, *gesticulations* may be constructed with deliberate intent at the time of speaking, however, both L1 and L2 speakers are highly likely to forget them. For this reason, they would require forms of visual stimuli in order to revisit both the form and meaning of the gesture and to remember its corresponding language use.

2.3.2 Gesture and speech as utterance

Kendon (1972) initially came to his conclusions about the nature of the relationship between gesture and speech through his research in anthropology. On the other hand, he is also pre-occupied with aspects of language acquisition and notes a lack of interest in the study of gesture within the field of linguistics. He blames the ‘chomskyan revolution’ and its limited focus on competence, rather than language use (seen as ‘performance’), as detrimental to allowing something as chaotic as gesture to become the focus of serious study (2004, pp. 64-83). Kendon’s (2004) definition of gesture is firmly rooted in the shared speech act in that gestures are situated within social interaction and are considered to form part of an utterance only when treated by those co-present as a “move, turn or contribution” (p. 7). To illustrate his point, Kendon multimodally analysed a series of episodes which took place between interlocutors who were engaged in everyday conversations. He describes the deployment of the *gesture-speech ensemble* as two components which are then orchestrated in relation to one another through the control which the speaker assumes (p. 127). Interestingly, he found that speakers modify and repeat their gestures, as well as their spoken language, as they attempt to clarify sources of information in order to establish mutual comprehension: “the gestured component, like the verbal component, is guided by the output the speaker is aiming at” (p. 128). Kendon’s analysis of conversations, occurring between native speakers, could equally be interpreted as a series of multimodal examples of the negotiation of meaning (Long, 1996; Pica, 1994; Varonis & Gass, 1985) where speakers attempt to clarify sources of information in order to establish mutual understanding between them. Forms of conversational clarification and modification are considered particularly important for language learners because, through these ongoing interactional adjustments, there can be perceived benefits to their acquisition of a second language (Long, 1996; Varonis & Gass, 1985). Conversely, the previously mentioned interactionist theorists, make no reference to the role which gesture might play within conversational adjustments in terms of its contribution to the negotiation of meaning and support for the acquisition of a second language.

2.3.3 McNeill's theory of hand and mind

Gestures have also been linked to speech through the conceptualization of gesture acting as 'a window into the mind' whilst simultaneously considering the hands act as a vital social and communicative resource: "both views must be taken if gestures are to be properly explained" (McNeill, 2000, p. 11). One of the most commonly accepted and influential theories within gesture studies is McNeill's theory of the GP (see McNeill, 1992, 2000). McNeill (1992) posits that speech and gesture form a holistic unit of meaning which must be analysed in its entirety. His approach to the analysis of speakers' meanings in their entirety is also reflected within multimodal approaches to research into SLA. Stam (2008) notes that "listeners see the gestures, hear the speech, and take the totality of the moment-to-moment communication into account in their interaction with speakers" (p. 253).

McNeill theorises that the hands act as an external manifestation of 'thinking-for-speaking' (for an overview see Slobin, 1996). Through the activity of speaking, thinking is manifested via our lived experiences which "are filtered through languages into verbalized events" (Negueruela et al., 2004, p. 117). Whilst not the focus of the present study, it has been noted that bi-lingual speakers have commented on how they engage in different forms of thinking when they shift languages and that there are differences in 'thinking-for-speaking' across different languages. These can be linked back to differences between languages in terms of whether they are satellite (English) or verb-based (Spanish) and in ways which can then go on to influence gesture (see McNeill & Duncan, 2000, pp. 141-152). Whilst there have been differences shown in the manner and amount of gesturing which various cultures engage in (Kendon, 2004; Seyfeddinipur & Gullberg, 2014), it has been demonstrated that gestural representations enhance learning because they are also capable of reflecting concepts "in the form of universal representations" in ways which are not directly tied to specific languages (Breckinridge Church et al., 2004, p. 303).

To more fully understand the origins of McNeill's theory of '*Hand and mind*' (1992), it is first necessary to reflect on Vygotsky's writings (1987). This work formed the basis of McNeill's GP theory and indicates gesture's role within the dialectical and unstable relationship which operates between language and thought. As Vygotsky remarks:

Speech does not merely serve as the expression of developed thought. Thought is restructured as it is transformed into speech. It is not expressed but completed in the word. Therefore, precisely because of the contrasting directions of the movement, the development of the internal and external aspects of speech form a true unity (1987, p. 251).

As McNeill depicts the role of gesture and language, he resonates with Vygotsky in that he suggests that people's utterances do not arrive 'fully-formed' but that gesture and speech arise together in unstable ways.

What does the growth point consist of? It is, theoretically, the utterance's primitive stage, the earliest form of the utterance in deep time, and the opening up of a microgenetic process that yields the surface utterance form as the final stage (McNeill, 1992, p. 220).

From McNeill's perspective, seemingly contrary properties of thinking, consisting of the visual/imagistic and linguistic elements of thought, dialectically integrate into an underlying unit which he denoted as the GP: "The growth point then should be a combination of image and word, of image and linguistic meaning" (1992, p. 220) as "two contrasting modes of structuring meaning co-exist in speech and gesture" (McNeill, 2000, p. 6). McNeill's theory of the ongoing, unstable interplay between the linguistic and gestural mode can be more fully understood through a recognition of key distinctions regarding the manner in which language and gesture are individually constructed as modes. It has been noted that language moves from smaller parts, which are meaningful in themselves, and which form "a standardized plan or syntax" (McNeill, 2000, p. 5). In other words, there is a predictable sequence of morphemes, words, and sentences which build in an upward manner in order to constitute meaning. Conversely, gesture is 'global' in meaning (McNeill, 1992) in that the individual parts of a gesture do not convey an independent meaning and can only be understood by deconstructing the whole (Norris, 2004, p. 28). Unlike language, gestures are also 'synthetic' in that a single gesticulation (gesture with accompanying speech) "concentrates into one symbolic form distinct meanings that can be spread across the entire surface of the accompanying sentence" (McNeill, 1992, p. 5-6).

In order to explain the aims, design, and findings of the present study and to discuss the findings on gesture and SLA in the subsequent section of this literature review, it is first

necessary to understand that gestures can be analysed in terms of the overall gestural phrase with an analytic distinction made between the various gesture types (see Kendon, 2004; McNeill, 1992, 2000; Norris, 2004). McNeill (1992, p. 83) depicts gestures as occurring within an overall gesture phrase which consists of a series of smaller movements known as the gesture phases. These are summarised in Table 2:

Table 2: The individual gesture phases within the overall gestural phrase

<i>Preparation phase</i>	The limb moves away from the resting position to a position in gesture space where the stroke begins.
<i>Stroke (Obligatory)</i>	The peak effort of the gesture where the meaning of the gesture is conveyed.
<i>Post-stroke hold</i>	The final position reached is held.
<i>Retraction</i>	The hand returns to its resting position.

The sequence which is shown next (Figure 1) is designed to support the reader to understand how the gestural phases operate in terms of the execution of a single gesture and its co-occurring language.

Figure 1: The gesture phases in operation in relationship to their co-occurring language (McNeill, 2006). See: http://mcneilllab.uchicago.edu/pdfs/cambridge_encyclopedia.pdf

The first image shown in Figure 1 (reading from top left to right) illustrates the hands of the speaker seated on the left assuming a resting position (also known as the pre-preparation position). In image two, there is a preparation phase which is shown in the transcription of the spoken language through a bracket (the onset of the gesture) and a transcription of the co-occurring words: [oak tree/ and he. The gesture reaches its content-based peak or expresses its meaning through the speaker's deployment of the stroke which is shown in the third image and is co-expressive with the relevant words: "**bends it way ba."** The post-stroke hold is shown in the final image which co-occurs with the final part of the word 'back.' In this final image, there is no representation of the hand returning to its resting position as, according to McNeill, the speaker then moves straight into the execution of another gesture. Underlined words indicate gestural hold phases.

2.4 Gesture types

In Table 3 (shown next), I illustrate the gesture types derived from McNeill (1992) and include an additional gesture type of particular interest in the current study as indicated in the previous multimodal research on technology and deictic gestures, evidenced in Jaworski and Thurlow (2011). Table 3 consciously excludes the type of gesture known as ‘beats.’ Beats are up-and-down or in-and-out movements which appear as if the speaker is beating musical time. These gesture types are not related to semantic meaning but are instead connected to aspects of discourse management. Due to the present study’s focus on meaning, they were excluded from the analysis. The first three images shown in Table 3 are taken from Kelly et al. (2011) with the final image derived from the present study.


Redacted Images	<p>ICONIC: “Possess a pictorial content, often mimicking what is conveyed verbally, describing specific objects or events, making them more vivid” (Norris, 2004, p. 28, based on McNeill, 1992).</p>
	<p>METAPHORIC: “Possess a pictorial content, however they present the invisible: an abstract idea or category” (Norris, 2004, p. 28, based on McNeill, 1992).</p>
	<p>DEICTIC WITH THE HANDS AND FINGERS: “Point to objects or people in the physical world or to abstract concepts as if they had a physical location” (Norris, 2004, p. 28, based on McNeill, 1992).</p>
	<p>PROSTHETIC DEICTIC GESTURE: Pointing is achieved “technologically or prosthetically by means of the camera” (Jaworski & Thurlow, 2011, p. 260).</p>

Table 3: Gestures of interest within the present study

It has been noted that iconic gestures represent their meanings ‘pictographically’ in that the form of the gesture is conceptually related to the semantic content of the speech which it accompanies (Krauss et al., 2000, p. 275). Iconic gestures have been deemed to be difficult to interpret without hearing the accompanying speech because they assume a clear relationship to the ideas which they convey within a narrative, for example, when a speaker exploits an iconic gesture as a way to illustrate how a character bends a tree back to the ground whilst uttering the words: “and he bends it way back” (see McNeill transcript shown previously in Figure 1). In the case of the deployment of metaphoric gestures, abstract concepts can be made more tangible for an interlocutor through the use of the speaker’s hands: “the gesture depicts a concrete metaphor for a concept, a visual and kinesic image that we feel is, in some fashion, similar to the concept” (McNeill, 1992, p. 14). Therefore, metaphoric gestures must also be considered to assume a degree of iconicity in that they allow the speaker to convey abstractions in more concrete and potentially understandable ways. An example of iconicity within a metaphoric gesture is illustrated when a social worker explains how a client’s words were literally “all coming out” as she simultaneously moves her hands outwards and upwards as a way to depict an image of a substance which is literally gushing out of herself (Kendon, 2004, p. 100). Unlike iconic or metaphoric gestures, deictic gestures instead convey their meaning in relationship to the entity which is being highlighted. In the use of iconic and metaphoric gestures, the hands themselves form a pictorial representation of a concrete object or convey an abstraction in concrete form. In contrast, deictics involve the ability to ‘indicate’ (Clark, 2003) and to accurately locate entities and actions in space according to a specific reference point (Bühler, 1982; Haviland, 2000).

To unravel even apparently simple ‘pointing’ gestures requires cognitive and sociocultural insight: about what entities exist to be pointed at, about how to reconstruct referents from indicated locations, about why an interactant points at all [...] [G]estures are fleeting but accessible cognitive exhibits, playing out with the body the actions, referential and otherwise, that constitute discourse (Haviland, 2000, p. 39).

As mentioned previously, it has been posited that the manner in which tourists have been shown to point with technology implies their use of video and cameras functioning as literal extensions of the human body (Jaworski and Thurlow, 2011). Pointing within technology-mediated contexts in real time may operate in ways which can only be

understood in relationship to social interaction and in terms of speakers' interaction with features which are embedded in the setting in which they are situated. For example, deictics, whether conveyed in gesture or language or both, serve as an attempt to draw attention to referents of interest and to establish joint attention within an actual geographical location. In technology-mediated forms of communication, the manner in which people convey information in order to establish a frame of reference may become disrupted when compared to the act of pointing within face-to-face contexts where people are normally situated within the same space. In discussion of the multimodal affordances of mobile devices, Kress and Pachler (2007) suggest that these devices represent 'a prosthesis of self' with digital technologies enabling content to be presented using an array of semiotic systems of representation which offer different ways of orchestrating meaning. From a multimodal perspective, "digital video allows learners to create representations of themselves and the way they see and interact with the world" (p. 13). As a consequence, learners' online engagement with themselves, one another, and with artefacts in the wider world may potentially transform our traditional notions of pointing as devices and cameras are deployed in order to possibly replace deictic gestures normally associated with the human fingers and hands.

In relationship to an interest in exploring the category of prosthetic deictic gesture within this study, I first considered that strategic forms of task design would prompt the learners taking part to engage in a series of pointing gestures which they would be required to achieve through their use of the device and camera. Clark (2003) refers to pointing as an indicative act where speakers enable addressees to focus on specific objects within their environment. It has been previously suggested that artificial technologies may be used as a substitute for the human finger and hand to achieve this (see Clark, 2003; Jaworski & Thurlow, 2011). I surmised that the language learners taking part in my own study would also need to be able to successfully direct their addressee's attention towards a specific object of interest; via indicating this physically and virtually through the deployment of a mobile device across Skype. From this perspective, I considered that deictic gesturing, achieved through the use of technology, would still be reliant on features which are normally associated with the construction of more-naturalistic deictic gesture types. For example, I considered that key features which are inherent in deictic gesturing include the following: the need to direct an addressee's attention towards an indicated referent in order to establish intersubjectivity; the requirement to position the human body in relationship to

the wider environment; and the speaker's ongoing awareness of the interdependencies which exist between the formation of a deictic gesture and the use of language.

Conversely, I considered that the act of pointing with a mobile device would somewhat disrupt language learners' construction of deictic gestures when compared to how these are achieved in face-to-face contexts. For example, gestures are ephemeral movements which are normally reliant on the 'close proximity' of participants (Hampel, 2019, p. 44). Having considered the principal features of deictic gesturing, whilst at the same time acknowledging the uniqueness of this particular technology-mediated context, I considered it necessary to draw on a separate category of deictic gesture in order to distinguish this from those formed with learners' use of the finger and hand in isolation. As a result, I refer to this particular gesture type as 'prosthetic deictic' (Jaworski & Thurlow, 2011).

This section of chapter 2 has shown that gestures can be considered to be closely related to speech as reflected in theories, classification systems, and research findings. Gestures can either constitute a language in itself, as in cases where speakers' use of these becomes conventionalised and shared between specific communities, or they can be more spontaneously formed in non-conventionalised ways through speakers' deployment of *gesticulations* within their emergent talk. The field of gesture studies is beginning to impact the field of SLA in that there has been discussion and research into how attention paid to this mode in relationship to speech may allow for an increased multimodal understanding as to how L2 gestures contribute to language learning processes in a variety of ways. Taxonomies of gesture importantly enable distinctions to be made between the various gesture types in terms of the establishment of their semiotic meaning in relationship to L2 speech. Furthermore, the ubiquitous use of communication technologies potentially challenges our conventionalised notion of gesture. This aspect of contemporary communication warrants further investigation due to previous work which suggests that pointing or deictic gestures can be achieved prosthetically via the use of devices, cameras, and video in ways which reflect people's interaction with the wider world as they construct meaning.

In the following section, the area of SLA is discussed in relationship to cited theories and perspectives. This is followed by a comprehensive review of studies in the field of gesture and speech within SLA.

2.5 Aspects of SLA

Language learners today acquire languages for a range of different reasons from passing exams, studying at university, forming online friendships, reading academic papers; and the pursuit of international employment and travel opportunities. Doughty and Long (2003) note that SLA when “naturalistic, instructed or both” is becoming crucial as second languages increase in importance (p. 4). They also highlight that SLA always takes place within a social context and can be impacted by this in “micro and macro ways” (p. 4). There have also been challenges to Krashen’s (1980) cited dichotomy between acquisition and language learning. He suggests that acquisition is an entirely naturalistic, unconscious process which is related to meaning, whereas, learning represents a conscious process which is the result of instruction and which has greater relevance to form. It has also been argued that learning processes can only be manipulated to a certain degree and that aspects of natural language acquisition must also be considered in foreign language instruction (Felix & Simmet, 1981). In the case of MALL, paradigms appear to become further blurred as language learners may be supported by their teachers to communicate in ‘real-life contexts’ as they interact with members of the public within settings such as supermarkets and the workplace (Ogata et al., 2006).

To position the study of gesture within SLA within the theory of the negotiation of meaning, it is perhaps useful to first offer a brief overview of some existing theories within the field of language acquisition. This is in no way comprehensive, given the considerable amount of diversity and theories proposed within this particular field. Chomsky (1965) proposes a nativist theory of acquisition which can be exclusively explained by biological factors. He argues that people possess an innate knowledge of language, known as universal grammar (UG), and postulates that this knowledge is illustrated in their ability to acquire language; despite oral input which is regularly inadequate and marred through false starts and lack of grammaticality. He uses a ‘poverty-of-the-stimulus’ rationale (see Berwick et al., 2013) to explain UG, arguing that children are exposed to insufficiently rich input in their care environments to account for their capacity to acquire all the features of a language. However, Lancaster (2012) demonstrates that very young children are also “actively acquiring parallel systems of representation” (p. 315). In the case of adult language learners, Skehan (1998) argues that UG poses problems for second language

acquisition and language teaching in that it is “good at describing a formal, underlying competence [...] but it is less convincing with second language learning, with real-time communication, and with the relationship between performance and change” (p. 80).

In his *Input Hypothesis*, Krashen (1985) proposes that the acquisition of a second language requires that learners are exposed to comprehensible input and that this needs to be one step above their current level: $i + 1$. Whilst comprehensible input is deemed to be important, Swain (1985) contests the idea that this is sufficient in itself and argues that learners additionally require comprehensible output or productive opportunities. Her study on Canadian immersion learners demonstrated that, despite seven years of exposure to the target language, language use was deemed less than successful due to the conclusion that there was little opportunity for “two-way negotiated exchanges in the classroom” (p.247). Schimdt (1990) proposes the ‘noticing hypothesis,’ indicating that nothing is really learned unless it is noticed by language learners. He contests that Krashen’s notion of subconscious acquisition versus conscious learning are of little help regarding actual language production and comprehension scenarios. Whilst these cited theories of SLA are useful, they fail to address the fact that L2 speakers do not necessarily deploy language in isolation. It has been shown that gesture and speech interact together in order to create more complex units of meaning which are formed by speakers in a holistic sense (Kendon, 2004, pp. 108-109) but which are rarely considered within theories of SLA.

Another important theoretical perspective in SLA is sociocultural theory: this stance offers a framework through which cognition can be analysed without detaching it from its wider social context (Lantolf & Thorne, 2006). Vygotsky (1978, 1987) depicts the mind as developed through our interactions with people as these are continually mediated by cultural artefacts which include the use of language but also the use of gesture. As previously outlined, McNeill (1992, 2000) draws on Vygotsky’s work in order to explain how a speaker’s utterance evolves from its earliest stage to ‘the opening up’ of a microgenetic process: “in keeping with Vygotsky’s conception of the minimal units of a microgenetic process, growth points must retain properties of the whole” (McNeill, 1992, p. 220).

Interactionist theories of acquisition continually emphasise the importance of enabling conversational opportunities in a second language (Hatch, 1978; Long, 1996; Mackey, 2007) in ways which are prevalent in communicative approaches to language teaching and learning today. However, studies within an interactionist tradition have also tended to

focus on language as an isolated, psycholinguistic phenomenon with little information on the social elements of communication, the context in which the communication occurs; and the multimodal dynamics of conversational interaction. Moreover, it is perhaps surprising that gesture has not received more attention within CLT and TBLT as enabling language learners to communicate effectively remains a central tenet of these types of pedagogic approaches within the TEFL classroom. Long (1996) notes that opportunities for conversation importantly allow for negotiation of meaning in that speakers are prompted to adapt and modify their language use in order to make themselves comprehensible to their interlocutors. On the other hand, his notion of ‘conversation’ is limited and immediately challenged if it is considered that language learners’ talk can be viewed as a multimodal phenomenon. Conversations today are also regularly mediated by the material aspects of technology itself in a manner which has transformed how we converse across an array of different platforms and devices (see Hampel, 2003; Hampel & Stickler, 2012; Jones, 2015; Kern, 2014; Kukulska-Hulme et al., 2017; Lamy & Flewitt, 2012; Lancaster, 2013; Satar, 2015).

Within Long’s interactionist hypothesis (1996), native to non-native conversations entail speakers adopting certain adaptations or strategies such as repetition, confirmation, reformulation, comprehension checks, and clarification requests. These modifications are considered beneficial to the acquisition of language as they are used “strategically, to avoid conversational trouble, and tactically to repair communication when this breaks down” (p. 418). Conversely, Varonis and Gass (1985) demonstrate that peer-to-peer interactions can also afford rich opportunities to negotiate for meaning in ways which involve learners checking, clarifying, and confirming various sources of information when incidents of communicative ‘non-understanding’ arise and when they attempt to resolve the communicative problem (p. 79). Learner-to-learner interactions are also more likely than learner-to-native speaker interactions to offer interlocutors an opportunity to incorporate feedback (Adams, 2007).

Foster and Ohta (2005) draw attention to the seemingly disparate approaches adopted by researchers within SLA as the authors choose to highlight commonality between psycholinguistic and sociocultural approaches when examining the negotiation of meaning: “both seek to understand data sets in which learners talk to one another in the language being learned. Both are interested in how it is that interaction promotes SLA” (p. 404). A focus on gesture within SLA involves examining the manner in which language learners

deploy gestures in order to better understand the ways in which their gesture shifts with their development of language. The approach can offer both social and cognitive insights into acquisition processes (Gullberg, 2006). Furthermore, the study of gesture allows for the researcher to examine features of acquisition such as how learners may deal with aspects of vocabulary or grammar within extended multimodal and contextually-rich paradigms:

We conclude by arguing that SLA involves the dynamic, interactive alignment of learners, teachers, and their ever-changing environments, suggesting that symbiotic gesture is but one example of humanly improvised resources guiding ecosocial alignment and participation in SLA (Churchill et al., 2010, p. 234).

In the following section, the role of gesture in SLA is discussed in relationship to multimodal forms of teaching and learning within face-to-face contexts as this is where the majority of research in the area of gesture studies and SLA has taken place to date.

2.6 Gesture: multimodal language teaching and learning

The idea that it is possible to exploit the body as a learning resource within second language learning can be seen in attempts from the 1960s onwards. An approach known as the total physical response (TPR), requires that students respond with actions to imperatives which are spoken by their teacher. Whilst multimodal, the approach is essentially behaviourist with the learner unable to assume any active role in the communication or expression of their own meaning. Wylie (1985) and Neu (1990) suggest that the body itself needs to be trained to communicate and should be taken into account when assessing what it really means to be a proficient communicator in a foreign language. Bachman (1990) put forward the communicative language ability model, which encompassed non-verbal behaviours, and was designed to be applied within L2 teaching and testing scenarios based on the assessment of features such as learners' social and strategic abilities. However, modes such as gesture are still largely ignored in the teaching and testing of communicative skills within language schools and other educational institutions.

In relationship to the specific exploitation of gestures in the language classroom, Carels (1981) proposed the deployment of pantomime gestures within the language learning process in relationship to the acquisition of vocabulary items. Within this multimodal approach, the teacher narrates a text and simultaneously gestures vocabulary items which

are unknown or challenging for learners to understand. Following this phase, learners then repeat the text and the same gestures in order to consolidate the acquisition of novel words (see also Macedonia & von Kriegstein, 2012). However, whilst theories of language acquisition have been debated for decades, there has been little information available for teachers as to how resources such as gesture might be usefully harnessed to support SLA processes within more learner-centred and communicative learning scenarios. As Belhiah (2013) notes, there has been a lack of research into gesture within qualitative or quantitative paradigms, however, several studies have examined gesture and they afford considerable evidence to suggest that it is “a legitimate component of L2 discourse and as such worth investigating” (p.417).

One of the most important elements to the study of gesture within L2, is to understand the ways in which this mode has relevance to aspects of language acquisition. There have been specific practical suggestions by Gullberg et al. (2008) as to how gesture might be studied in relationship to language use:

- 1) We can examine the role gestures play in interaction to mediate the acquisition of spoken language, their general role in communication, in providing the socio-cognitive pre-requisites for the development of language, in conveying and possibly entrenching meaning, and their connection to cognitive capacities such as working memory etc.
- 2) Gestures as a reflection of language development. We can further investigate the way in which gestures develop and change in parallel to spoken language development, and the ways in which they shed light on both the product and process of acquisition.
- 3) Gestures as language development itself. This approach studies the acquisition of gestures as an expressive system in its own right.

(p. 150)

In reference to suggestion one, research into the role of gesture in relationship to learners’ language use may shed light on processes within SLA from communicative and cognitive perspectives. Suggestion one implies that the conveyance of meaning is no longer restricted to a learner using or comprehending disembodied sentences but of them understanding the ongoing, holistic interplay which occurs between gesture and speech in the social construction of their meanings. Suggestion two, implies that the researcher, language teacher, and learner might usefully observe how gestures unfold and evolve in

relationship to the meanings which participants are attempting to express via their use of language within isolated learning tasks and over more extended periods of time. As a result, the study of L2 gestures may afford 'a window' onto learners' thought processes as they attempt to jointly solve problems or move through a range of linguistic options as they negotiate for meaning in multimodal ways. Suggestion three may reflect the notion that learners acquire certain gestures as they acquire a second language. McCafferty and Ahmed (2000) implemented a study to find out if language learners from Japan and America would appropriate and internalize metaphoric gestures via two different conditions: naturalistic (residing in the target country) and instruction only. Findings revealed that metaphoric gestures were appropriated by the naturalistic learners only and were deemed to represent a significant aspect of acculturation for them. Conversely, it has also been demonstrated that it is possible to teach certain aspects of culture, for example, through the presentation of idiomatic French expressions and their corresponding emblematic gestures (Allen, 1995).

The scope of studies within SLA has also broadened over time to include the fine-grained analysis of learners' and teachers' use of the resource of gesture, and other relevant modes such as gaze, under a range of different learning conditions (McCafferty & Stam, 2008). The detailed study of L2 gesture within face-to-face contexts has explored a number of diverse areas which include the following: vocabulary acquisition (Allen, 1995, Eskildsen & Wagner, 2015); repair work (Olsher, 2008; Mortensen, 2016); facilitation of comprehension (Dahl & Ludvigsen, 2014; Sueyoshi & Hardison, 2005); gesture as a communicative strategy (Bialystok, 1990; Gullberg, 1998; Kellerman, 1992; Olsher, 2008); gesture in communicative competence (Neu, 1990); gesture and grammar (Churchill et al., 2010); gesture and private speech (McCafferty, 1998; McCafferty and Ahmed, 2000; Lantolf & Thorne, 2006; Lee, 2008); gesture and meaning-making (Negueruela & Lantolf, 2008); gesture within problem-solving (McCafferty, 2002; Platt & Brooks, 2008); the use of gestures by teachers within instructional contexts (Allen, 1995; Belhiah, 2013; Faraco & Kida, 2008; Lazaraton, 2004; Sime, 2008); and the establishment of intersubjectivity through embodied forms of action (Mori & Hayashi, 2006).

2.6.1 Incorporating gesture into the teaching of language

One problem for language educators is that students, even when they appear to know the required vocabulary and grammar, inevitably battle to express meaning in communicative situations (Allen, 1995). Areas of focus on gesture within instructional contexts have included the analysis of gestures in relationship to teachers' vocabulary explanations (Lazaraton, 2004); definition talk with regard to use of gestures which involve emphasising the meaning of the verbal channel and disambiguating the meaning of specific lexical items (Belhiah, 2013); attributions which learners themselves make in relation to a review of a teacher's gestures (Sime, 2008); and the teaching of emblems in relationship to the learning of French expressions (Allen, 1995).

Allen (1995) examined the influence of emblematic gestures on the development and access of mental representations of a range of French expressions among one hundred and twelve university students studying at a French university. The premise of the study was that if fifty emblems were taught with simultaneously presented linguistic expressions, this would entail an increased depth of processing and a greater ability for learners to later recall the expressions. Firstly, results of an analysis of the post-test scores revealed that learning emblematic gestures with French expressions did lead to greater recall. Secondly, the groups that witnessed the emblematic gestures by the teacher forgot significantly fewer sentences than the group who did not have access to the gestures. Allen's conclusions have multimodal pedagogic implications as she emphasises how mental representations created through resources such as gesture can help learners to better understand how language works and enable them to internalize it: "educators must gain an understanding of mental representations and develop methodologies that will facilitate the development and access of the mental representations" (p. 527).

Sime (2008) adopted a more learner-centred approach to the examination of teachers' gestures in the EFL (English as a foreign language) classroom and found that, through a stimulated recall approach, learners were afforded an active role in the research process as they became active interpreters of the gestures. From this perspective, gestures were not examined in isolation but were viewed as a shared communicative and cognitive resource which learners could be encouraged to interpret and respond to as the conversational addressees within their instructional setting. The categories of gesture exploited in the

study were drawn from McNeill (1992) and encompassed use of iconics, metaphors, beats and deictics. Volunteer learners were asked to watch a video extract derived from their classroom lessons and asked to comment on the teacher's body language. The research also drew on the work of Gass and Mackey (2000) who had previously suggested that video can act as a reminder of an event and can be deployed to stimulate recall of the cognitive processes active at the time. It was found that the learners taking part in Sime's study considered that the teachers' gestures contributed to the interaction and conveyed meaning to them in relevant ways. From cognitive perspectives, gestures were found to contribute to comprehension and meaning of lexical items, supported learning processes in multimodal ways; and could demonstrate a teacher's reaction to the learner's original output via forms of feedback. A section from a learner interview within Sime (2008) depicts the value of the learner analysing a deictic gesture in relationship to their understanding of a linguistic concept:

"It makes concrete an abstract situation and concrete things are always clearer. Pointing to himself means he is one character in the example, and his hands move to Manuelle to show that he gives her his house, like a transaction."

(p. 265)

In contrast, Faraco and Kida (2008) found that gestures within the negotiation of meaning were not generally related to linguistic items. Conversely, it was also concluded that gestures could cause ambiguity for learners if a teacher co-ordinated their gestures in relationship to the learner's discourse, rather than within the portion of the teacher's corrective restatement (p. 294). It was found that gaze was the most important multimodal resource for participants as it afforded an indicator from the teacher as to how to manage the interaction. One of the limitations of this particular study, is that gesture is never clearly defined, for instance, the mode is sometimes confusingly illustrated through examples of gaze and wider embodied movements such as shoulder shrugs. Nevertheless, the study cites the potential value of examining gesture within the Varonis and Gass (1985) framework of the negotiation of meaning in a more learner-centred way. The study also notes that gaining an understanding of particular sets of social conditions are critical to address within the analysis of learning sequences such as the negotiation of meaning.

Belhiah (2013) highlights how the teacher's hand can choreograph their instruction within forms of definition talk within the physical classroom. Gestures can thus be used by the

teacher to underline meanings, to disambiguate vocabulary, and to establish forms of cohesion. As a result, gesture is not viewed as a compensatory strategy but as an integral part of co-present conversation. However, unlike the approach adopted in Sime's study, the learner and teacher taking part in this small-scale study assume clearly defined roles with the teacher 'teaching' and seen to be firmly in charge as he offers input but little opportunity for negotiated discussion with the learner. As Belhiah notes, "[o]ne may disagree with his pedagogical choice because it does not promote active learning nor engage the student in a negotiation of meaning" (p. 420). In drawing attention to the pedagogical choices inherent in his study on gesture, Belhiah makes a salient point in consideration of the rationale behind the value of studying gesture in SLA within more communicative and learner-centred approaches to pedagogy.

2.6.2 Gesture as communication and meaning

Several studies have explicated the role of gesture in relationship to aspects of L2 communication and language learning (Eskildsen & Wagner, 2015; Gullberg, 1998; McCafferty, 2002; Mori & Hayashi, 2006; Negueruela & Lantolf, 2008; Neu, 1990; Olsher, 2008). Findings illustrate that gestures can aid in the learning of new vocabulary (Eskildsen & Wagner, 2015); sustain communication and support listeners (Gullberg, 1998); contribute to people's perceptions of L2 proficiency (Neu, 1990); enable multimodal repair moves (Olsher, 2008); aid in the establishment of intersubjectivity (Mori & Hayashi, 2006); and contribute to L2 meaning-making (Negueruela & Lantolf, 2008). Studies have varied in their approach to the study of gesture within forms of communication, for example, L2 communication strategies have formed one important line of inquiry (Gullberg, 1998; Olsher, 2008). Gullberg (1998) explored gesture within communication strategies within experimental conditions whereas Olsher (2008) chose to video L2 elementary learners whilst they were engaged in interaction in a more spontaneous manner as they completed a task within the classroom. There have also been examples of the fine-grained analysis of learners' and native speakers' gestures within embodied forms of conversation situated within informal settings such as a local coffee shop (Mori & Hayashi, 2006).

Gullberg (1998) was interested in the function of gesture as a communication strategy for intermediate learners of French and Swedish. She notes that one of the principal features of learner language is the deployment of communicative strategies (CSs) which are exploited in order to resolve problems in real situations. Therefore, the study of gesture is

consciously positioned in relationship to existing theories of language learners' use of communication strategies (see Bialystok, 1990; Faerch and Kasper, 1983; Tarone, 1980) in ways which can contribute to our multimodal understanding of SLA. Conversely, Gullberg (1998) is critical of these existing frameworks in terms of how they approach gesture: separating its use from oral CSs, a lack of definition of gesture types, and no indication of their relationship to speech (p. 32). Within her exploratory study, seventeen participants' gestures were elicited via the vehicle of a narrative task which involved them recounting a cartoon (following McNeill, 1992). It was found that speakers produced more gestures in their L2 than L1 when they recounted the same task. Gestures in L2 were also not restricted to pantomime gestures and the use of this type of mimetic gesture was rare across the data. More common gestures observed across the data involved the learners' deployment of metaphoric and deictics, rather than iconic gestures. This is in contrast to findings which have shown that iconic gestures are the most prevalent gesture type within L2 interactions (McCafferty, 2002, 2004; Olsher, 2008).

Gullberg concluded that iconic and metaphoric gestures are exploited by learners in order to solve lexically-related problems with deictic gestures used to overcome discourse-related difficulties and to enable forms of co-reference and coherence within discourse. McCafferty (2002) found that an L1 speaker and an L2 speaker used deictic gestures to point to hypothetical but also concrete places within their spoken interactions. Participants in his study exploited aspects of their immediate surroundings, for example, they used buildings and objects as material resources to stimulate discussion beyond the classroom. Pointing acts as a multimodal resource which is designed to establish a shared interactional space as people draw attention to objects of interest within their environment (Mondada, 2014). Gullberg's study (1998) also encompassed the native speaker's assessment of the L2 learners' gestures and she concludes how the use of gestures within CSs appear to positively contribute to an overall evaluation of an L2 learner's proficiency.

Bachman (1990) has long argued that non-verbal behaviours should be introduced into the teaching and testing of L2 communication, however, this idea is not normally taken into account in consideration of an assessment of oral proficiency levels within language institutions today. Conversely, Neu (1990) assessed the role of non-verbal communication, including the use of gestures, within a wider framework of communicative competence (Canale & Swain, 1980). She used the vehicle of an L2 oral proficiency interview conducted by a Japanese and Saudi student in order to conclude that movements such as

facial expressions and gestures are key to creating the impression of a successful second language speaker. Taking non-verbal modes into account within the analysis resulted in a participant being evaluated as more communicatively competent and proficient than their counterpart, even when the level of verbal proficiency was deemed to be lower.

It could be considered that adults who are just embarking on their language learning journey may rely more heavily on gestures as they have not acquired enough language in order to fully express themselves. For example, it has been noted that language learners in the earlier stages of acquisition tend to gesturally reflect some of the same issues which they experience within their use of language (McCafferty & Stam, 2008). For instance, adult learners, at the earlier stages of acquisition, demonstrate a tendency to verbally repeat the full referent rather than to adopt the use of pronouns within their talk; and this phenomenon is also evidenced in their use of gesture which can also appear overly-explicit. L2 learners have been found to repeat gestures using the original hand position and same portion of gestural space, instead of indicating what they are referring back to in a more abbreviated fashion through, for example, pointing in a general direction (see Gullberg, 2003; McCafferty, 2004). However, Gullberg (1998) argues that the relationship between linguistic proficiency levels and a reliance on gesture remains somewhat unclear.

Olsher (2008) examined L2 elementary level learners as they were collaboratively engaged in designing and creating an annotated map. The task importantly allowed for the more interactional elements of gesture to be studied between three language learners as they conversed within a classroom space. Olsher argues that the spontaneous nature of the learners' talk can be likened to more naturalistic aspects of SLA. He draws on Tarone (1980), noting that CSs can be defined as interactional attempts by interlocutors to agree on meaning when meaning does not appear to be shared. Gestures have been previously shown to contribute to the creation of zones of proximal development (ZPD) in that second language learners create a 'shared history of signs' which crucially allows them to scaffold one another as they co-construct meaning in ways which enhance the communication and comprehension process (McCafferty, 2002).

Olsher's study was interested in finding out whether combinations of gesture and speech within an interactional context could facilitate acquisition in terms of learners' attention to input and multimodal forms of processing. Gestures enabled the learners to repeat aspects of language within the repair turn with the conclusion drawn that learners attend to input in multimodal ways. Mori and Hayashi (2006) found that gestures which were shared

between L1 speakers and L2 learners of Japanese could help interlocutors to establish critical levels of intersubjectivity within their conversations.

Drawing on Krashen and his *Comprehensible Input Hypothesis*, Gullberg et al. (2008) posit that gesture assumes a role in input from multimodal perspectives: “interlocutors are known to attend to and make use of gestural information” (p. 163). This idea is supported in Olsher’s findings in that the learners’ use of ‘gesturally-enhanced repeated repair turns’ demonstrated how they paid attention to input from both verbal and gestural perspectives when communicative issues arose within their real-time conversation. In terms of gesture type, the repair turns involved learners’ deployment of iconic gestures which were connected to their lexical affiliates in meaning:

[T]hese repair moves via repeat *plus embodied action* reflect an interactional choice to elaborate or specify aspects of the trouble-source turn through gestural displays. In this way, they can be seen to address problems of understanding as well as hearing (Olsher, 2008, p. 125).

Negueruela and Lantolf (2008) exploited McNeill’s gesture types and the theoretical construct of the GP (McNeill, 1992) in order to establish the role of gestures in meaning-making processes within L2 oral narratives by more advanced speakers. The gestures under investigation were iconic and deictic. Negueruela and Lantolf conclude that the principal difference between gestures of the L1 and L2 is that, within the former, gestures often carry meaning which is not expressed in the co-occurring spoken language. Conversely, in the L2 examples, the speakers produced the same meaning in their speech as in their gesture.

In the following transcript, the native speaker gesturally illustrates that a salad was brought on a tray, however, this information is notably absent within their use of the speech mode. In contrast, the L2 Spanish speaker combines their use of language with the use of gesture in ways which appear to indicate aspects of gestural redundancy in terms of meaning.

The difference in how gestures and speech are exploited by L1 and L2 speakers in their study is illustrated next:

L1 speaker: [he’s the guy who brought **the salad**]

(Iconic, co-expressive with arm bent at elbow extending upwards hand with palm up, fingers extended in horizontal position)

L2 speaker: Pero esta ensalada está [encima de una **bandeja**]

But this salad was on a tray

(Iconic, regulatory with arm bent at the elbow extended upwards; hand palm up with fingers extended in a regulatory fashion)

(Negueruela and Lantolf, 2008, p. 98)

The overall conclusions drawn in the previously cited study suggest that the exploitation of L2 gestures assume a self-regulatory function because the speaker uses gesture to support their thinking processes, rather than to add meaning-based information in a more communicative sense. Conversely, the findings are also contradictory in that there are examples of language learners adding meaning-based information via gesture which is not evidenced in their use of language. The findings also show that the L2 learners execute gestures in an asynchronous manner in order to engage in lexical searches, for example, a learner's gesture can sometimes emerge before the linguistic part of their utterance in ways which appear to support the learner to remember a relevant word. Negueruela and Lantolf suggest that within examples of lexical retrieval (lexical searches), the holistic meaning created between gesture and speech breaks down with the two modes seemingly operating in more independent ways than is made apparent within examples of McNeill's GP. Krauss et al. (2000) suggest that through the use of gesture "the input consists of features of the source concept represented in motoric or kinesic form" in ways which support the individual speaker (p. 269). However, there have also been findings which illustrate that an interlocutor will often 'join in' the lexical search having witnessed the relevant gesture from their interlocutor: the concept is made visually accessible to enable the interlocutor to try to discern its meaning (Breckinridge Church et al., 2004; Hayashi, 2003). Hayashi (2003) found that Japanese speakers' vocal and visual searches for specific words can be understood as a shared social practice which enables 'a window' into what are normally considered to be private cognitive processes.

McCafferty (2004) concludes that L2 participants make a conscious effort to present identical content in their use of gesture and speech within their construction of meaning. Drawing on McNeill and Duncan (2000), and their notion of co-expressivity within the GP, McCafferty suggests that gestures may assume a different role in L2 learning to L1 speech in that the gestural channel is designed to support or 'concretize' the verbal channel; rather than to represent the same underlying idea in a different way.

Gesture has also been shown to contribute in a positive manner to comprehension of a second language (Dahl & Ludwigsen, 2014; Gullberg et al., 2008; Kellerman, 1992; Lazaraton, 2004; Sueyoshi & Hardison, 2005). This includes factors such as support for learners via their access to multimodal forms of input (Gullberg et al., 2008); the importance of taking account of kinesic behaviour and the perceived limitations in using audio-only materials with L2 learners (Kellerman, 1992); and improved levels of comprehension when gestures are present within communication (Dahl & Ludwigsen, 2014). Dahl and Ludwigsen's (2014) study aimed to understand the role of gesture in terms of aspects of language learning such as recall and comprehension. They also considered the learner perspective within their research and implemented a Visual Cue Preference Questionnaire (VCPQ) (see Sueyoshi & Hardison, 2005). This was in order to examine learners' own perceived beliefs with regard to the value they placed on access to visual resources such as gesture in their understanding of English. Learners preferred developing their listening skills via access to gesture-rich forms of media such as TV or movies and appreciated opportunities to practise speaking skills through talking, which was viewed as 'a gesture-rich activity.' Conversely, they preferred to learn vocabulary from reading texts and did not perceive gesture as of value as an aid in the comprehension of language. This is in contrast to Sime (2008) who found that participants paid attention to a teacher's gestures when they were perceived as supportive of their comprehension of specific language items.

2.6.3 Gesture and cognition

The Russian psychologist, Vygotsky (1896-1934), assumed that the development of cognition and language arose as the result of social interaction. In sociocultural theory, speaking is said to mediate thinking in that individuals assume control of the mind via the process of internalization. In the field of gesture in SLA, McCafferty (2008) has argued that learners' internalization begins with their embodied experience around object-related activity in the wider world and that gesture provides insights regarding developmental processes which are involved in SLA (p. 48). Lantolf (2000) posits that opportunities for learners' spoken collaborations are critical for their language acquisition as he notes that internalization suggests that the source of consciousness is rooted in social activity (p. 13). He also considers that speakers' gestures play a role in these processes and positions the mode in direct relationship to the analysis of L2 speech. Lantolf (2000) claims that McNeill's work (1992) has been crucial in drawing our attention to the role of gesture within communicative activity with others but also as a way to understand key aspects of

communication with ourselves (p. 16). Gesture has been shown to support forms of private speech (Lee, 2008; McCafferty, 1998; McCafferty & Ahmed, 2000); enhance L2 comprehension (Dahl & Ludwigsen, 2014; Kellerman, 1992; Lazaraton, 2004); help learners to memorise words (Kelly et al., 2009; Macedonia & von Kriegstein 2012); and to problem-solve (McCafferty, 2004; Platt & Brooks, 2008).

There have been notably few examples within the literature as to how learners exploit gestures within their private study, rather than via the process of communication with others. However, Lee (2008) illustrated how gestures could contribute to the scaffolding of a reader's interactive understanding of an L2 text in relationship to their use of private speech. Findings showed that learners pointed at materials and represented abstract concepts with their hands in order to essentially teach themselves as they supported their own learning in autonomous ways. It was concluded that these movements enabled the learners to both study and learn language. McCafferty (1998) elucidates how L2 speakers (Japanese and Spanish) synchronise their gestures to co-occur with their use of inner speech, for example, they use gesture as a self-regulation strategy when they experience difficulty in completing learning tasks.

Platt and Brooks (2008) drew on microgenetic analysis as a method to capture developmental processes as they occurred (Vygotsky, 1978; Wertsch, 1991) within the unfolding of a collaborative L2 task. Through this approach, they demonstrated how learners exploited combinations of hands and the mind in order to support them to speak. It was found that the use of embodied resources could support cognition and promote forms of self-regulation in important ways. It has been suggested that gestures help learners because L2 speakers are pressed to deal with problems inherent in thinking through one language, their L1, but in speaking via a different one, their L2 (Lantolf & Thorne, 2006; McCafferty & Stam, 2008). McCafferty (2004) argues that because learners do not have full control over discourse in their L2, they use their hands to support their cognition; and that gestures can enable speech as they contribute to the structuring of the verbal channel.

2.6.4 Gesture and the learning environment

Kita (2000) theorises that gestures assume a portion of the physical environment which acts in close combination with cognition. She explains that this *spatio-motoric* form of thinking helps speakers to organise information within action-based schemas. This is the type of thinking which occurs when people have to organise information as they interact

with affordances within their environment. Mori and Hayashi (2006) examined interactions from within a local coffee shop in the US as L1 (English) and L2 speakers (Japanese) orchestrated vocal and gesture/non-vocal resources as a way to attempt to achieve intersubjectivity. Affordances present within this more informal learning environment included the participants' spontaneous use of material artefacts, such as a random page of advertising found within a local newspaper, which happened to be lying on the table. For example, a native speaker drew his interlocutor's attention to this page and the material artefact of the newspaper itself then led to a complex discussion about the price of cars within the US and Japan. Mori and Hayashi (2006) note that L1 participants' use of gesture critically allows them to convey particular ideas in a 'global/synthetic' manner (McNeill, 1992) which then helps the L2 interlocutors to gain an approximate understanding of an idea or concept; and supports them in their ability to then effectively respond. Mori and Hayashi (2006) refer to these incidents within their data as 'embodied completions.' It was concluded that these exchanges are especially relevant to SLA in that they allow the L2 speaker to examine connections between linguistic forms and concepts expressed via forms of embodied and situated interaction.

McCafferty (2002) discusses gesture and speech in relationship to interlocutors' interaction with both their immediate setting and hypothetical surroundings. The two participants in this study regularly referenced buildings and objects via the use of deictic pointing gestures which were exploited in order to co-construct meaning. This was achieved through combinations of gesture, language; and the speakers' exploitation of available affordances inherent in everyday settings. The interlocutors formed their deictic gestures via use of their index finger and thumb which led the addressee to pay close attention to the direction which the speaker was indicating in relationship to their co-occurring language use. Deictic gestures were also harnessed to represent aspects of the world, for example, spaces such as the ground were transformed into a carpet in a house and an armchair was exploited to virtually reference an entire city. One of the most interesting conclusions in this study is the idea that gestures between an L1 and L2 speaker can transform in meaning across a more extended series of interactions. For example, a gesture which was originally used to signify the plan of the university campus was later deployed to represent the meaning of 'everything' for the two speakers.

Whilst there was no designated task within the previous examples of interaction, Olsher (2008) notes the importance of the implementation of L2 tasks which are designed to

facilitate forms of multimodal learning via the interaction of gestures and language use for learners. He notes that, from pedagogical perspectives, collaborative tasks which involve learners' interaction with material objects are deemed to be conducive to their deployment of L2 gestures within incidences of repair; suggesting that learners' interaction with material objects creates a rich learning environment which allows for opportunities to deploy gestures in relationship to semantic meaning. There are also particular advantages to the implementation of pair work or small group work tasks which actively promote the use of vocabulary in connection to the use of images or material objects within a task-based learning environment.

This section has presented and critiqued some of the literature in SLA in relationship to the study of gesture and speech in face-to-face contexts. Perhaps due to the more extended history of face-to-face classroom research, this has resulted in a more comprehensive and structured analysis of gesture in relationship to L2 speech than has been evidenced in the majority of technology-mediated teaching and learning contexts to date. The literature on the role of gesture within instructional learning contexts indicates potential benefits to learners if teachers introduce gesture into their explanations of language. Learners themselves have been found to deploy gestures to convey meaning and to communicate within their spoken interactions in ways which can support their communication and cognition across many different aspects of language learning. There are also perceived differences between the manner in which L1 and L2 speakers may bring gesture and speech together with the suggestion that language learners exploit gesture as a way to support their use of the verbal channel. It is also not entirely clear from the literature as to whether L2 learners will exploit gesture in ways which involve the conveyance of meanings, not made apparent in their co-occurring language use, or whether they will demonstrate examples of gestural redundancy. Gestures have been shown to help learners to comprehend language and to disambiguate problems in their understanding of meaning-based concepts. Learners have been found to engage with CSs and repair moves in multimodal ways which are not normally taken into account within theories of SLA. Learners also use gesture to support processes such as private speech in a manner which may support their cognition and help them to internalize language. The role of the learning environment has also been shown to play a role in the shaping of gestures in terms of how speakers harness aspects of their surroundings as a way to support their communicative processes. From this perspective, gesture and speech do not occur within a vacuum but

involve multimodal processes within SLA which occur within, and in relationship to, the meaningful worlds in which the learners interact. In the following section, the framework of the negotiation of meaning is illustrated.

2.7 The negotiation of meaning and the Varonis and Gass framework

From the 1980s onwards SLA researchers have examined the impact of conversational interaction on learners' development of language. Hatch (1978) originally suggested that language learning takes place during communication and is the result of learners understanding how to engage with conversations. Long (1996), in his updated version of the interaction hypothesis (IH), remarked that opportunities to interact can support language development in a number of different ways through aspects of collaboration. Within the IH it is argued that learners are exposed to comprehensible input and forms of feedback (Long, 1996; Pica, 1994; Varonis & Gass, 1985), and, in combination with opportunities to adjust their output (Swain, 1985), they are ideally supported to learn. Features of collaborative, negotiated interaction include the notion that "through exposure to positive evidence and negative feedback which contains positive evidence, incorrect associations are weakened" (Long, 1996, p. 430). Learners are also afforded opportunities to test out various hypotheses through forms of clarification and confirmation of information (Pica, 1994). Therefore, the premise behind the negotiation of meaning is that interlocutors provide valuable forms of feedback which can prompt L2 speakers to make the necessary adjustments or modifications to their language use when their input is deemed to be difficult to understand.

Foster (1998) explains that the creation of opportunities for learners within tasks "to clarify problem utterances (negotiating for meaning) enables them to receive comprehensible input and to produce comprehensible output, both of which have been claimed to be critical to SLA" (p. 1). Foster and Ohta (2005) adopted a simultaneously psycholinguistic and sociocultural stance in order to conduct their investigation into the negotiation of meaning within a face-to-face classroom setting. They found that there were very few instances of negotiation within their data but that learners actively assisted one another towards task completion via forms of co-operation and prompting. However, they note how learners cannot be separated from their environments and that knowledge construction becomes an integral part of the 'setting' itself, with language development seen to occur in relationship to the learners' interaction with people and artefacts.

Varonis and Gass (1985) originally defined negotiation of meaning as conversational episodes which serve one of two functions: (1) negotiation of non-understanding and/or (2) continuation of the conversation. In this sense, the learner may or may not notice the opportunity to negotiate for meaning or may instead decide that the conversation should continue, regardless of any perceived communicative problem which may have arisen between interlocutors. For example, whilst Block (2002) acknowledges that negotiation takes place within learner communication, he notes that interlocutors may be simultaneously engaged in acts of identity affirmation and forms of face-saving during communication (p. 124). Varonis and Gass demonstrate that native speakers equally pretend to understand aspects of language, rather than to admit their non-understanding to an interlocutor. An example of this can be seen in a native speaker who lets an opportunity pass rather than avail themselves of an opportunity to clarify their non-understanding of the word “monolithic” (see Varonis & Gass, 1985, p. 74).

Varonis and Gass (1985) define a negotiation routine as the following:

Non-understanding routines are operationally defined as: *those exchanges in which there is some overt indication that understanding between participants has not been complete* [...] These can stem from an exchange in which there was a misunderstanding, no understanding, or incomplete understanding (p. 73).

The above definition of ‘non-understanding’ is broad in nature and implies that it is not necessary for interlocutors to experience a complete breakdown in their communication in order to generate opportunities which may allow them to modify and adapt their talk as they attempt to establish mutual understanding. For example, Varonis and Gass (1985) confirm that: “[t]he *sine qua non* of a non-understanding routine is that within the exchange there are embeddings of one or more clarifications” (p. 73).

Varonis and Gass (1985) based their recommendations for language learning principles on their data which demonstrated that NNS—NNS pairings spent more time negotiating than the NS—NNS pairs; and they also indicate that the peer exchanges showed more work around the resolution of communicative difficulties before the conversation was then allowed to return to a forward trajectory (p. 83). Varonis and Gass (1985) were interested in comparing the interactional patterns observed within dyadic interactions which took place between native—non-native speakers versus non-native—non-native speakers. Their study consisted of fourteen conversational dyads of non-native speakers; four conversational dyads consisting of a native and non-native speaker; and four

conversational dyads consisting of native speakers (p. 72). It is implied that the most productive way to form dyads in terms of the creation of negotiation opportunities is if interlocutors do not share an L1 in common. This is also in line with more contemporary teaching and learning scenarios in that private language schools and universities in the UK frequently consist of mixed nationality classes who are learning English. However, there are also numerous possibilities for teachers working abroad with learners of the same L1 as they can harness technology in order to connect speakers from different L1 backgrounds from all over the world.

However, Varonis and Gass (1985) too readily attribute their overall findings to the notion that there is some acknowledged 'incompetence' which disrupts the conversational flow between L2 speakers (p. 73). In contrast to their depiction of the non-native speaker, it is highly questionable whether language learners, who demonstrate both the need and the skill required to negotiate for meaning, can ever be seen to represent incompetent speakers. This is a derogatory and undermining label which the researchers use to depict the normal difficulties which most adult language learners will experience in their attempts to acquire a second language later in life. From a more contemporary pedagogic perspective, adults who choose to express their non-understanding represent motivated learners who understand the knowledge-sharing opportunities made available through opportunities for conversational interaction. Furthermore, they represent 'risk takers' who are, importantly, unafraid to draw attention to problematic aspects of communication in order to progress their language learning. Nevertheless, Varonis and Gass (1985) highlight some important pedagogic points as they describe the 'equal footing' which exists between non-native speakers in ways that may consequently impact a range of affective issues for them. For instance, many adult learners do feel embarrassed and undermined when native speakers or teachers interrupt them and 'error correct' in an overt manner; rather than engaging the learner in forms of negotiated discussion about language.

Through the Varonis and Gass (1985) framework, it is possible to break a conversation down into a series of coded speech turns which fall within the boundaries of a designated 'routine.' This operates as shown in Figure 2:

T → I → R → RR

Figure 2: The coding categories for speech turns within the negotiation of meaning from Varonis and Gass (1985)

The negotiation can be divided into two broad sections which consist of a trigger phase: (T) which is followed by a resolution phase. The resolution phase is made up of an initiation (I), resolution (R), and an optional reaction to resolution (RR). The trigger forms the section of the discourse which prompts the initial non-understanding for the listener. During the resolution phase, the portion of the discourse which originally triggered the non-understanding is signified by the same listener via an indicator move (I). This then leads to a response (R) from the speaker in recognition of the previous indication of non-understanding with the optional reaction to response (RR) signifying the end of the routine. The model also suggests that speakers may attempt to further clarify sources of information through comprehension checks (see also Long, 1996) which may occur at any point within the conversation.

To illustrate how the coding scheme operates, Varonis and Gass (1985) provide numerous examples from their data. Triggers can assume a variety of forms for speakers which include asking questions, stating information or attempting to rephrase, reduce or elaborate on information.

Examples of triggers:

A: What is your name? TRIGGER
B: My name?
A: Yeah
A: yeah. How long...will you be? will you be staying?
B: I will four months TRIGGER
A: four months?
B: stay four months here until April

(Varonis and Gass, 1985, p. 75)

Indicators are the signal that an episode or 'routine' of non-understanding has been triggered. This can be achieved through an echo of the trigger, an explicit statement of non-understanding, a non-verbal response or an inappropriate response. The Varonis and Gass framework reduces definitions of non-verbal communication to features such as silence. There is no recognition of the possible role of gesture or other embodied cues within their version of the negotiation of meaning.

Examples of indicators:

A: What is your name?
B: My name? INDICATOR
A: Are you a student in your country?
B: In my class?
A: In your country?
B: I don't understand INDICATOR

(Varonis and Gass, 1985, p. 76)

Responses are achieved via forms of repetition, expansion, rephrasing, acknowledgement; and reduction of the input.

Examples of responses:

A: This is your 2 term?
B: Pardon me?
A: 2 term, this is this term is term your 2 term RESPONSE

A: Yeah, how long will you be? will you be staying?
B: I will be four months
A: Four months?
B: Stay four months here until April RESPONSE

(Varonis and Gass, 1985, p. 77)

Reactions to responses are an optional portion of the routine and are viewed as the closing feature which is seen to 'tie up' the routine before the main flow of the conversation can again be resumed.

A: My father is now retire

B: retire?

A: yes

B: oh yeah

REACTION TO RESPONSE

(Varonis and Gass, 1985, p. 77)

Comprehension checks manifest themselves within any of the four categories previously shown. They normally take the form of questions and can appear within the trigger, indicator, response or reaction to response:

A: I was born in Nagasaki.

TRIGGER

Do you know Nagasaki?

COMPREHENSION CHECK

(Varonis and Gass, 1985, p. 78)

The previously shown routines are linear and brief in length. This is in contrast to the case of an illustrated example of data which demonstrates a sample of a conversation which took place between two intermediate level language learners within the Varonis and Gass study. In this instance, the coding scheme had to be adapted by the researchers in order to accommodate the levels of complexity and length of the types of negotiation which were prevalent amongst the patterns from the non—native dyads across the data. These instances of negotiation are categorised as extended examples which can last for over two minutes and consist of a series of “embedded non-understanding routines” within the wider picture of the overall negotiation. The linear coding pattern is disrupted as one turn can assume more than one function with the coding categories deployed in a seemingly more flexible manner. A section from this data is shown next:

140 S: But he work with uh uh institution	T
140 J: institution	I
140 S: Do you know that? The name is . . . some thin like eh control of the state	CC/R/T
140 J: aaaaaaaah	RR
140 S: Do you understand more or less?	CC
140 J: State is uh . . . what what kind of state?	I
140 S: It is uhm	R
140 J: Michigan State?	I
140 S: No, the all nation	R/T
140 J: No (back channel) government?	I
140 S: all the nation, all the nation, Do you know for example is a the the institution mmm of the state mm of Venezuela	R
140 J: ah ah	RR
140 S: had to declare declare? her ingress.	T
140 J: English?	I
140 S: No, English no (laugh) . . . ingress, her ingress	R
140 S: Ingress?	I
140 S: Ingress, yes, INGRESS more or less	R
140 J: Ingless	I
140 S: Yes, if for example, if you, when you work you had an ingress, you know?	R/CC
140 J: uh huh an ingless?	RR/I
140 S: yes	R
140 J: uh huh OK	RR
140 S: yes, if for example, your homna, husband works, when finish, when end the month his job, his boss pay—mm—him something	R
140 J: aaaah	RR
140 S: and your family have some ingress	R
140 J: yes ah, OK OK	RR
140 S: more or less OK? and in this in this institution take care of all ingress of the company and review the accounts	CC/R
140 J: OK I got, I see	RR

(Varonis and Gass, 1985, p. 79)

SLA researchers have continued to evaluate the potential value of negotiation opportunities suggesting that learners are exposed to processes, as illustrated in the Varonis and Gass (1985) data, which involve opportunities for clarification, repetition, and the rephrasing of language. It has been suggested that negotiation can enable learners themselves to notice potential discrepancies between the input and their own interlanguage (Mackey, 2007, p. 13). Negotiation of meaning may therefore help to raise learners' awareness of target language forms, for example, when interlocutors make repair moves based on negative feedback as a way to indicate issues. The learners' output may foster second language development by increasing features such as accuracy and comprehensibility (Pica, 1994).

The principles from face-to-face research such as Varonis and Gass (1985), and from the interactionist account in a broader sense, have also been introduced into the area of CALL. Whilst the learning principles remain the same, interactions and processes such as negotiation will be invariably altered and mediated by a variety of different tools such as text-based chat, voice chat, audio and videoconferencing. In the following sections, studies which focus on the negotiation of meaning and draw on the Varonis and Gass (1985) framework across desktop forms of communication are presented and critiqued.

2.8 Research in CALL

Chapelle (2009) adopts a broad and pragmatic perspective on SLA as she notes that “SLA theory needs to encompass the learner, teacher, language, technology and institution” and argues that the communication that the learner engages in via peers and others exploiting the target language is critical in that “communication succeeds when multiple contextual factors come together” (p. 747). Online exchanges and peer-based opportunities for telecollaboration across linguistic and cultural boundaries have addressed a wide range of areas which include development of multiliteracies, the negotiation of meaning, peer feedback; and possibilities for groups of learners to collaborate via a sharing of their social and cultural identities (Guth & Helm, 2010; Hampel & Hauck, 2006; Hampel & Stickler, 2012; Kern, 2014; Lamy & Flewitt, 2012; Tudini, 2012; Van der Zwaard & Bannink, 2014; Ware & O’Dowd, 2008). The negotiation of meaning has continued to prove a ubiquitous, theoretical framework which is exploited in order to support interaction within a wide range of computer-mediated learning contexts. These include eTandem exchanges (Bower & Kawaguchi, 2011); telecollaboration (Van der Zwaard & Bannink, 2014); online classrooms (Hampel & Stickler, 2012); possibilities to learn through technology use outside the classroom (Blake, 2000); and the deployment of technology in face-to-face classrooms (Yanguas, 2010).

CALL is a rather broad research area which initially involved studies in desktop communication and which explores the potential of particular tools in relationship to their potential to support forms of interaction. Studies have examined features of interaction such as negotiation of meaning via email and text-based chat (Bower & Kawaguchi, 2011); text-based chat (Blake, 2000; Lee, 2001; Smith, 2003, 2005; Tudini, 2012); text-based and voice-based chat (Jepson, 2005); videoconferencing and text-based chat (Van der Zwaard & Bannink, 2012); video and audioconferencing (Hampel & Stickler, 2012; Wang, 2006; Yanguas, 2010). Features of negotiation have encompassed opportunities to engage with clarification requests, used as a way to initiate forms of repair (Jepson, 2005); the use of comprehension checks, clarification requests, and self-repair (Lee, 2001); negotiation as a means to overcome communicative problems (Bower & Kawaguchi, 2011); negotiation based around vocabulary items in TBLL (Smith, 2003; Yanguas, 2010); multimodal forms of negotiation (Hampel & Stickler, 2012; Wang, 2006); and abandoned routines and a lack of negotiation seen across videoconferencing tools (Van der Zwaard & Bannink, 2014).

Studies in CALL have also found that learners may avoid attention to form or drawing attention to errors due to a number of reasons which include: an absence of repair moves (Jepson, 2005); the pressures of real-time communication (Lee, 2001); and learners' current proficiency level which causes them to be unable to focus on form (Blake, 2000). In his study, Blake found that within a text-based chat environment, intermediate learners of Spanish negotiated over lexis but less so over grammatical form. He found that the patterns of negotiation which emerged between learners most closely followed the Varonis and Gass (1985) framework. In his data, there were examples consisting of trigger, indicator, resolution; and reaction to resolution patterns but also more complicated and extended negotiation patterns (see Varonis and Gass, 1985). Patterns of negotiation can reveal how peers scaffold interlocutors in important ways in order to help them to comprehend the meaning of new words. For example, Ma (2017) notes how interactionist theories such as Long (1996) and Swain's (1995) output hypothesis can support our understanding as to how forms of 'technology-mediated communication-based lexical learning' can help L2 learners to communicate through the negotiation of meaning.

It is notable from a review of studies in CALL, that aspects of language use in isolation remain the default conceptualisation of the negotiation of meaning, for example, there is no recognition of the work in gesture studies or attempts made to formally categorise gestures and link them directly to examples of L2 spoken language use. One possible reason for this stance is that research into negotiation has tended to focus more extensively on text-based chat rather than to examine learners' negotiation of meaning via tools such as videoconferencing where learners potentially gain access to gesture in relationship to their spoken communication through access to visuals.

2.8.1 Studies in audio and videoconferencing tools

Hampel (2003) first proposed a theoretical framework for language learning via audioconferencing which introduced the notion of multimodality into an examination of second language acquisition within the field of CALL. She suggests how theories such as the negotiation of meaning and TBLL can be understood in relationship to multimodality, leading to new forms of practice. Hampel's conclusions emphasise the importance of building on pedagogic rationale, to learn from previous research; and the necessity to take account of the challenges and drawbacks of any particular technological medium (p. 34). Wang (2006) attempted to extend paradigms of negotiation through positing a new taxonomy of CMC (computer-mediated communication) which was based around

interactions which encompassed the following: written, oral, and oral-visual interaction. She notes how research into visual-oral interactions enabled by VC has remained less common and lacks depth. However, there have been subsequent studies which have explored aspects of negotiation through the detailed analysis of learners' and teachers' exploitation of multimodal forms of communication in terms of how these can be harnessed to support interactions within online classrooms (see Hampel & Stickler, 2012). There have also been innovative frameworks of social presence put forward which are conceptualised from expansive multimodal perspectives in CMC (Satar, 2015). Lamy and Flewitt (2012) depict online conversations as 'multimodal' with analysis as to how participants collaboratively orchestrate various modes of communication, including gesture, within their desktop exchanges in order to establish forms of intersubjectivity across distance.

Wang's (2006) study of VC in *NetMeeting* explored five learners studying Chinese on distance programmes with a teacher/researcher. She examined attention to focus on form during meaning-based tasks. The study exploited the Varonis and Gass (1985) framework of the negotiation of meaning as a scheme for coding and analytic purposes. Importantly, learners were also interviewed about their online experiences and she notes that via the available tools "a multimodal learning environment was offered to the learners" (pp. 138-139). She immediately highlights the shortcomings in the original Varonis and Gass' (1985) coding categories by positing that "visual indicators of non-understanding constitute another type of indicator which was not covered in Varonis and Gass' study" (Wang, 2006, p. 129). Wang cites examples such as puzzled expressions by learners as representative of visual indicators of non-understanding in ways which can also help interactants to establish mutual understanding due to the affordances of the video. In further broadening the scope of the original Varonis and Gass framework, she also suggests that learners may experience communicative issues around the request for new vocabulary.

Wang offers an example of negotiation over lexis:

- A: Do you have many colleagues who speak Chinese?
- B: Colleagues? What does colleague mean?

The learner also appeals for assistance via a direct request for new words:

A: Do you have any friends who speak Chinese?

B: No, but I have a ... what is the word for describing people who work with you?

A: Colleagues

(Wang, 2006, p. 125)

Despite the claim to have approached the negotiation of meaning from multimodal perspectives, Wang's transcripts of the online interaction somewhat belie this research aim. They principally, and almost exclusively, focus on the linguistic utterance, coded according to the Varonis and Gass (1985) framework. As a result, the 'visual-oral' component within negotiation is left underexplored in the analytic process with gestures reduced to examples such as the researcher holding up three fingers in order to convey the number three or a learner seen tracing written characters into the camera. These are ill-defined and cursory references to gesture which cannot be used to draw any firm conclusions as to how the mode operates in relationship to participants' speech within their negotiation of meaning. Other than a few cited non-verbal cues, the transcripts of the negotiation routines serve as a mono-modal, rather than a multimodal, representation of the negotiation of meaning across technology.

Yanguas (2010) also examined the negotiation of meaning, involving classes of intermediate learners of Spanish studying at a university in the US. He placed fifteen learners into dyads in order to extend opportunities for oral computer-mediated interaction (OCMC) interactions via the implementation of a vocabulary task enabled by desktops and encompassing a comparison of audio, video, and face-to-face forms of communication. Learners completed a jigsaw task with the rationale that this can "force learners to equally share their (different) information to achieve a particular goal" (pp. 75-76). The inquiry focussed on negotiation around target lexical items which were then found to trigger routines across the three cohorts. The negotiation routines were identified using the Varonis and Gass (1985) framework. The results are interesting in that Yanguas quotes the percentage of learners' use of signs as evidenced in over fifty-five percent of turns within the video group. He illustrates a negotiation routine which contained a learner's use of a sign as a response move in VidCMC as follows:

A: y tambien tengo un toldo
[and I also have an awning]
B: Que es esto?
[What is this?]
A: es como...[MAKING SIGNS FOR SHELTER]
[it's like ...]
B: ¡oh! como casa
[Oh! like a house]
A: si
[yes]

(Yanguas, 2010, p. 83)

The researcher does not choose to categorise or analyse what type of gesture the learner is using in their response move within the negotiation routine. Therefore, the reader cannot ascertain, from the verbal transcript alone, whether the gesture co-occurred with the words “es como” or whether it followed these words. Firstly, the gesture could be interpreted to represent a communicative strategy, designed to depict the linguistic concept of an “awning” when this is apparently unknown to an interlocutor. Secondly, it could be acting as a form of lexical search as the individual speaker attempts to hold the visual concept in their thoughts as they try to remember the word for “awning.” However, the interlocutor immediately proffers the word “house,” which, it would seem, is different to the original trigger to the negotiation which was based instead around the word “awning.”

Yanguas comments on learners’ use of visual cues in somewhat negative terms in relationship to a perception of their value regarding their development of language within the interactionist account. Nevertheless, he does question the possible value of a linguistic concept being made visually manifest by another learner, noting that these signs were strategies designed to clarify and resolve a communicative problem when learners had access to visuals via their use of the video channel on Skype.

It seems that the impossibility of using visual cues in the AudCMC dyad pushed student A to elaborate on the lexical item that caused the non-communication. On the contrary, Student A in the VidCMC group used signs to tackle the communicative problems. In both cases, the result is the same but in terms of both output and linguistic input the scenario [...] does not seem to be ideal for L2 learning [...] However, these more linguistically elaborated responses had negative consequences, namely, they translated into a higher percentage of partially understood target lexical items [...] The question is then what is more important for L2 acquisition, to linguistically elaborate on the item augmenting the possibility of non-communication to occur, or to fully and unequivocally understand the concept being focussed on as a source of non-communication?

(Yanguas, 2010, p. 83)

This is an intriguing question as it raises the issue as to whether it is more productive for learners to continue to linguistically negotiate in order to establish understanding or whether their acquisition is better served through their ability to ‘see’ a concept being described, for example, through an interlocutor’s use of an iconic gesture. However, the previously discussed literature on the role of gesture in SLA (which is never referred to in this study) illustrates that, in settling for non-visual forms of communication such as audioconferencing, we are perhaps depriving language learners of a valuable resource which has been shown to support their acquisition of language in a number of different ways in relationship to their use of co-occurring talk.

Hampel and Stickler (2012) investigated learning interactions in German, including negotiation of meaning, on the desktop platform *FlashMeeting* in exchanges between teachers and learners which enabled use of visuals, text, audio and video. They note dichotomies between sociocultural and psycholinguistic approaches to acquisition and instead adopt a socio-interactionist approach in their analysis which includes observed patterns in negotiation such as clarification and confirmation requests (Long, 1996). They were interested in how interactions are influenced by the affordances of the learning environment itself; and identify the increasingly wide range of communicative modes available to interlocutors within an online classroom setting. Whilst identifying gesture as a mode within VC, they conclude that technical issues, regarding clarity of visuals, significantly impact the use of body language and gestures from being exploited as an

effective communicative tool. Conversely, participants adapted their communication and exploited an array of different tools in order to achieve their goals, for example, text-based chat was used to engage in clarification requests over vocabulary items and audio served the purpose of enabling discussion of technical issues. Another especially notable aspect of the interactions described is the considerable dominance of the teacher who “spoke most and took turns very frequently” (p. 122). However, it is important to note that speech was only one aspect of the multimodal communication documented, for example, participants were seen to bypass the teacher via their use of chat tools in order to have private ‘off-task’ conversations; and were seen to strategically adapt modes to their needs and the communicative environment itself whilst developing a “new culture of interaction” (p. 133).

Van der Zwaard and Bannink (2014) draw conclusions about the advantage of chat modes over VC, using the Varonis and Gass (1985) framework. They conclude that differing forms of communication in CMC can result in differing patterns of interaction and negotiation:

The negotiated interactions in our data seem to be shaped and influenced by the mode of communication. Since all participants communicated through both video call and chat during task-performance, the unique, distinctive features of the specific mode of SCMC seem to model the pattern of negotiation of meaning episodes.

(p. 145)

Drawing on Goffman (1959), Van der Zwaard and Bannink (2014) note that the speakers interacting via video are faced with “complex, often contradictory, sign activities in which the L2 learner would give verbal signs of understanding but give off non-verbal signs of confusion or non-understanding” (p. 145). It is not clear from this study exactly which features of communication are considered to be ‘non-verbal,’ other than the placing of sparse details such as learners ‘leaning towards the camera’ within the written transcripts of the negotiation routines. The authors also mention participants fidgeting with clothes and hair but do not distinguish these inadvertent movements from learners’ seemingly deliberate movements in their analysis. It is concluded within their findings that chat is preferable as it focusses “the entire burden of communication on written characters which allows for more explicit and unambiguous statements of non-understanding” (p. 145).

In the final section of this chapter, research into language learning with mobiles is discussed and critiqued in relationship to the aims of the present study.

2.9 Language learning with mobile devices

Learning with mobiles has enabled new possibilities to acquire languages in communicative ways. Mobile devices such as tablets and smartphones represent ubiquitous tools which can be used to encourage learners to capture aspects of the world around them whilst moving through the spaces of their everyday lives. From this perspective, social media platforms and video-based forms of communication offer participants possibilities to capture the world in personally meaningful ways in order to contextualise their language use and to share their experiences with others. Al -Shehri (2011) argues that the majority of EFL practices can be characterised as traditional language learning settings in which teachers direct the learning process, with learners merely seen as “passive receptors of knowledge” (p. 278). He views mobile technologies as capable of providing ‘context in our pockets;’ offering the potential for learners to both use language and contextualise its use from a range of real-world settings beyond the language classroom (p. 278). Conversely, the area of MALL has sometimes borrowed from more traditional notions of learning and has somewhat focussed on individualistic and behaviourist approaches to the detriment of learning in more collaborative and contextualised ways. Yet there are strong arguments that language learners require opportunities to communicate and interact as learning a language is conceptualised as a social event (Dörnyei, 2001). MALL has also been defined as the use of mobile technologies in language learning, where device portability offers specific advantages (Kukulska-Hulme, 2012, p. 1). As a consequence, there would appear to be little purpose in simply transferring passive learning activities from a coursebook onto a mobile device. Especially in earlier research, there has been a failure to engage with features of devices such as their mobility and peer connectivity; with areas such as collaboration and the development of speaking skills neglected (Kukulska-Hulme & Shield, 2008).

Areas of research interest in the field of MALL have encompassed aims such as vocabulary learning (Kennedy & Levy, 2008; Lai & Zheng, 2018; Song, 2008; Stockwell, 2007); listening skills (de la Fuente, 2014; Nah et al., 2008); reading (Hellermann et al., 2017; Wu et al., 2011). Approaches to research in MALL are also changing as affordances of devices such as their mobility, personalisation, connectivity, multimodality, and context-

sensitivity become increasingly recognised. These types of affordances can be exploited to enable language learners to collaborate and communicate with one another from a diversity of settings beyond the language classroom (Al-Shehri, 2011; Comas-Quinn et al., 2009; Kukulska-Hulme et al., 2017; Ogata et al., 2006; Ros i Solé et al., 2010; Wong et al., 2012).

In terms of positioning the area of language learning with mobiles within technology-mediated forms of communication, there have been arguments that this specific field also overlaps with CALL and M-learning (Mobile learning): “The research, insights, and experiences from its cognate disciplines of CALL and other domains of mobile learning provide valuable, perhaps even crucial, inputs to the field” (Stockwell & Hubbard, 2013, p. 5). Figure 3 (shown next) illustrates the interface between these related yet distinct fields.

(Image redacted)

Figure 3: Relationship & overlap within the 3 fields (from Stockwell & Hubbard, 2013).

The present study draws on a number of tenets from MALL such as the requirement to consider available affordances which include device portability, learner mobility, connectivity; and opportunities for context-sensitive forms of task design. However, it was also influenced by the research in CALL where multimodal forms of interaction via tools such as videoconferencing via desktops have been previously demonstrated to support language learning through recognition of theories and frameworks such as the negotiation of meaning (Hampel & Stickler, 2012; Wang, 2006).

Mobile technologies also offer the possibility to transform learners’ multimodal communicative experiences across a variety of formal and informal learning contexts. It has been suggested that characteristics of mobile devices such as their portability can be linked to learners’ deployment of every-day life-situations with mobiles allowing people to capture these situations, as well as creating the possibility to stand back and reflect on them (Pachler et al., 2010, p. 7). Kukulska-Hulme and Lee (2020) argue that formal and informal aspects of language learning within MALL operate along a flexible and nuanced continuum; and that teachers may be required to help to support language learners to understand how best to avail of their collaborations which take place beyond the language classroom. From this perspective, digital technologies forge a path for new pedagogies but also the adaptation of existing ones (Pegrum, 2014, p. 25).

This is not an easy pedagogic balance to achieve, as learning beyond the classroom ideally provides authentic and autonomous language learning experiences but may equally expose learners to interactions within unpredictable settings. It has also been argued that it is critical for teachers to prepare students for the activities and to provide follow-up in the classroom (Nunan & Richards, 2015).

2.9.1 Mobiles and SLA

The previous section of the literature review, which discussed learners' interaction via tools such as chat, video, and audioconferencing, has shown that the promotion of theories such as the negotiation of meaning is still prevalent in approaches which aim to encourage interactive forms of communication across desktop technologies. Conversely, there has been more limited discussion as to the importance of SLA within approaches to the design of learning activities based around mobile devices. Nevertheless, Stockwell (2016) highlights that "SLA theories can do a great deal to inform researchers about what can be achieved when learning through mobile devices" (p. 297). A review of Burston's (2015) meta-analysis of studies in MALL across a period ranging from 1994 to 2012 gives an overview of the types of studies conducted in the field. It is notable that there is an absence of more communicative approaches to learning, descriptions of theoretical frameworks within SLA or recognition of learners' multimodal forms of communication across mobiles.

Baleghizadeh and Oladrostam (2010) conducted a study where forty female pre-intermediate learners of English on courses at a university in Iran were asked to record their speech on their mobile phones as a way to elicit grammatical forms. The out-of-class assignment involved a review of the recordings and an opportunity for the learners to report their analysis in a subsequent classroom session. The aim was to test two groups, an experimental and control group, on grammatical forms which had been previously under review in the face-to-face classroom. The findings illustrate that the multiple choice post-test showed a significant improvement in grammatical accuracy within the experimental group. Conversely, Lan et al. (2009) exploited constructivist approaches to learning with tablet devices in order to observe how young Taiwanese learners studying English deploy these in order to complete reading activities. It was found that the learners collaborated with one another in face-to-face scenarios in order to share what they had learned and that peers took control and temporarily assumed the role of teacher.

Think-aloud protocols and the theory of 'noticing' (Schmidt, 1990) have also been deployed in order to find out the extent to which participants pay attention to aspects of form within their listening activities on iPods (de la Fuente, 2014). However, the use of iPods already appears outdated as devices such as tablets and smartphones would presumably replace the use of this stand-alone technology through offering an array of multimodal possibilities for learners to engage in listening activities. The research compared two learning conditions: MALL versus instructor manipulated forms of learning via computers in a lab. It was found that the MALL group demonstrated significantly higher levels of noticing and comprehension of linguistic forms. The study focussed exclusively on linguistic outcomes, however, Guichon and McLornan (2012) have argued within CALL that learners' comprehension of L2 benefits most from sources of information which are conveyed through multiple modes. It was found in their study that exposure to audio-visual videos was particularly valuable in terms of learners' successful comprehension of language.

Nah et al. (2008) directly link their discussion of SLA theories to forms of collaborative learning supported by mobile phones. They exploited a combination of sociocultural and interactionist perspectives on SLA: "[t]he study is based on input, interaction, output, and sociocultural theories [...] as well as collaborative, learner-centred, constructivist task-based approaches" (p. 331). That study was interested in exploiting the internet via mobile phones in order to improve listening skills for a group of intermediate level learners studying English at a university in Korea with participants encouraged to collaborate with teachers and peers in a communicative manner. Learners themselves were given questionnaires and interviewed about their attitudes to using mobile devices in order to complete the listening tasks. Responses showed that the majority of students found the use of mobiles both convenient and interactive, however, the research also highlights some disadvantages such as one particular learner struggling to concentrate enough in order to complete their assignments on the train or bus: "I could not concentrate on them" (p. 343). The findings also claim that learners benefitted from increased opportunities to negotiate for meaning via mobiles. However, whilst an interesting research conclusion, the negotiation of meaning is never defined. There is an absence of an analytic framework to elucidate how the research chose to conceptualise the negotiation of meaning; and a lack of data evidence which could have revealed information as to how learners go about achieving this via their use of mobile technologies.

2.9.2 Language learning opportunities from beyond the classroom

It has been noted that learning beyond the classroom (Benson, 2011) or ‘out-of-class’ learning can enhance classroom work; and that learners benefit from the use of technology and the internet as resources which can be used to support this (Nunan & Richards, 2015). Learning beyond the classroom is a field in MALL which is attracting increasing attention as the affordances of devices offer the potential for researchers and teachers alike to create a range of collaborative learning opportunities. Kukulska-Hulme et al. (2017) note how the deployment of mobile devices “potentially extends learning beyond the classroom setting [...] where learners are continuously on the move [...] in outdoor settings or in places where everyday life and leisure activities merge with learning” (p. 217). Mobile blogs have been previously exploited as social tools which can be used to build ‘communities of practice’ (Lave & Wenger, 1991) in ways which support learners to use language in informal, situated, and contextualised ways (Comas-Quinn et al., 2009). Participants in that study deployed mobile phones, digital cameras and MP3 players in order to record examples of their encounters with the local culture *in situ* which could then be shared with peers online. Whilst the researchers position the type of learning which occurred within their study firmly within the category of ‘informal,’ there is an acknowledged encouragement for the learners taking part to avail themselves of their surroundings, to capture interesting images, and to share these with peers. For example, learners were required to give their images titles and annotate them with comments. Whilst the approach is defined as ‘accidental learning,’ it could equally be argued that the ideas suggested by the researchers could be seen to represent a loosely-structured pedagogic task which reflects the opportunities created by mobile devices when used from beyond the classroom. The study recognises the importance of fostering learner autonomy and the enabling of opportunities for learners to become active content creators. It inadvertently suggests how teachers might help learners to avail of the multimodal and social features of their mobile devices in relationship to capturing aspects of their immediate surroundings.

In contrast, Lai and Zheng (2018) evaluated learners’ self-directed use of mobiles, without intervention from researchers or teachers. They found that the learners taking part principally used their devices to personalise their learning, rather than to enhance authenticity or to exploit key features such as social connectivity. The technology was used in valuable, but exclusively individualistic, ways in order to develop skills such as learning words, googling information, reading, and watching videos, rather than to engage in

collaboration. Lai and Zheng conclude from their findings that “educational mediation and interventions are needed to enhance this aspect of mobile learning in the context of foreign language learning” (p. 312). It is suggested that teachers can initially help learners to build communities across classes as a way to eventually engender more autonomous networks within daily life (pp. 312-313). Without pedagogic intervention, learners may be denied important opportunities to communicate and, in the case of learning, may exclusively perceive their devices as pocket dictionaries or resources to support individual tasks.

Learning beyond the classroom also implies that learners will operate in more autonomous ways where the teacher may not be present at the time of the interaction or task. This may represent a new experience for many learners who may be more used to the discipline and structure of a face-to-face or online classroom where the teacher is always present.

Kukulska-Hulme (2013) argues that learners need to be ‘re-skilled’ in order to equip them for a world where learner autonomy will be increasingly valued and supported. In terms of the creation of learner-centred social spaces, language learning with mobiles entails opportunities to test out language, to acquire knowledge; and to form social relationships without the ‘authoritative voice’ of a teacher (Ros i Solé et al., 2010). Teachers operating within MALL may choose to adopt a less authoritative role, however, they will still need to design tasks, set up activities, discuss problems, provide forms of feedback; and enable learners to achieve their goals across a diversity of settings from within but also beyond the classroom. This approach requires a re-evaluation of pedagogy with the development of some potentially new skills for language teachers. These skills include a recognition of support for learners’ digital and multimodal literacies, the ability to integrate activities within and beyond the classroom; and forms of effective task design with an identification of available resources for learners (see Kukulska-Hulme & Traxler, 2005; Pegrum, 2014; Kukulska-Hulme et al., 2017 for an overview).

2.9.3 Multimodal literacies

The New London Group (Cope & Kalantzis, 2000) has argued that technology and society are constantly evolving in ways which reflect increased levels of global connectedness and cultural diversity. As a result, the group put forward ‘a pedagogy of multiliteracies’ designed to recognise the multiplicity of ways that people exploit and reshape the representational resources around them in order to make meaning. They suggested that a focus on modes needs to include a recognition of linguistic meaning but also gestural, visual, audio, and

spatial forms of meaning-making. Within the field of TEFL, Royce (2013) has posited that traditional notions of L2 ‘communicative competence’ should be extended. This is in order that teachers support learners to comprehend the interrelationships which operate between language and the deployment of other semiotic modes in terms of the creation of holistic meaning.

Stickler and Hampel (2015) note that the use of digital technologies has created spaces which are “multimodal, multicultural, and multilingual” (p. 63) and that teachers require new skills in order to support their learners. Lancaster (2013) remarks how the use of digital technologies has extended opportunities for people to create “new kinds of literate texts,” for example, mobiles can be used to convey “images, color, sound, music, digital images, and writing in multiple combinations” (p. 2) in ways which language teachers need to begin to recognise. Portable digital devices can be exploited in communicative scenarios “on the street, in public transport, and in shops and cafes” in ways which position the L2 learner in “a digitally multimodal world, creating new schema for participation and meaning-making” (Lotherington & Jenson, 2011, p. 227).

Taking account of modes such as gesture, in relationship to the use of digital technologies, poses new challenges for language teachers in that their training and skills set has been traditionally focussed on the notion that language use operates in isolation. Within teacher training programmes such as the DELTA (Diploma in Teaching English to Speakers of Other Languages) the mode of gesture is left completely unaddressed. Furthermore, there has been a lack of information within training programmes within TEFL as to how portable digital technologies used in the everyday world might offer new possibilities for multimodal meaning-making which are not available from within the L2 classroom setting. Whilst discussing the question of transformation in the area of literacy practices, it has been suggested that teachers need to negotiate with their learners and to ask relevant questions such as whether as educators we want people to be able to communicate effectively or to adhere to a narrow, imperialist view of English which fails to reflect second language learners’ lives (see Lea and Lillis, 2016, p. 388). Traxler (2013) discusses language as a social practice in relationship to how our society is being transformed by mobiles via features such as their connectivity and mobility. Through their use of mobiles, language learners are now exposed to authentic ‘out-of-class’ experiences in which they “interact and communicate with the context in which a mobile task is taking place” (Al-Shehri, 2011, p. 278). The multitude of learning resources and contexts which learners

interact with in their everyday lives can also be harnessed to assume a role in more formal learning situations, however, these are too often overlooked by both learners and teachers (Wong et al., 2012).

2.9.4 Context-sensitive paradigms of learning

In their discussion of ‘context-sensitive learning’ with mobiles, Pachler et al. (2010) highlight how interactions which involve conversations, including non-verbal forms of communication such as gesture, can be used to support learners to engage in meaning-making processes. Within the field of mobile learning, they discuss high-tech systems which have been specifically designed for use across mobile devices such as ‘CAGE’ (Context Aware Gallery Supporter) as part of the ‘MOBIlearn’ project. CAGE enabled participants to interact with people, technology, and artefacts as they learned directly from their immediate surroundings (p. 44). Conversely, most language teachers and learners do not have access to these kinds of sophisticated interfaces, however, participants may be enabled to interact in more achievable ways via their mobile devices if photos, video, and freely available social-media platforms are fully exploited from a range of settings beyond the classroom. Within models of situated learning (Lave & Wenger, 1991), it has been posited that participants acquire knowledge in social ways and within specific ‘contexts’ as learning is embedded in real-world activities and achieved through gradual degrees of participation. Several studies in MALL have illustrated how settings beyond the classroom operate in combination with the strategic use of devices in ways which enable peers to communicate in their L2 (Al-Shehri, 2011; Comas-Quinn et al., 2009; Ogata et al., 2006; Ros i Solé et al., 2010; Wong et al., 2012).

Ros i Solé et al. (2010) adopted an ethnographic research lens in order to investigate how learners relate to and exploit MP3 players/recorders in their daily lives. They emphasise the importance of charting the learner experience itself. Two-way opportunities to produce spoken language were enabled as learners were asked to conduct interviews with native speakers. The speaking tasks in this study occurred within settings such as cultural and religious spaces as it was noted how technology allows learners to become “immersed in authentic contexts” (p. 48). Learners were then afforded opportunities to re-evaluate their recordings from within a classroom setting. It was concluded that the mobile devices importantly facilitated personalised, contextualised, and social forms of learning.

Vocabulary learning in MALL has traditionally relied on so-called ‘push’ mechanisms for learners such as the teacher sending text messages in Italian (see Kennedy & Levy, 2008). Conversely, a study of thirty-three EFL university students who were learning vocabulary as part of a course in Saudi Arabia were instead encouraged to create their own learning resources via text-based comments and photos from beyond the classroom. The advantage to the latter approach within MALL is that learners become engaged as active content creators. Furthermore, developments in technology have ensured that the multimedia capabilities of devices have become more sophisticated and accessible to learners than, for example, a decade ago. Al-Shehri (2011) found that adopting a more communicative and contextualised approach to learning with mobiles: “reflected the learning context and students’ own choices” and “allowed them to create meaningful learning resources via their daily interaction with the external environment” (p. 283). Learners will already be familiar with capturing information in visual formats on their mobile devices from settings beyond the classroom, however, they may need support to understand how interactions within more informal settings can also be harnessed as opportunities to learn a second language.

The LOCH (Language learning outside the classroom with handhelds) project (see Ogata et al., 2006) illustrated how language learners of Japanese were supported by their teacher to integrate knowledge which they had previously acquired in the classroom through subsequently highlighting learners’ real-life language needs out in the world. The project exploited mobile technologies as a way to shift the learning environment beyond the classroom in order to place learners in situations which they would regularly encounter in their daily lives: from places such as streets, transport hubs, supermarkets, and the workplace. Teachers implemented a range of authentic tasks which included asking learners to interview the public, make transactions at the local station; and to buy local produce and ask how to make certain recipes. The rationale was that learners could develop their spoken communicative skills in productive and contextualised ways with the reassurance that the teacher was available for them at different stages of the process. Learners actively participated, recorded events, and took photos. The multimodal data was then brought back to the classroom and the smartboard was exploited by learners and teachers as a collective space in which to share the captured data, to stimulate discussion; and to offer solutions to any problems which the learners had encountered in their everyday scenarios out in the world.

Wong et al. (2012) aimed to harness learning resources and contexts which learners of Chinese had already exploited in their daily lives via a model for an artefact-orientated approach in analysing these types of learning settings. The researchers aimed to highlight how concepts such as mediation via objects and forms of distributed cognition can be exploited to illustrate how participants can be enabled to bring their vocabulary learning to life through harnessing *in situ* resources within ‘real-life contexts.’ The researchers highlight their pedagogic stance as they question why vocabulary is frequently delivered via forms of explicit instruction within the classroom and, as a consequence, taken out of its natural context of use. The researchers encouraged the learners in the study to reflect on their environment, and to learn to associate contexts (or creating contexts with the aid of physical objects and people) with a range of previously learnt language items (p. 412). They argue that language items such as idioms are particularly complex and require examples of usage within real-life contexts. The study concludes that artefacts transform their roles through learners’ appropriation for carrying out different learning tasks and that the teacher can support learners through forms of scaffolding.

Hellermann et al. (2018) also reconsidered the seemingly isolated activity of reading as ‘embodied practice’ when learners were given the opportunity to participate in an augmented reality game based around the notion of ‘place.’ It has been posited that the most profound aspect of learning with mobiles is not ‘anytime, anywhere’ but in the experience of place itself (Squire, 2009). The English language learners taking part in this study were given gaming tasks which involved them locating various different physical destinations around a US campus. As a group, the learners were prompted to read, walk, and talk to one another in embodied ways as they read aloud, listened to instructions, negotiated information and attempted to achieve intersubjectivity. As the learners only had access to one device between them, they were pushed to share information and to collaborate in order to play the game and to work together in order to achieve the task goals. The importance of understanding the issue of context for language learners reaches beyond the realm of the individual researcher: it is “born out of deep pedagogic concern, as teachers around the world grapple with complex communicative situations” (Lillis, 2008, p. 354).

2.10 Conclusion to the literature review

As new forms of technology-mediated communication emerge, such as the use of VC via mobile devices, the requirement to examine learners' multimodal forms of communication in terms of SLA increases. Whilst there has been an important body of research conducted into the effectiveness of desktop VC to support learners' multimodal interactions, there has been less focus on multimodal forms of communication within MALL. There is also a growing amount of work focussed on the specific role of L2 gesture in SLA within face-to-face language learning contexts but this is not evidenced in the research on learners' use of technology. In this chapter, I have demonstrated that gesture is a prevalent feature of spoken communication and, as a result, an analysis of the mode is relevant to the development of our understanding of SLA within communicative language teaching and learning processes. Conversely, there is a dearth of studies which have chosen to ask L2 learners to reflect on their gestures as a way to potentially support them to understand aspects of their multimodal communication. In terms of the negotiation of meaning, the majority of previous research has been pre-occupied with revealing psycholinguistic processes in mono-modal ways with the issue of context largely ignored. Yet this chapter has illustrated how the field of MALL, and its corresponding pedagogies, highlight a perceived need to focus research on enabling opportunities for language learners to communicate from beyond the classroom, with the importance of context increasingly emphasised.

As a concluding note to this chapter, the negotiation of meaning is defined as pertaining to examples of task-based dialogue where language learners engage in the attempted clarification and resolution of triggered incidents of non-understanding. The triggering of these incidents through to their resolution is described in an holistic sense as the negotiation of meaning. This particular theoretical stance on negotiation, is evidenced in much of the earlier interactionist literature in SLA within face-to-face and technology-mediated language learning (see Blake, 2000; Long, 1996; Varonis & Gass, 1985; Van der Zwaard & Bannink, 2014; Yanguas, 2010). Long's (1996) IH suggests that learners who engage in conversation will ideally create the right conditions for L2 learning to occur; particularly when they experience a range of problems in understanding one another. For this reason, the negotiation of meaning is evidenced in the researcher's isolation of examples of dialogue where incidents of non-understanding (forms of problematic communication) are triggered. Following a perceived incident of non-understanding,

interlocutors then face a choice as to whether to ignore the problem or to ask for forms of clarification (through lines of questioning, forms of repetition, elaboration etc.) in order to indicate that there is an identified problem and to then resolve this. The relationship between incidents of non-understanding, and the subsequent opportunities for speakers to attempt to resolve the problem, is defined by Varonis and Gass (1985) as “exchanges in which there is some overt indication that understanding between participants has not been complete” (p. 73). The series of conversational steps, including the original move which first triggered the non-understanding, are deemed to represent episodes or examples of the negotiation of meaning (Varonis & Gass, 1985, pp. 73-79). This particular conversational process is considered beneficial to SLA because it ideally prompts language learners to notice their issues and to then try to resolve them through the ongoing modification of their problematic communication. Conversely, this particular study has chosen to expand on the original concept of the negotiation of meaning beyond an examination of language use in isolation and towards a multimodal reconceptualisation of the original model by Varonis and Gass (1985).

In consideration of the literature and an identification of the existing gaps in research led me to formulate the following questions:

- 1) *What is the role of gesture in the negotiation of meaning when learners are communicating via Skype VC accessed on mobiles from beyond the language classroom?*
 - a) *How are gestures exploited and responded to between learners during their negotiation of meaning?*
 - b) *What are the affordances and constraints of mobile technologies, and the settings they enable, in relationship to learners' deployment of gesture during their negotiation of meaning?*
- 2) *To what extent can a supported focus on gesture and language use help learners to understand aspects of their multimodal communication during the negotiation of meaning?*

Chapter 3 Methodology

3.1 Introduction

The principal research question in this thesis pertains to an interest in understanding the role of gesture in terms of how gesture and speech may be harnessed intermodally by language learners within the negotiation of meaning. This would be supported through their deployment of mobile technologies in settings beyond the traditional language classroom. Another aim of the research is to understand how learners might be supported to verbalize and reshape their experiences through the creation of opportunities to enable them to reflect on, and ideally learn from, their collaborative orchestration of multimodal forms of communication. An existing framework of analysis for the negotiation of meaning (Varonis & Gass, 1985), previously exploited to analyse face-to-face and CMC communication in language learning contexts, was extended in order to allow for the multimodal transcription of learners' negotiation of meaning. These data were then triangulated with the learners' stimulated recall interviews to aim to answer the research questions from a more in-depth and holistic perspective and as a way to take account of the learner experience.

The overall approach to the research first involved video-capturing the learners' task-based VC interactions from a range of settings in a city in the UK. These included the following settings: restaurants, cafés, a public garden, a museum and art gallery, and a stately home. The pedagogic task design for the purposes of enabling speaking activities beyond the language classroom aimed to harness learners' use of VC across mobile technologies with opportunities for both the learner and the technology itself to 'become mobile.' The aim was that the dyads would collaboratively explore a diversity of situated artefacts which were physically located within informal spaces and which would then be shared in virtual ways across VC in a manner which would potentially stimulate the negotiation of meaning. In the present study, learners were therefore asked to operate across a hybrid space which involved them exploiting aspects of both the material and virtual world simultaneously. Keating and Sunakawa (2011) explain how the issue of multimodality operating across real and virtual spaces entails people's use of gesture and talk but also their engagement with features of the context which include interaction with physical spaces and artefacts (p. 194).

In this chapter, the approach to the current research study is first discussed in relationship to key philosophical aspects within qualitative approaches. This is followed by a description of the roles of the researcher and ethical considerations around the implementation of a research study which was designed in relationship to learners' use of mobile devices and the settings which they enable. Next, there is an outline of methods in the field of multimodal research and then my approach to the research study is explained as I link the methodology to the research questions. In the following section, the chosen framework of the negotiation of meaning is described. I then give an overview of the chosen data collection instruments, describe the pilot study, and explain the learning curve which this process entailed. I next present the task design in terms of how I identified and harnessed a range of material artefacts which were contained within a series of informal settings beyond the classroom as a way to support opportunities for the negotiation of meaning to occur between learners. Following this, other possible methodological approaches are discussed and the main research study is outlined with critical discussion of the chosen data collection instruments in terms of the aim to capture and analyse learners' gestures and speech within the negotiation of meaning. The data analysis procedures focus on a multimodal methodological toolkit which I used to code and transcribe gesture and speech in intermodal interplay; and I explain the role of the data analysis software ELAN. I follow this with a section which describes the final phase of the design of the multimodal transcripts. Finally, I outline my approach to coding and analysis of the stimulated recall interviews and explicate the purpose and aims of the qualitative content analysis.

3.2 Philosophical perspectives on qualitative research

In the present study, I subscribe to a qualitative paradigm and adhere to its philosophy due to the exploratory nature of the research. Within a qualitative paradigm, there is no distinct set of methods or practices which are entirely its own and, therefore, qualitative research represents different things to different people (Denzin & Lincoln, 2011). It has been found that qualitative work in the field of language teaching and learning is distinguished by considerable "methodological eclecticism" and a lack of adherence to established tradition (Benson et al., 2009, p. 79). Richards (2003) also argues that qualitative research involves "rigour, precision, systematicity and careful attention to detail" (p. 6). As there had been no systematic multimodal framework in existence to examine gesture and speech within the negotiation of meaning in a mobile learning context, I drew on several methodological influences to address the specific aims of this exploratory research study. I therefore adopt

an interdisciplinary approach to the methodology which exploited, adapted and integrated existing approaches within the fields of applied linguistics (Varonis & Gass, 1985) and gesture studies (McNeill, 1992; McNeill & Duncan, 2000). I also adopt the stance that people's engagement with multimodal forms of communication does not take place in isolation but is situated within a wider ecology or context (Streeck et al., 2011).

The present study adhered to many of the suggested principles within qualitative research paradigms through various stages of the process. While there are multiple approaches to qualitative studies, there are specific guiding principles which supported this research. Creswell's (2013, pp. 1-25) description of qualitative research acted as a useful starting point regarding my awareness of the philosophical underpinning of the present study and in terms of enabling me to reflect on my own stance as a researcher around aspects of ontology, epistemology, axiology, rhetoric, and methodology. An awareness of the philosophical domain of ontology is concerned with the nature of reality which goes on to shape and influence the epistemological stance adopted by the researcher. Epistemology entails their ability to critically evaluate and to justify any claims for knowledge. Within the tradition of SLA, researchers regularly observe an event, frequently a spoken language task, and subsequently implement coding procedures to enable them to organize, account for, and try to explain the data in terms of establishing and interpreting patterns based around a particular theory (Long, 1996; Varonis & Gass, 1985). However, I also addressed aspects of learners' multimodal forms of negotiation through enabling them to talk about their experiences following their collaborative speaking tasks on VC. One of the purposes of interviewing them was to provide a secondary viewpoint on negotiation as I had wished to gain a different perspective on my own observations of the learners' speaking tasks. I also aimed to uncover the ways in which a reflective focus on gesture and language use could potentially support language learners to explore and ideally learn from their forms of multimodal communication on mobiles. For example, the previously presented literature review demonstrated some potential ways in which gesture can be exploited as a way to support forms of language acquisition in multimodal ways.

Claims for validity and reliability within a qualitative research paradigm have been defined by guiding principles which helped to support the various phases of my own research process. These included aspects of Lincoln and Guba's (1985) taxonomy:

- *Credibility* or the 'truth value' of a study, which is the qualitative counterpart of 'internal validity.'
- *Transferability* or the 'applicability' of the research to other contexts, which is the qualitative counterpart of 'external validity.'
- *Dependability* or the 'consistency' of the findings, which is the qualitative counterpart of 'reliability.'
- *Confirmability* or the 'neutrality' of the findings, which is the qualitative counterpart of 'objectivity.'

Methods which helped to support Lincoln and Guba's qualitative taxonomy in the present study included the triangulation of data sources (Mackey & Gass, 2012) with the use of thick description (Geertz, 1973). Aspects of 'dependability' aimed to be achieved through my detailed explanations of the stages of research process and I also subjected the data from the tasks and interviews to interrater reliability checking; with incidents of disagreement in coding decisions highlighted. Descriptive validity (Maxwell, 1992) was considered more important than aspects of 'neutrality' in terms of the design and aims of the present study. This is reflected in my own chosen multimodal representation and interpretation of the participants' negotiation of meaning excerpts (inclusive of routines and more extended forms of negotiation) which had been derived from their wider speaking tasks. This also involved decisions as to how best to accurately analyse and describe the participants' interpretative stance on their own use of multimodal forms of communication within negotiation and to balance this with my own interpretative observations.

As a researcher, I adopt the stance that reality is multifaceted and complex, necessitating multiple methods and data source approaches. This entailed collecting and coding sections of data from the task-based interactions but also introducing stimulated recall methods as a way to attempt to account for a learner-centred interpretation of these same excerpts of negotiation. In triangulation, the researcher exploits many different sources and methods (Creswell, 2013, p. 251). I therefore aimed to triangulate the task-based data with this secondary learner-centred data source to help to answer my questions more fully but to

also balance potential limitations and researcher bias. It has been remarked that data from learning tasks can be enhanced by listening to how learners personally interpret their own interactive opportunities and experiences (Mackey, 2002). Whilst Mackey does not specifically mention how forms of reflection and discussion can serve as learning opportunities, the area of mobile learning has more recently highlighted the importance of audio or video recording learners in order to enable forms of structured reflection. The approach encourages teachers to collaborate with learners, following their engagement with autonomous tasks from beyond the classroom (see Kukulska-Hulme et al., 2017). Researchers and teachers can enable learners to extensively talk about what they know about how they process their L2 (Gass & Mackey, 2017), however, this is rarely considered to be closely related to their deployment of gesture.

Researcher credibility and issues around trust also assume paramount importance in qualitative research in that there must be transparency in areas such as rigour of methods, ethics; and the presentation of findings in relationship to the researcher's interpretation of the data (see Creswell, 2013; Hammersley, 1992; Lincoln & Guba, 1985). In addressing philosophical aspects of research, axiology concerns itself with the issue of ethics and whether research can ever be considered 'value free.' In qualitative work notions of 'truth' and 'worth' are reflected through the researcher in terms of "how their interpretation flows from their own personal, cultural and historical experiences" (Creswell, 2013, p. 25). My years of working with language learners, my teaching philosophy, and my own personal experiences of being an adult language learner will have all influenced my interpretation of the learners' negotiation of meaning in the present study. In the case of educational domains, the research itself must always entail attention to ethics (Cohen et al., 2000). This is to protect all participants and sensitive forms of data which, in the case of the present study, were captured from mobile devices used beyond the classroom. In my role as researcher, there was also a responsibility to the participants in terms of comprehending different forms of rhetorical choices and the possible implications of these. Rhetoric raises questions as to how the researcher can most accurately represent the voices of the participants but also balance this with their own personal researcher's voice. In the case of studies which exploit methods such as multimodal analytic transcription, these rhetorical choices now entail questions as to how to visually represent the participants and their multimodal interactions; and consideration of the ethical implications of this for the participants.

Within a qualitative paradigm, the researcher needs to make transparent their worldview and theoretical stance in terms of how this will invariably shape and influence aspects of the research aims and design. It has been recently noted that paradigms within the social sciences have ensured that clear distinctions between the research *insider* and *outsider* have become somewhat problematic. It is argued that many researchers are shifting combinations of both these roles as qualitative research methodologies in education increasingly seek to encompass aspects such as inclusiveness and participation (McNess et al., 2015, p. 295).

3.3 The roles of the researcher: ethical consideration around use of mobiles

The current study proved challenging to implement in the beginning as institutions had been reluctant to take part. It appeared from feedback that one of the principal problems was the research description itself. This had been sent out to higher education institutions in this country and abroad and had explained that language learners would be engaging with peer-to-peer speaking tasks from a range of settings which lay outside the boundaries and locus of control of the educational institution. I therefore made the research decision to return to a familiar language school where I had previously taught English for several years and where task-based learning within communicative paradigms was the norm. The advantage to this was that I was known, trusted; and familiar with the overall pedagogic approach and learning ethos of this particular institution. I considered myself an ‘insider’ who was now returning as someone who was working as a researcher at the same institution. I received ongoing access to participants combined with high levels of co-operation from the principal gatekeeper and some of the teachers. This was important in that I was initially concerned that the implementation of activities based around learning with mobiles in autonomous ways could be perceived as a potential challenge to the authority of teachers and the institution itself. In the case of research around mobile learning, Pachler (2009) highlights a shifting paradigm in that there has been “a change in locus of control” in that participants own their mobile devices (p. 4).

Ethical approval for the present study was gained in February 2015 from the Human Research Ethics Committee of the Open University after useful consultation and advice from them. As a member of and presenter at the British Association of Applied Linguistics (BAAL), I followed their guidelines alongside adherence to the British Educational

Research Association (BERA) ethical guidelines for educational research. In the present study, I adopted the approach that ethical decisions are made on the basis of problems and issues applicable to specific projects (Robson, 2011). For example, during the process of conducting this study, I came to the conclusion that researching mobile learning beyond the classroom differs in certain respects to researching learners completing tasks on desktops from within the boundaries of an institution. It was therefore necessary to consult sources around ethics which pertained to the specific area of researching learning with mobiles (Kukulska-Hulme & Traxler, 2005; Vavoula et al., 2009). This was helpful in that these publications had included the sets of ethical problems which learning with mobiles in my own research context involved: “What is novel, is the evaluation of technology-enabled learning on the move, with the learner travelling while learning, or learning that spans locations and times” (Sharples, 2009, p. 20). Heath et al. (2010) also raise awareness of the practicalities and ethical issues which can arise when qualitative studies are conducted from public or informal settings such as museums.

At the beginning of this research journey, I had fully anticipated that learners participating in the study would exploit their own devices which had included the deployment of smartphones, 2-in-1 devices, and tablets to access Skype in order to complete a speaking task. The rationale behind the initial decision to enable learners to exploit their own device was that learners would assume greater autonomous control; also that they would possibly adopt some of the ideas from the study in order to pursue language learning goals via the use of their own devices in the future. Learners taking part were familiar with the process of storing their video data on their devices when communicating via video and, as a result, several asked if they could keep the data before the tasks had even taken place. It was clear that some had wanted a record of the task on their device for reflective learning purposes and other participants had simply wanted to keep a video souvenir of their learning experience; similar to how a person might store a video of a birthday party on Facebook. It occurred to me that some might decide to publicise or share the data on social media in ways which would breach confidentiality and possibly cause embarrassment to certain participants. On further consideration and in trialling how the video capturing tool Supertintin operated, a compromise was reached where it was decided that one interlocutor would use the researcher’s own tablet device, with the pre-installed tool used to record the video data internally, and the interlocutor would exploit their own device which would not have the tool installed and therefore would not capture any of the video data. This ensured

that the collection and the storage of confidential data could remain under the control of the researcher at all times and would adhere to the strict guidelines in the data protection act (1998).

An initial information letter (see APPENDIX 1) was sent out to potentially interested parties through the help of the teachers at the school with the opportunity offered to attend a forty-five minute meeting about the research. It is important to note that as a researcher I took the decision not to reveal the specific focus on gesture until the point of the stimulated recall sessions due to this possibly influencing how the participants might behave during the tasks. Therefore, they were told that their overall communication was of interest to the researcher and that the researcher would need to use visual screenshots (still images from the recorded video) of their Skype communication for educational research purposes. Potential participants were openly invited to discuss the research, to ask questions; and to raise any concerns with me that they might have had at the meeting. The participant consent form (see APPENDIX 2) aimed to ensure that learners fully understood the implications of their taking part and that they could withdraw from the research at any time. One consideration was to clearly explain to participants that their data would be published for educational research purposes and that the decision had been taken to use images of learners but that all names would be changed.

It was then necessary to contact the public establishments where the speaking tasks were to be implemented in advance and, in one case, the director of the local museum had to be consulted with the necessity for the study's aims to be explained in detail. Most establishments were supportive, for example, the local museum and gallery expressed a strong interest in how the cultural artefacts in their building could be better exploited to support visiting international students and speakers of other languages to help them to learn. Many of the café and restaurant owners were also sympathetic to the study and recounted their own entertaining stories and memories of learning English. When they understood what was involved, the owners agreed that it was acceptable to bring the participants to these settings to interact on Skype VC. These businesses naturally wanted some added value for their co-operation and we agreed that participants would contribute to the business or institution through the purchasing of some food, a coffee, buying a small souvenir etc. The cost of this was covered by the researcher and also represented a small gesture of thanks to the participants for their time and valuable contribution.

There were other obstacles to the research in that learners were residing in a foreign country away from home and sometimes had personal issues to deal with, for example, they might experience unforeseen stress surrounding problems with their accommodation. This meant that I had to listen to their personal concerns and at times try to support them through offering a more flexible research timetable. It became evident that whilst several of the participants had various problems, I was also dealing with a co-operative group of adult learners and, as a result, it became possible to renegotiate the day or even the month of tasks. This did have a considerable impact on the overall research timetable with the data collection extending over a period of nine months. As the tasks were implemented from a range of public settings, there was an occasional domino effect caused by participants altering the day of the tasks as venues such as the local museum or café might not then open on the revised day or would unexpectedly have closed early just as we arrived at the door. This happened on one occasion. When the speaking tasks were rescheduled, the stimulated recall also had to be scheduled to take place two days later due to the suggestion that learners can forget information beyond this timeframe.

There were other considerations in scheduling the stimulated recall sessions two days after the tasks. This firstly involved the necessity for myself as the researcher to gain some time to look through the data to see if and where learners had used gesture and speech within sections of negotiation before I played the video of tasks for them. The second involved concerns from the principal of the school that the group should not have to complete a speaking task and a lengthy interview within the same school day. Learners who took part in the study were all studying full-time on their English language immersion programmes in the UK for around seven hours a day and so I became aware that they became tired and, as a result, there was a conflict between my ideal research timetable and the needs and welfare of all the participants. The study ran for an extended period from January 2016 to September 2016.

3.4 Multimodality as research methodology

In her work on learning with technology, Jewitt (2006) explains the multimodal character of this type of learning in that new technologies are constantly reconfiguring what it means to learn in that students “point, gesture, gaze at the screen [...] Students learn from all the modes present on the screen and around it” (p. 7). In contrast, the field of linguistics and SLA research has generally tended to reduce a depiction of learning through spoken interaction in a second language to the linguistic ‘turn’ (Long, 1996; Mackey, 2007; Varonis & Gass, 1985) with no account taken of other modes and how these may be co-orchestrated with language use.

In terms of embarking on a multimodal research path, the scope for methodological innovation becomes apparent through studying O’Halloran and Smith’s (2011) overview of this field: “[T]here is a diversity of viewpoints and approaches that seems inherent in multimodal studies;” with the notion of diversity relating to the array of human resources which have been developed for communication and the sites and contexts in which multimodal communication occurs today (p. 10). Some of the more proceduralized methodological approaches to multimodal work include: multimodal interactional analysis (Mondada, 2014; Norris, 2004, 2011; Scollon & Scollon, 2003); a social semiotic approach to multimodal analysis (Kress & Van Leeuwen, 2006); multimodal approaches to discourse analysis (Baldry & Thibault, 2006). Multimodal research approaches have encompassed contexts which include the analysis of everyday interactions in the world (Norris, 2004; 2011); multimodal literacy (Lancaster, 2012, 2013); multimodal forms of language learning in classrooms (McCafferty & Stam, 2008); language learning scenarios using digital technologies in desktop and mobile contexts (Austin et al., 2017; Hampel, 2003; Hampel & Hauck, 2006; Hampel & Stickler, 2012; Kukulska-Hulme et al., 2017; Lamy & Flewitt, 2012; Satar, 2015); the analysis of meaning-based signs indexed in the material world (Scollon & Scollon, 2003); and the use of cameras and video-based technologies conceptualised as ‘prosthetic deictic gesturing’ within tourist spaces (Jaworski & Thurlow, 2011).

Multimodal research also acknowledges and draws on a range of other methodologies such as the deployment of conversational analysis (Schlegloff., 2007; Sefeddinipur & Gullberg, 2014; Streeck et al., 2011) and forms of analysis derived from gesture studies (McNeill, 1992) which have subsequently been implemented within a wider multimodal SLA research paradigm (Gullberg, 2011; Negueruela & Lantolf, 2008). Goodwin (2014)

focuses on the deployment of gestures and context-specific language use as embodied interaction in the material world; explicating how different semiotic resources are always partial and incomplete until they are brought together by participants. Norris (2004) draws on earlier systems of multimodal analysis as she exploits methods derived from gesture studies (Kendon, 1982; McNeill, 1992) within her methodological framework: *Multimodal interactional analysis*. In this guide, she offers a range of systematic procedures around the study of gesture and details these in a highly practical manner in ways which personally helped me to design the present study; despite not choosing to exploit CA. She clearly explains how to break gestures down according to McNeill (1992) and demonstrates a series of potential ways to represent gestural interaction through the construction of visually-based sequential transcripts which demonstrate how gesture unfolds in relation to language use but also in relationship to the wider context of the interaction. Her work successfully employs the use of video stills to depict the unfolding gestural phrase. Conversely, her approach does not offer a transcription procedure to fully explore intermodal relationships between gesture and speech as they unfold, for example, within her multimodal transcripts it is sometimes difficult to discern which part of the gesture relates to which specific words.

Whilst communication has always been multimodal, the requirement to develop a range of workable systems of analysis within multimodal methodology has become pressing due to the rapidly shifting communicative landscape around people's deployment of ubiquitous technologies. Analysing language learning interactions within digital environments from a multimodal perspective, now entails the researcher developing a skills set to understand, for example, how but also why learners collaboratively create 'modal ensembles' (Jewitt, 2011) online. These ensembles could include their use of language, gesture, gaze, proxemics, and haptics in order to learn. There are also parallel questions as to how interested language teachers can best harness these same modes in relationship to learners' use of ubiquitous forms of technology in order to support language learning in potentially new and extended ways. According to Scollon and Scollon (2011), the analysis of language is rapidly changing in that studies in the fields of language and multimodal research are now "developing along somewhat convergent lines" (p. 177). For example, there is increasing evidence in CALL research that demonstrates that learners exploit a range of modal resources to construct and achieve their contextualised L2 learning

interactions via the use of desktop VC technologies (Austin et al., 2017; Hampel, 2003; Hampel & Stickler, 2012; Lamy & Flewitt, 2012; Satar, 2015).

A multimodal theoretical stance undoubtedly raises serious methodological concerns for researchers interested in how groups of learners come to co-orchestrate a diversity of modes in their L2 online learning scenarios. Methodological challenges in the current study included aspects around the collection of multimodal data and procedures to analyse and then elucidate the intermodal interplay between the modes of gesture and speech. For example, I was interested in how to analytically transcribe the points where the learners' use of the two modes might converge around the GP (McNeill, 1992; McNeill & Duncan, 2000) in relationship to the learners' use of mobile technologies from beyond the classroom. Flewitt et al. (2014) provide one of the most comprehensive methodological overviews of the primary research aims and emergent issues within the field of multimodal data and transcription procedures. Certain issues which they raise in their chapter guided this study as they posed questions such as 1) what are acceptable as data. 2) suitability of chosen descriptive and analytic tools, and 3) what counts as research evidence in this particular realm.

3.4.1 Approach to the research

Multimodal considerations around the study of gesture and speech in a mobile learning context led to the methodological consideration as to how research within negotiation might be extended to explore and reconsider an established theory and model within second language acquisition. I was particularly interested in how the Varonis and Gass framework for the negotiation of meaning (1985) might be operationalized when the systematic analysis of gesture was introduced and integrated into learners' speech turns through a unit of analysis which could take account of the modes operating together (McNeill, 1992; McNeill & Duncan, 2000). The purpose was to elucidate the role of gesture for learners and to better comprehend how dyads orchestrated gesture and speech in interplay across their use of mobiles and an array of settings. For example, would gesture perhaps be enabled or possibly constrained by the use of mobiles beyond the classroom? Whilst some would argue that gesture is a 'non-verbal' mode which assumes little influence on language acquisition, McCafferty and Stam (2008) remark in their work on gesture within SLA that "virtually all researchers working in the field of gesture studies accept that there is a close connection between language and gesture in relationship to meaning-making" (p. 3).

As outlined in the presented literature review, it has been demonstrated within gesture studies that this mode needs to be analysed and understood in direct relationship to speech (see Kendon, 2004; McNeill, 1992, 2000; McNeill & Duncan, 2000). This led to a realisation that the area of gesture studies, with its extended history and defined systematic and analytic procedures, could offer a methodological toolkit in which to explore the negotiation of meaning via mobiles as a multimodal phenomenon in order to answer the research questions posed. It is important to note that Hampel and Stickler (2012) previously exploited Long's (1996) framework of the negotiation of meaning to investigate desktop-mediated language learning interactions from multimodal perspectives; suggesting that the deployment of SLA theory can be exploited as a way to understand how learners co-construct meaning across various modes. As Flewitt et al. (2014) note, the study of gesture and spoken language has been an area of research which has taken account of modes of communication for decades with gestures having been associated with both cognition and the multimodal organisation of people's actions.

Within multimodal research paradigms, Jewitt (2011) exploits the term 'multimodal ensemble' to indicate the interrelationships evidenced between modes as participants use these as resources in complementary but also dissonant ways (p. 301). Kendon (2004) had previously used the term 'ensemble' in order to specifically depict speakers' co-orchestration of gesture and speech. His conclusions about the multimodal nature of 'utterance' is due to findings which indicate that speakers can repeat and revise the gestural as well as the verbal component within their talk (p. 128). In terms of L2 meaning, it has been shown that McNeill's theory of the GP (see section 2.3.3 of the literature review) can be exploited to elucidate the ways in which gesture either exactly mirrors learners' speech or to reveal how speakers add information which is not present in their speech. It is also possible that the deployment of L2 gesture can result in examples of lexical searches where examples of asynchronicity occur and where gesture precedes speech in ways which challenge the inevitability of the GP (Negueruela & Lantolf, 2008).

The close relationships forged between language and other modes is also evidenced within Scollon and Scollon's (2011) comments on multimodality in relationship to their discussion of the study of language. The researchers position themselves as linguists; remarking that it is important not to confuse multimodal research aims with any side-lining of language nor to adopt the assumption that multimodality is simply 'a rephrasing' of the field of study of non-verbal forms of communication (p. 171). I adopted a gesture-speech

analytic unit (McNeill, 1992; McNeill & Duncan, 2000) which would ideally enable me to comprehensively analyse and explain how and why learners exploit gesture and speech together or in divergent ways in relationship to their use of technology and settings within a framework of the negotiation of meaning, according to Varonis and Gass (1985). How to indicate the potentially complex semantic relationships between gesture and speech within the framework of the negotiation in relationship to learners' use of mobile technologies beyond the classroom is the challenge to be met when addressing the following research questions, as first introduced in chapter 2:

- 1) *What is the role of gesture in the negotiation of meaning when learners are communicating via Skype VC accessed on mobiles from beyond the language classroom?*
 - a) *How are gestures exploited and responded to between learners during their negotiation of meaning?*
 - b) *What are the affordances and constraints of mobile technologies, and the settings they enable, in relationship to learners' deployment of gesture during their negotiation of meaning?*
- 2) *To what extent can a supported focus on gesture and language use help learners to understand aspects of their multimodal communication during the negotiation of meaning?*

Adopting a multimodal approach towards language use and meaning, implies that modes which include, but also move beyond language, must be considered worthy of fine-grained analysis and can then become the subject of interpretation by the researcher. Flewitt (2006) suggests possible reasons for the neglect of modes other than language which include problems with a lack of systematicity of analysis and the misguided assumption that these modes are somehow superfluous to learning processes.

In the present study, there was a perceived need to explore methodological procedures from gesture studies in an in-depth manner to begin to interconnect the mode of gesture and relevant features of learners' language use via mobile technology as a way to understand the role gesture may play. I realised that transcription systems from face-to-face contexts would also have to be adapted to account for the manner in which learners might exploit VC on mobiles. It has been suggested that modes are not random entities but instead form complex semiotic systems with their own rules and regularities attached to them (Kress & Van Leeuwen, 2001). However, to exclusively view modes as disconnected systems implies that they are removed from their inevitable social contexts of use. Benson

(2011) has posited that settings beyond the classroom offer opportunities to learn in “social spaces” with learners gaining access to “affordances for and constraints on language learning” (p. 13). Moreover, the negotiation of meaning in language learning contexts frequently involves forms of collaborative task design which aim to support learners to generate social interaction in ways which are considered beneficial to their acquisition of language (Ellis, 2003). Context-sensitive task design may assume even greater importance when learners are requested to leave the traditional classroom and to connect with one another from a range of settings which are located in the world beyond the classroom.

3.4.2 The coding of the multimodal negotiation of meaning

The negotiation of meaning operates along the premise that learners’ acquisition of language is supported as they share information together through being prompted to modify, restate, clarify, and confirm aspects of this to try to establish mutual understanding. In the present study, I aimed to exploit this framework from a multimodal perspective as a way to examine the gestures and speech which occurred together within the language learners’ negotiation of meaning when enabled via mobiles from beyond the classroom.

As discussed previously (see chapter 2 section 2.7), a routine is proceduralized by Varonis and Gass (1985) as follows:

Non-understanding routines are operationally defined as: those exchanges in which there is some overt indication that understanding between participants has not been complete. *The sine qua non* of a non-understanding routine is that within the exchange there are embeddings of one or more clarifications. These can stem from an exchange in which there was a misunderstanding, no understanding, or incomplete understanding.

(p. 73)

As first described in chapter two, a routine falls into two broad categories which consist of an initial trigger and then a resolution phase which entails an indicator, resolution, and reaction to resolution. The trigger (T) is the designated part of the speaker’s utterance which is first identified as causing the initial communicative problem. A negotiation trigger can be identified and coded by taking account of the follow up turn by an interlocutor. In encountering a trigger from the first speaker, the interlocutor then has the

choice to either ignore and bypass the communicative issue or to react to it in some way and begin to solve the problem of ‘non-understanding.’ For example, an interlocutor who makes the choice to draw attention to an issue may respond to another learner’s trigger with a comment which indicates that they have not fully understood something and perhaps wish to clarify the original source of information. The second phase of the routine is identified and coded as a resolution phase which can consist of up to three parts. These parts constitute an indicator (I), demonstrating that the learner has noticed that there is a problem and, as a result, draws attention to the need for interactional modification from their interlocutor through a variety of strategies, for example, they could use a repetition of the original word causing the problem or exploit a more overt expression such as: “sorry?” or “I don’t understand.” The response (R) by the original speaker who created the trigger is then coded as an acknowledgment of the difficulty and serves as an identified attempt by learners to try to repair and solve the issue. The RR (Reaction to response) is optional and is often indicated as an affirmative reaction such as: “yeah” or “ok,” following the previous response and, according to Varonis and Gass (1985), this move indicates that understanding between speakers is now complete.

Whilst the researcher’s coding decisions may superficially indicate that a speaker has acknowledged the end of the negotiation with an affirmative response such as “yeah,” I considered that it could not be assumed that full understanding had necessarily been achieved between a dyad without asking them to comment. As a result of this perceived limitation, the framework calls into question how observations alone can be exploited as a method to illustrate that issues of non-understanding have been resolved, particularly within complex digital learning environments. I considered that learners themselves should be afforded further opportunities to talk about aspects of their communication and to be enabled to elaborate on what level of understanding they felt had been achieved between themselves and an interlocutor. Varonis and Gass also cite another coding category known as CC (Comprehension checks). These checks can occur in any of the four places (trigger, indicator, response, reaction to the response) and demonstrate that learners are trying to further check or confirm levels of understanding with an interlocutor when they perceive that this may be challenged.

As demonstrated in the section on the negotiation of meaning in the previous literature review, Varonis and Gass (1985) also identify examples of exchanges between non-native interlocutors which include complex and extended examples of negotiation, referred to by

the authors as ‘extended.’ This type of embedded sequence within negotiation can contain “multiple layers of trigger-resolution sequences” (Varonis and Gass, pp. 78-79). This is where the previously shown linear patterns of trigger, indicator, resolution and reaction to resolution appear to become disrupted (also see Varonis and Gass, 1985, p. 79). For this reason, I will adopt the term negotiation excerpts in order to signify both simple routines and extended examples of negotiation observed within the data which I collected.

3.5 Introduction to the data collection and analysis

Within this section, I outline the data collection and analysis procedures which I implemented and exploited in order to elucidate the role of gesture within the negotiation of meaning; including a focus on the experience of negotiation as interpreted by the learners themselves. I began the study with the concept of a pedagogic task based on techniques to stimulate and elicit negotiation of meaning within the learner dyads who were all operating within communicative scenarios which took place beyond the classroom. I video-recorded the learner dyads as they completed specifically-designed speaking tasks during the first data collection phase of the research. The subsequent approach to the stimulated recall data collection phase involved my deployment of combinations of video stimuli from the tasks based around the learners’ negotiation and the use of semi-structured interview questions. These were designed to encourage participants to verbalize aspects of their gestures and speech in relationship to their use of mobile technology and setting (see APPENDIX 4). In Table 4, I present an overview of the data collection and analysis methods and their purpose in relation to the research questions.

Table 4: An overview of the data collection and analysis procedures and how they relate to the research questions

Research questions	Data collection	Analysis
<p>1) What is the role of gesture in the negotiation of meaning when learners are communicating via Skype VC accessed on mobiles from beyond the language classroom?</p> <p>a) How are gestures exploited and responded to by learners during their negotiation of meaning?</p> <p>b) What are the affordances and constraints of mobile technologies, and the settings they enable, in relationship to learners' deployment of gesture during their negotiation of meaning?</p>	<p>Video-capturing learners' task-based interactions via Skype VC across mobiles from beyond the classroom.</p> <p>Recording of learners' reflections. Use of stimulated recall methods exploiting semi-structured interview.</p>	<p>Multimodal/descriptive approach. Gestures and speech examined in relationship to wider aspects of context via multimodal analytic transcription.</p> <p>Descriptive/ interpretative approach. Qualitative content analysis approach in order to establish themes and draw inferences. Aim to view the data from different vantage points and to triangulate the data.</p> <p>Use of multimodal transcripts combined with qualitative content analysis approach.</p>
<p>2) To what extent can a supported focus on gesture and language use help learners to understand aspects of their multimodal communication during the negotiation of meaning?</p>	<p>Stimulated recall methods, exploiting semi-structured interviews.</p>	<p>Use of multimodal transcripts combined with qualitative content analysis approach.</p>

3.5.1 The pilot study

To explore the subject of gesture and speech operating together within a framework of the negotiation of meaning via learners' use of Skype VC on mobile devices, it was first necessary to conduct a pilot study. At the time of conducting the research, there had been little available information as to how the negotiation of meaning might operate across Skype VC when accessed by learners from mobile devices deployed beyond the classroom; and insufficient information as to the potential role gestures might play for learners in online interactions within SLA. Firstly, I considered video-capturing the speaking tasks from the mobile device itself to see and hear how learners were employing the multimodal resources of speech and gesture in close orchestration. It was also important to determine the impact of mediating aspects of technology such as the size and boundaries of the screen and to understand gestures in relationship to wider contextual factors such as the learners' interaction with material objects. Secondly, I wanted to explore a learner-centred perspective on negotiation and to understand ways to help to support learners to reflect on gesture and speech within their multimodal communication. I had to therefore consider the suitability of stimulated recall as an instrument for multimodal research purposes. As Dörnyei advises: "always pilot your research instruments and procedures before launching your project" (2007, p. 75).

In the present study, the pilot was small-scale (one learner dyad) but this played an important role in determining the future approach for the main study. The pilot firstly allowed me to trial the task type to understand if it would elicit negotiation of meaning between two learners interacting across Skype within two different geographical settings. In the pilot, the settings encompassed a nearby stately home (a historical building) and a garden. The task I had designed for this interaction contained photos of interesting objects situated in the settings themselves which were included in two different paper-based task sheets for each learner with the aim to encourage them to collaborate to find and 'piece together' the missing sources of information as they negotiated for meaning. I had photographed a series of artefacts on my mobile phone in advance of the interaction and designed a task around them in ways which would ideally prompt the learners to attempt to locate a series of different objects situated within their individual settings.

In completing the task, both learners appeared to grasp the general concept of the task but mostly ignored the detailed instructions. However, the interaction did elicit excerpts of negotiation (both routines and more extended examples). Another reason to set up the pilot was to test Skype VC from mobile devices and to trial the technology itself; including my deployment of a video-based capturing tool which needed to function on a mobile device. It is never possible to guarantee that Skype will work and my research set-up required that the video-capturing app would simultaneously record both learners' video streams as they walked around and explored their different settings in order to negotiate for meaning around the designated objects.

The most crucial elements of the pilot study aimed to enable me to begin to understand how gesture and speech might operate together and in relationship to the dyad's use of Skype VC on two devices (a tablet and a 2-in-1 device) as the learners interacted with both the technology and their individualised settings. Two learners took part in the pilot and the data was then integrated into the wider study. The learners produced a variety of hand gestures with one exploiting this mode more than their interlocutor. Due to the task design, the device and camera were also regularly pointed at artefacts in the setting by one learner in particular and this appeared to operate in a similar manner to the notion of the device itself acting as 'prosthetic deictic gesture' (see Jaworski & Thurlow, 2011). At this stage, I did not understand this aspect of gesture and first began to consider how I would proceduralize the action of pointing with a device and camera in relationship to the existing analytic system which had been designed for the analysis of gesture and speech formed with the hands alone (McNeill, 1992; McNeill & Duncan, 2000).

I also needed to trial the instrument of stimulated recall in terms of accessing learners' thought processes around gesture and speech at the time of the task. Having worked with the learners following the task through introducing stimulated recall, I became more aware of the advantages and constraints of asking language learners to talk about gesture through a combination of video stimuli and interview questions. I made several mistakes in the pilot, such as the unrealistic expectation that learners would be able to freely talk about gesture without replays of the video extracts or more direct lines of questioning and prompting when they struggled. Furthermore, I realised that questions which elicited extended and elaborate responses from the learners appeared to be generated via semi-structured approaches which had the flexibility to allow me to then probe the learner further for increased amounts of information. At times the line of questioning transformed

instead into a more spontaneous dialogue between researcher and learner. One learner in particular, who had taken part in the pilot, demonstrated a high level of awareness and was able to comprehensively link their gestures to specific examples of their language use in ways which surprised me and which I wanted to pursue further. The pilot study also highlighted that the stimulated recall sessions might offer possibilities for learners to reflect on aspects of their gesture and language use from multimodal perspectives as they achieved their negotiation of meaning.

3.5.2 A collaborative task designed around learners' use of VC on mobiles

Connections between the development of learners' multimodal meanings in relationship to their interaction with artefacts situated in the world are reflected in the context-sensitive task design implemented within the present study. My approach to the task design ideally aimed to elicit and stimulate negotiation processes as learners collaboratively interacted via technology but also via some everyday objects, as well as more exotic objects, which were authentically situated within the settings around them. Through the exploitation of strategic task design and guidance, mobile technologies suggest new possibilities for learners to communicatively interact out in the world. Each Skype VC session between the learner dyads taking part in the study was therefore enabled through the design and implementation of 'a pedagogic task' (Ellis, 2003; Pica et al., 1993; Skehan, 1998). Information gaps (see Skehan, 1998; Yanguas, 2012) were strategically engineered as a way to prompt certain types of language use and, therefore, the approach cannot be claimed have generated entirely naturalistic forms of SLA. The task design allowed me to explore questions about gesture and speech within the negotiation of meaning with the additional aim to understand the affordances and constraints of mobile technologies and the types of settings which they may enable for language learners today. The captured task data would also ideally support learners to later reflect on what they had been attempting to achieve during the process of their negotiation.

Information gap and jigsaw approaches to tasks are prevalent within design for digital language learning environments (Hampel, 2006; Thomas & Reinders, 2012; Yanguas, 2010). Hampel (2006) has created a useful framework in which to design and evaluate digital tasks from perspectives which encompass the following principles: theories in SLA, constructivist stance on learning, multimodal elements of learner communication, and

identification of available affordances for learners. In terms of the task design in this study, features associated with the ‘Interaction Hypothesis’ were implemented in my chosen re-interpretation of Pica et al.’s (1993) framework for spoken learning tasks.

Pica et al.’s (1993) task analysis framework is as follows:

- 1) Each interactant should hold a different portion of the information which must be exchanged or manipulated in order to reach a task outcome.
- 2) Interactants should be required to request and supply this information to each other.
- 3) Interactants should have the same or convergent goal.

This task design had been previously used in face-to-face and CMC communicative learning contexts (Skehan, 1998; Yanguas, 2010), however, it had to be reconfigured in order to stimulate negotiation of meaning from real-world settings via mobile technologies. In order for the tasks to work effectively, learners had to be placed in settings which they had never visited before so that they would then be reliant on sources of information conveyed by their interlocutor in order to achieve the task goals. At the beginning of the information gap task design process, I visited each of the locations which the participants had suggested and then searched for affordances around which to design a set of contextualised speaking tasks. Identified affordances included stimulating and curious objects which included models and photos of animals, people, and places (see Figure 4). With the owners’ prior permission, I took a series of photos on my phone from within a series of settings located near the school.



(Image redacted)

Figure 4: Artefacts for task design via mobiles to stimulate the negotiation of meaning

The first image shows a portion of a much larger photo of Taxime Square which was located on a large wall within a local Turkish restaurant. The idea for this section of the task was that one learner would have this image placed in their task sheet and would be prompted by the task instructions to describe the scene and encourage their interlocutor, situated in the Turkish restaurant, to then locate the correct photo with their mobile device via the process of negotiation. The second image represents a model depicting a ubiquitous scene from the film ‘The Shining’ which was located in a bohemian café. The first learner describes this and encourages their interlocutor, situated in the café itself, to locate the authentic object with their mobile device as they negotiate for meaning to try to find the object and then continue to negotiate around features of the object itself. The café space was particularly large and, as a result, finding objects was likely to take some time.

(APPENDIX 3 shows some examples from the task sheets for the learners)

3.5.3 Other possible methodological approaches

It is important to note that this research could have been conducted in various different ways had I employed the use of other methodological procedures designed to explore the learners’ interactions via VC. These approaches could have included conversation analysis (CA) (ten Have, 2007; Schegloff, 2007); grounded theory (Creswell, 2013; Glaser & Strauss, 2017); ethnography (Dicks et al., 2011; Hammersley & Atkinson, 2007; Lillis, 2008) and other methodologies within the field of multimodality such as *geosemiotics* (Scollon & Scollon, 2003). The deployment of a CA approach, rather than the chosen Varonis and Gass framework, would have implied that the talk within the VC interactions was seen as a naturally-occurring phenomenon with no pre-determined theory. A CA approach could have illustrated the sequential organization of the learners’ interaction from a different perspective, for example, through transcribing considerable levels of detail which could have highlighted adjacency pairs, pauses, overlapping speech, and intonation patterns (Schegloff, 2007) within the final transcripts. After careful evaluation, these levels of detail would have portrayed the interactions differently but might also have left very little room to transcribe the gestural phrase. A CA approach could have illustrated relationships between gesture and speech but did not offer a solution as to the direct integration of the two modes in terms of where the meaning-based gestural stroke coincided with specific words; as in the identification of the GP (McNeill, 1992). Moreover, a multimodal reconfiguration of the SLA theoretical framework, when analysed to take account of gesture and speech, was considered of greater interest than the capturing

of microscopic levels of talk. It has been noted that methods such as CA are not normally combined with interviews (Mackey & Gass, 2012), however, there have been strong links forged between methods within gesture analysis and CA (Mondada, 2014; Streeck et al., 2011) with observations that CA publications which incorporate information about participants' bodily conduct have increased dramatically in recent years (Mortensen, 2012). Norris (2004) also adopts a CA approach within her multimodal framework, however, it was her rigorous approach to the analysis and transcription of gesture as a mode within communicative interactions situated within the world which was of principal interest within the present study.

Had I chosen to pursue a grounded theory approach, I could have generated a theory which would have emerged from the data itself in order to explain a particular concept. Grounded theory would have entailed a more inductive approach to research where I would not have known what patterns I was probably going to encounter within the task-based data. In the case of the present research, I instead chose to select a specific theory in advance and had a clear idea which questions, based around a multimodal version of this theory, I wished to answer. Furthermore, the pursuit of research into gesture within SLA is still an emergent field and I considered that it was vital to build on existing theoretical findings and analytic frameworks in order to understand what learners were trying to achieve within their multimodal mobile interactions beyond the classroom. Exploring an established theory within SLA from a multimodal perspective was also related to my interest in linking my findings to practical aspects of pedagogy such as task design, the affordances and constraints of the technology; and elucidation as to the relevance of notions such as setting in relationship to the deployment of gesture within learning tasks.

My choice of methodology influenced earlier decisions as to why I had chosen to design a pedagogic task based around the theory of the negotiation of meaning. For example, I could have instead asked the learners to meet one another online and spontaneously talk or perhaps discuss a much more general topic. Conversely, the latter phases of the analysis involved working with raw interview data and this was approached from both deductive and inductive perspectives. For example, it was considered that some codes would be drawn from the research questions and previous research but that other categories would be allowed to emerge from the interview data itself.

There have also been innovative suggestions that it may be possible to position research in a realm where multimodality operates in conjunction with more established methodologies such as ethnography. Dicks et al. (2011) discuss the idea of a ‘third space’ which could enable researchers to work at the intersection between multimodality and ethnography as the two fields “enrich the insights of each other” (p. 232). The authors highlight the area of childhood literacy studies as entailing the study of meaning (semiotics) alongside a focus on the need to address social context (ethnography) (p. 229). Whilst certain methods associated with ethnography such as observations and interviews (see Lillis, 2008) were deployed in the present study, ethnography involves the researcher’s sustained engagement with a site (Hammersley, 1992) and does not normally entail the exploitation of a pre-defined theory within SLA. However, finding an intersection where disparate methodologies are seen to converge may provide a possible solution to some of the perceived limitations of any particular methodological stance; including studies within SLA. The field of multimodality demonstrates considerable levels of diversity as researchers choose to focus on systems, and to isolate incidents of semiosis, or to instead adopt a more contextualised viewpoint as to how multimodal interaction occurs. The latter approach suggests that a research focus on people’s deployment of multimodal systems of communication does not exclude an analysis of the wider contexts or ‘ecologies’ in which they occur (see Goodwin, 2003; Gullberg, 2011; Keating & Sunakawa, 2011).

After a further review of other methods within the area of multimodal research, the framework of *geosemiotics* (Scollon and Scollon, 2003) appeared to be able to accommodate the analysis and transcription of learners’ use of gestures and speech within the ‘interaction order;’ and potentially enabled their analysis within the wider notion of ‘place.’ However, if I had taken the choice to implement this alternative multimodal methodological framework, the subsequent analysis and transcription phases would have resulted in a quite different interpretation and presentation of the learners’ interactions and would have influenced the overall research outcomes. The framework of *geosemiotics* was originally designed to analyse face-to-face contexts but has been demonstrated to have been flexibly and effectively utilized for the multimodal analysis of learners’ interactions via their use of desktop technologies (see Austin et al., 2017; Lamy & Flewitt, 2012). Scollon and Scollon theorise how the human body is indexed within the world as part of the ‘interaction order’ category (pp. 45-46) and, as a result, it was considered that gestures and speech could be analysed within this particular ‘order.’ However, I could not

comprehend how the category of ‘place’ could so easily be separated from learners’ interaction with one another and their deployment of gesture and speech from beyond the classroom. Conversely, due to the specific research context and task design situated within MALL, it was considered that aspects relating to ‘place’ would continuously constitute part of the ‘interaction order’ as these were interwoven by participants in ways which would confusingly overlap and imply boundaries which perhaps could only be artificially drawn. This research dilemma is complex and represents a considerable challenge which relates to how “we define multimodality itself” (Dicks et al., 2011, p. 228).

3.6 The research study

In this section, I first outline the details of the participants and the settings from which the tasks took place. Following this, I describe the data collection and procedures which were used; including tables which outline the types of hardware and software which made the data collection possible. Next, I outline the implementation of the videoconferencing sessions via mobile devices. In the final part of this section, I describe the data collection and procedures based around the stimulated recall sessions with the learners.

3.6.1 Participants and settings

Data for this study were captured (which includes the pilot data) and collected from ten volunteer adult language learners who completed speaking tasks from ten informal settings beyond the classroom. Nine female and one male participant took part who were all between the ages of 22-27. They came from the countries of Columbia, Hungary, Italy, Kazakhstan, Poland, Spain, Switzerland, Thailand, and Turkey (see Table 5). At the time of the study, the participants were all studying on formal English language programmes at a private language school in a city on the South coast of the UK and had been evaluated at B1 and B2 level by this institution according to the Common European Framework for Languages. The rationale for using participants from a wide range of different cultures and backgrounds is reflected in suggestions by Varonis and Gass (1985) around the increased opportunities for learners from differing backgrounds to negotiate for meaning.

Participants were told during a meeting about the research held at the school that if they wanted to take part they would be asked to participate in a shared speaking task via Skype, accessed on a mobile device, with a peer from a different linguistic and cultural background. During this first meeting, the volunteers were paired with someone whom

they had not met before. This was possible as this was a school with a constantly shifting population of students and a wide range of classes at Intermediate level. It was made clear that one half of the dyad would use the researcher’s tablet but that the other half would be offered the opportunity to select their own device. Participants attending the meeting were invited and encouraged to put forward their favourite spaces or settings to the researcher; provided they were all located in close proximity to their language school. The suggested settings had to be near their school building due to practical considerations such as travel costs, the amount of time required to set up the learners and technology, and time restrictions regarding the closing of some public buildings for the evening at around five thirty to six o’clock. The learners would have an opportunity to order refreshments (with the costs covered by the researcher) and this also ensured that after an entire day at school, they had something to eat and drink if they needed this.

Table 5: Participants and settings

Dyad 1 (Pilot)	Dyad 2:	Dyad 3:	Dyad 4:	Dyad 5:
Fay (F) (Kazakh, 27) Andrea (F) (Swiss, 23)	Nadia (F) (Italian, 26) Simone (F) (Spanish, 22)	Paul (M) (Columbian,22) Lily (F)(Swiss,26)	Angela (F) (Turkish, 27) Bobbi (F) (Thai, 23)	Jane (F) (Hungarian, 24) Sara (F) (Polish, 27)
Fay: public gardens Andrea: stately home	Nadia: museum Simone: café	Paul: café Lily: café	Angela: restaurant Bobbi: gallery	Jane: café Sara: gallery

The tasks were designed to reflect the potential levels of fluency and relative confidence normally associated with learners at intermediate levels studying on programme in the UK. It was a policy at the school to actively encourage learners to verbally communicate as much as possible and this participant group would have been familiar with task-based language learning practices from the wider communicative approach employed by their everyday language teachers. The study could not have been implemented with participants at a beginner or elementary level of language, given the task design, and the demands and degree of fluency required to share and negotiate around artefacts within two different settings via Skype VC on mobile devices.

The research plan involved myself as the researcher setting up one learner with the mobile technology and task sheet and then moving across to the second location to repeat the same process again. It had been ensured before the task that all the devices had the Skype app installed; if that was not the case I would help learners to download this if required.

Considerable research time was spent moving across both locations to ensure that participants felt adequately supported and that there were no emerging problems which required immediate intervention by myself. If there was a significant problem, for example, a location unexpectedly deciding to close when the researcher was elsewhere, then I could be phoned on a mobile and take immediate action.

The implementation of the first phase of the research entailed participants working together in mixed nationality dyads in order to complete one speaking task each on Skype VC which lasted up to a restricted maximum of one hour. One interlocutor in the dyad used the researcher's device (a Windows tablet) which had the video-capturing tool Supertintin pre-installed to internally video-capture the screen. Whilst often functioning successfully, it also became apparent that, due to aspects such as slow connectivity caused by the wi-fi, the video-capturing tool could sometimes stall and fail to capture some portion of the interaction.

To understand how learners pointed to referents in their setting with the device acting as prosthesis, I had wanted to set up an external video camera in some of the locations. This became highly problematic as most establishments did not want a video camera set up in any formal sense, as for example, other people might be inadvertently filmed as they dined. I also realised that I could not simultaneously check and monitor two devices or two cameras in two separate locations without another researcher being present. I had not wanted to follow the participants with a roaming camera as this would have entailed the learner possibly feeling inhibited and might have influenced their behaviour. As a result, only a small section of film which focussed on one learner from an external perspective was achieved. The information gap speaking tasks were followed by eight individualised stimulated recall sessions (one learner fell ill and could not reschedule and another went home unexpectedly) which took place two days after the initial task back in a quiet classroom space within the school building and lasted for up to one hour each. Whilst it is recommended that the scheduling of the stimulated recall sessions occurs as close to the actual event as possible (Gass & Mackey, 2017) the earliest the research timetable could

offer was always two days after the task. The following tables (0& 7) illustrate the use of different technologies in terms of the phases of research.

Table 6: Use of technology employed in the first data collection phase

Devices for the tasks	Platform	Data-capturing tools
<p>Researcher's tablet:</p> <ul style="list-style-type: none"> • Window Surface tablet: Screen size 10.6 inches and weight: 1.5 lbs <p>Participants' devices:</p> <ul style="list-style-type: none"> • Samsung Galaxy tablet: Screen size 10.1 inches and weight: 1.23 lbs • Samsung Galaxy tablet: Screen size 7 inches and weight: 0.76 lbs • HP 2- in-1 tablet and laptop device: Screen size 10 inches and weight 2.82 lbs • iPad pro: Screen size 10.5 inches and weight: 1.57 lbs • iPhone 6: Screen size 4.7 inches and weight 0.4233 lbs 	<p>Web 2.0 Skype app: for download on mobile devices. Freely available. User name and password required.</p>	<p>Supertin recorder: Download cost of \$30 from website. Requirement to activate link via an email. Originally designed for use with <i>Microsoft windows</i> systems and could be downloaded onto researcher's tablet. Currently incompatible with other systems such as <i>Android</i> and <i>Apple</i> but can capture <i>Skype</i> on these mobile devices from a <i>Windows</i> tablet. Use of external video camera employed to record one of the learner's interaction.</p>

Table 7: Use of technology employed in the second data collection phase

Stimulated recall	Devices	Purpose
Implementation of video recordings for the purpose of stimulated recall in order to support learners to understand their multimodal communication.	Transfer the <i>Skype</i> recordings of the tasks from the researcher's tablet onto a USB and then onto a larger laptop to enable their comprehensive review back in a quiet classroom space within the institution.	A video recorder is pointed at the laptop screen whilst playing the video of the task during the stimulated recall interviews. This is in order to film the laptop screen as the gestures occurred, allowing learners to reflect on and talk about these. The approach is in order to recapture the video clips of the relevant sections in order for the researcher to interconnect these to what the learners are saying about specific aspects of their multimodal communication during their interviews.

3.6.2 Introduction to data collection instruments and procedures

In this section, I will describe the data collection instruments as illustrated in previously shown Tables 4 -7. The first instrument for data collection purposes consisted of 'visual-audio video' (Heath et al., 2010) of the mobile VC sessions to represent the collaborative dyadic interactions in order to enable multimodal analysis of these. I discuss the specific problems which I experienced regarding visual data collection from the learning tasks with the aim of focussing on the mode of gesture in considerable detail within the settings of learners' use of VC on mobiles from beyond the classroom. From the video recordings of these earlier tasks, I had to first locate some excerpts of multimodal negotiation which were observed from within the wider captured data for the purposes of enabling the second phase of data collection based around the stimulated recall interviews.

3.6.3 The videoconferencing sessions on mobiles

The VC sessions on mobiles consisted of the first phase of the data collection *in situ*. There is no single prescription for what data collection instruments to use but the issue here is of ‘fitness for purpose’ (Cohen et al., 2000, p. 146). In the present study, video-recordings were considered a gateway to enable the subsequent multimodal analysis of gesture and speech. It has been noted that qualitative research can draw on multiple methods and sources to achieve triangulation, to increase validity, and to offer interpretative perspectives on the phenomenon under investigation (Mackey & Gass, 2012, p. 186). Analysis of video data and multimodal transcription procedures are analytic methods which can be employed across a wide range of multimodal research contexts (Flewitt et al., 2014; Jewitt, 2011; Mavers, 2012; Norris, 2004).

At the point of the task, I was unfortunately unable to position myself in both learners’ settings simultaneously and, as a result, made the decision to spend the first twenty minutes of the task in the first location and then walk over (normally five to ten minutes) to the second location. As the learners regularly walked out of sight and sound with their devices, I often had little indication as to how they were using gesture and speech; although I could watch the learner for short periods of time as some placed the device on a nearby surface in order to communicate or they stood near me as they pointed at objects with the device. I would smile at them to reassure them that I was co-present but never intervened in the task unless the technology failed or if there was a problem with the venue such as an owner deciding to close earlier than usual. Both these incidents occurred at various points in the research.

It would have been impossible to deconstruct a task-based interaction which involved analysing gestures and speech within the negotiation of meaning without first video-capturing the VC sessions in which the learners had taken part. It has been suggested that the captured video of a learning activity can act as a video-based ‘artefact’ (Henderson et al., 2010; Kukulska-Hulme et al., 2017). In the case of collecting and analysing data of L2 gestures, video recording of learners allows for meaningful, but also fleeting movements, to become the subject of intense hours of multimodal scrutiny for the researcher. The audio channel operating in conjunction with the visual images of gestures enable the researcher to begin to link the modes of speech and gesture together to attempt to better understand their holistic meaning or possible dissonance within negotiation. These video-based

'artefacts' also afford data examples around which learners themselves can reflect as a way to understand how they shape and reshape modes such as gesture and spoken language. In sharing the data with the students after the task, through stimulated recall, they are recognised and included as stakeholders within a research project and given an opportunity to revisit, interpret, and talk further about their interactions from beyond the classroom. It was considered that stimulated recall, questioning, and forms of discussion might be exploited as further resources designed to help learners to comprehend the interplay between gesture and speech.

It is important to consider even at the data collection phase that multimodal video-captured interactions will be transposed once again via the researcher's subsequent choices of modes designed to re-represent the learners' use of the original modes. These choices can entail combinations of written and image-based systems which are exploited in order to illustrate speech and gesture in a series of multimodal transcripts. These research choices will also result in the interaction being re-represented in a certain way, based on the researcher's focus of inquiry and their interpretation of the interaction (Flewitt et al., 2014; Mavers, 2012). It was considered essential to the data collection process to capture the VC sessions via the mobile screen from inside a tablet device to understand the gestures but also to transcribe these in relationship to the possible affordances of the technology and the physical settings.

Throughout the process of collecting the data, I found that the Skype screens of both interlocutors can be recorded simultaneously through installing the video-capturing software on just one device. In the case of the analysis of gestures, further decisions around which video angle or 'view' to capture the data from became particularly complex because the video-capturing tool Supertintin offered a variety of differing visual perspectives on the learners' original interaction. These 'views' had to be considered in relationship to the data collection aims with attention paid to the subsequent analysis of the mode of gesture with speech; bearing in mind that these images would form a large portion of the final phase of analytic transcription. There were advantages and constraints to all the decisions which I made as to how to collect the data in relationship to the learners' exploitation of mobile technologies and settings throughout this first data collection phase of the study.

When two learners communicate via Skype VC, one learner appears in the larger window whilst their interlocutor appears in a smaller version of this window. I discovered that when the Skype screen is captured exactly as it appears for the learners at the time of the

task, there can be problems for a researcher in terms of the ability to clearly discern the gestures occurring in the Skype window of the learner which appears as a smaller space on the screen. The smaller size of this window, combined with features such as a blurred image of learners' moving hands, can prevent the researcher from seeing the gestures clearly enough to analyse them fully. Conversely, if the software setting in the video capturing tool is changed from the default setting to a 'side-by-side view' then both learners' gestures will visually appear to the researcher across two Skype windows of equal size in ways which may enable the researcher to analyse these fleeting movements more easily. The prosthetic deictic action with the mobile device itself would result in a series of images of the designated object being simultaneously conveyed to an interlocutor situated in the other setting via their Skype screen.

One of the aims of the data collection with the intention to capture learners' gestural interactions around objects or artefacts within a public space, as they simultaneously talk, is to understand how the captured video of the VC interaction can later enable or constrain the researcher to subsequently explore modal 'interdependencies.' Heath et al. (2010) discuss the methodological considerations involved in the exploitation of video recordings of participants when communicating from a range of public settings within qualitative research in the field of multimodality as follows:

Firstly, we can begin to consider the ways in which participants themselves invoke features of the physical environment within the course of activity and thereby render a particular object(s), artefact(s) noticeable. [...] The mutual 'encounter' with the objects, their looking together and being able to talk about the 'same things,' is systematically accomplished in and through their interaction [...] [R]ather than consider the conversation between the participants as somehow independent of the environment in which it occurs, we can explore how talk and bodily conduct not only enable certain things to be seen and seen in certain ways, but provide a resource for discussion. Video enables us to explore these interdependencies and how talk [...] can be embodied within, and dependent on, the participants' orientation to and engagement with features of the immediate environment. (p. 92)

In the present study, the prosthetic deictic gestures with the mobile device itself, would result in a series of images of the designated object being simultaneously conveyed to an interlocutor situated in the other setting via their Skype screen.

I next illustrate the differing perspectives on the learners' gestural interactions via a table of differing 'views,' adopted in the data collection phase (see Table 8).

Table 8: Possibilities for capturing gestures when learners use *Skype* VC on mobiles


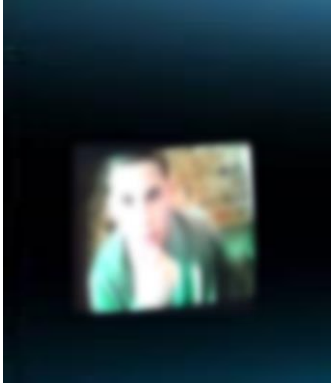
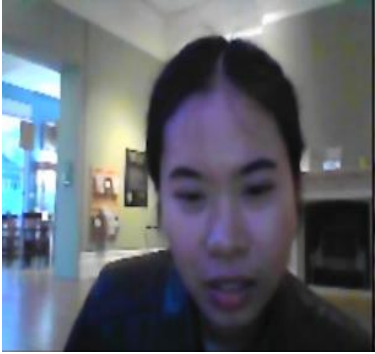




		<p>The ‘<i>picture-in-picture</i>’ setting on Supertintin. (1)</p>
	<div data-bbox="751 1059 1099 1126" style="border: 1px solid black; padding: 5px; text-align: center;">Image redacted</div>	<p>The ‘<i>side-by-side</i>’ setting on Supertintin. (2)</p>
		<p>The ‘<i>side-by-side</i>’ setting on Supertintin. (2)</p>
		<p>Externally captured image (3)</p>

Image (1) shows the capturing tool Supertintin for VC interactions on mobile devices operating in the ‘picture-in-picture’ or ‘default setting.’ This view enables the researcher to observe gestures within the Skype screen as it appears on one of the mobile devices at the time of the interaction. This ensures that there are a set of clearly observable gestures in the main Skype window but it also illustrates the problem of distinguishing gestures which the second learner uses within the smaller Skype window. The second window is much smaller and is surrounded by a wider area of black on mobiles. When images from the smaller window are enlarged, the gestures can be distinguished but the overall resolution quality can be compromised.

The next set of images (2) shows the capturing tool Supertintin for VC interactions on mobile devices operating in the ‘side-by-side setting.’ This enables the researcher to observe images of gestures formed with the hands with screens of equal size; making it easier to access the images of the gestures in higher resolution and enabling the analysis of the shared communicative nature of gestures in terms of whether learners noticed these and responded to them. The approach also more clearly conveys the visual results of learners exploiting the device to form prosthetic deictic gestures as they attempt to point at relevant objects in their environment.

The final image (3) exemplifies the researcher’s use of an external static camera to further analyse how the learner gestures. There are certain disadvantages to this approach in that learners constantly move around a space and frequently disappear from view. Other camera angle decisions might include trying to follow the learners with a roaming camera in the researcher’s hand. However, the approach could result in learners feeling self-conscious and, due to the limitations of the present study, both these approaches became unworkable.

3.6.4 Stimulated recall interviews

The second phase of the data collection consisted of stimulated recall methods (Ericsson & Simon, 1996; Gass & Mackey, 2017; Sime, 2008). This instrument was used to explore the learners’ own perspective on their use of multimodal forms of communication but also as a way to help them to learn from opportunities to reflect following the tasks beyond the classroom. Various methodological options within established proceduralized approaches towards introspection were considered before the decision was made to exploit the video from the tasks, combined with a series of interview questions based around learners’

negotiation, in relationship to the questions I had wished to answer. For example, there are methodological choices around introspective instruments which consist of think aloud, self-observation and prompted interview (Gass & Mackey, 2017, p.17). These suggested approaches were reviewed before the pilot study and in direct consideration of the multimodal aims and specific questions which I had wanted to answer within the main study.

Excerpts of video were discussed in micro detail, following a series of semi-structured interview questions, with learners encouraged to verbalize how and why they had used gesture in relationship to the mobile device and setting. I played a brief section of video which normally contained a sequence of gestures. Next, I began to highlight individual gestures, through replaying the video and pausing on each gesture. This process was repeated as many times as necessary in order that learners were enabled to focus on both modes and the interplay between them. I found that different learners required different numbers of replays and, for this reason, there was a flexible approach adopted in terms of the way in which the video was exploited. Learners were also encouraged to comment on their interlocutor's multimodal forms of communication. I had aimed to help learners to understand where combinations of gesture and speech had formed coherent meanings but to also raise their awareness of problematic aspects of communication, for example, where gesture and speech had seemingly operated in dissonant and potentially incoherent ways for interlocutors. It is important to mention that there were also some examples of negotiation of meaning within the data which had been achieved through language without the use of gesture.

For the interviews I decided to use a quiet reflective space back in the language school to allow learners to explore how they themselves might learn from an examination of their own and one another's multimodal forms of communication derived from their speaking tasks which had taken place beyond the classroom. The approach was in contrast to the task-based interaction beyond the classroom where learners were operating under conditions which could involve complex levels of multi-tasking, potential background noise; and possible difficulties with the technology itself.

The approach which I adopted to the interview questions, in combination with the use of the video-based 'artefact' was the semi-structured interview (Cohen et al., 2000; Dörnyei, 2007; Kvale, 1996). Contrary to Henderson et al. (2010) who advise that instructions to participants as to what information to recall can compromise the data, it was necessary in

the case of the recall of gestures to offer the participants a considerable degree of structured direction. I offered support through lines of questioning and enabled forms of dialogue in order to help them to revisit what they were seeing on the video in relationship to what they may have been thinking at the time of the task. As it is acknowledged that stimulated recall can result in highly interpretive and descriptive forms of data (Mackey, 2002; Gass & Mackey, 2017), the aim of my use of this particular instrument was not derived from a positivist but a qualitative research stance. Therefore, the present study was pre-occupied in exploiting this instrument to uncover learners' subjective experiences around negotiation; and to understand in what ways the approach could support them to build awareness of their use of the modes of gesture and speech in interplay. Lillis (2008) highlights the importance of enabling students themselves to talk about texts as a way to offer the participant both an 'analytic lens' and 'perspective' in ways which are critical in discerning "what may be significant and important in any specific context" (p. 359). I realised during the pilot sessions that learners required support to achieve this and that I needed considerable levels of flexibility within the semi-structured question paradigm to enable the interviewee to "elaborate on the issues raised in an exploratory manner" (Dörnyei, 2007, p. 136). The conduct of the qualitative interview is important and includes background preparation, pacing and timing, supporting the participant, and eliciting knowledge (Lincoln & Guba, 1985). I had initially designed a set of semi-structured questions based on the pilot study sessions (see APPENDIX 4). However, due to the exploratory nature of this study, I also allowed discussions to spontaneously emerge with the aim that these might lead to insights which I had not predicted or which might emphasise areas which learners themselves felt were especially important to them.

3.7 The data analysis procedures

The data analysis procedures in the study included mono-modal and multimodal analytic transcription methods based on the data from the collaborative VC sessions derived from the learners' tasks. For the analysis of the VC sessions, I first exploited the Varonis and Gass (1985) framework of the negotiation of meaning to code the speech turns and then within the second phase introduced multimodal analytic methods around the use of a gesture-speech unit of analysis (McNeill, 1992; McNeill & Duncan, 2000) which was integrated into the wider framework of negotiation. The implementation of the Varonis and Gass framework involved a second round of coding to check and confirm if the original coding decisions based on the speech turns alone could apply when this additional unit was

introduced into the coding categories originally exploited to depict the learners' negotiation. I also subjected several examples of negotiation to inter-reliability checks where there were some areas of disagreement. The coder had to be trained by myself to identify the gesture phrase in relationship to the learners' speech turns but was familiar with how to code data according to the negotiation of meaning.

The majority of the multimodal transcription work around gesture and speech was enabled by ELAN (see <https://tla.mpi.nl/tools/tla-tools/elan>) which is discussed in detail within section 3.8.1 of this chapter. My approach to multimodal analysis also draws on the analysis of gestures derived from Norris (2004). As mentioned previously, I did not choose to exploit CA methods, however, she uses forms of analysis derived from earlier work in gesture studies (Kendon, 1982; McNeill, 1992) within her multimodal interactional analysis framework. Unlike Kendon and McNeill, who exploit line-based drawings of gestures, she adopts the contemporary use of sequences of photographic images of gestures which are seen to occur within situated examples of spoken language use. I therefore considered that this more visual approach could also support an examination of learners' use of gesture via Skype VC on mobile devices as this unfolded within examples of learners' interactive forms of online negotiation.

The latter phases of the analysis involved working with and coding the learners' stimulated recall interview data via a second qualitative software tool: ATLAS.ti 8 <http://atlasti.com/>. ELAN and ATLAS.ti are both categorised as Computer Assisted Qualitative Data Analysis Software (CAQDAS) and supported me in differing ways during the various phases of the analysis. It is relevant to note that they are never a substitute for knowledge of a subject area and have to be combined with the analytic and interpretative skills of the researcher. From this perspective, it has been suggested that the researcher still requires a sound knowledge and training in SLA qualitative research methodology (Baralt, 2017).

In the following table (Table 9), I illustrate the nature and scope of the data collected for the purposes of this study.

Table 9: The nature and scope of data collected

Data collected in the study	The nature and scope of data
<ul style="list-style-type: none"> Five task-based interactions from five learner dyads enabled by Skype were captured on video. The dyadic interactions generated approximately five hours of raw data. 	<ul style="list-style-type: none"> When the negotiation of meaning examples were isolated from the wider data, these comprised of approximately three hours of data which was transcribed multimodally. There were examples of data where learners did not negotiate for meaning. Examples of the negotiation of meaning which contained no gestures were excluded from the analysis. The majority of the negotiation of meaning examples were achieved multimodally; and contained the use of iconic, deictic, prosthetic deictic but also beat gestures (beat gestures were used frequently but were excluded from the analysis). There was very little evidence of the use of metaphors. Gaze was a particularly prevalent mode across the data but this was also discounted from transcription and the final analysis due to the focus on gesture and speech.
<ul style="list-style-type: none"> Eight one-to-one interviews (two learners were not available) were captured. This resulted in another eight hours of data. 	<ul style="list-style-type: none"> The interview data was transcribed and coded in its entirety in order to arrive at abstracted themes. Smaller sections of verbatim interviews were also isolated and triangulated with their multimodal transcription counterparts.

I next explain the phases of analysis in the present study. These consisted of the following:

Phase 1:

- Transcription of the negotiation of meaning excerpts in terms of first taking account of learners' speech turns within ELAN.
- Initial coding of the data according to negotiation of meaning framework (Varonis & Gass, 1985).

Phase 2:

- Isolation of learners' gestures in terms of the identification of the individual gesture types through identification of the gesture phrase (see McNeill, 1992; Norris, 2004).
- Implementation of gesture-speech unit of analysis (McNeill, 1992; McNeill & Duncan, 2000) to enable the multimodal transcription of turns with the aim of taking account of the interrelationships between gesture and speech according to the GP.
- Second round of coding by myself to examine the data from a more holistic, multimodal perspective. Second coder to then carry out interreliability checks on an individual basis. We then met to discuss the outcome of the coding and to talk about where any discrepancies had occurred.

Phase 3:

- Production of a final set of transcripts illustrating the coding of the multimodal excerpts of negotiation; incorporating written transcription of gesture and speech (McNeill, 1992; McNeill & Duncan, 2000) and visual transcription of the learners' gestures (Norris, 2004) within the wider framework of the negotiation of meaning (Varonis & Gass, 1985). The use of images as data also designed to explicate how the learners' gestures related to aspects of technology and setting.

Phase 4:

- Initial use of verbatim sections of learner interviews examined together with multimodal transcription examples of the same excerpts of negotiation with the aim to draw conclusions from different vantage points, for example, to understand where there had been convergence and divergence once the written transcriptions of the stimulated recall interview data had been placed in ATLAS.ti 8 to enable coding within a qualitative content analysis approach (Mackey & Gass, 2012; Gass & Mackey, 2017; Schreier, 2013).
- Initial round of coding of the interview data.
- Second round of coding of the interview data.
- The coding scheme is checked for fitness of purpose and reliability by a second coder who was asked to code two interviews using the agreed categories. The second coder raised issues such as where my coding definitions could not be accurately matched to segments of the data due to overlap between categories or a lack of clarity in my coding definition. I then revised or refined some of the problematic categories.
- A final round of coding to re-assess the levels of coding consistency.

Phase 5

- Corroborating findings through triangulation: examination of data as a whole with attempts to view and interpret the data from various different perspectives. Exploration of similarities and divergences between the multimodal transcripts and learner interviews in order to attempt to answer the research questions from a broader perspective.

3.7.1 Data management procedures

As Creswell (2013) notes, there is the initial task of managing the data within qualitative research paradigms. In the present study, this involved initially working with approximately twelve hours of data; consisting of the data collected from the tasks and the individualised stimulated recall sessions. Two learners had also dropped out of the interviews and two tasks included problematic sections of data where either the visual channel had frozen, leaving the audio running, or where the screen had gone black. Using the identification of the boundary of a negotiation through identifying routines and examples of more extended negotiation, I was able to reduce the amount of raw video data from the VC speaking tasks from approximately five hours to very brief sections of video data which each lasted from twenty seconds to approximately four minutes. These were consciously chosen as they contained gesture-speech examples with the hands or the device and involved learners' interaction with the setting itself. In the case of the present study, the coding process also helped me to manage the data through enabling me to break this down into smaller units for the purposes of analysis. This process was regularly iterative in that I went back and forth through several stages of coding procedures and engaged with a series of mono-modal and multimodal transcription phases before arriving at the final stages of analysis. In qualitative terms, coding is the activity that the researcher engages in, while codes are "names or symbols used to stand for a group of similar items, ideas, or phenomena that the researcher has noticed in his or her data set" (Le Compte & Schensul, 1999, p. 55). However, Silver and Lewins (2014) suggest that codes can be used for a broad range of purposes and that the researcher can also exploit these to transcend more conventional methodological ways of working.

My use of ELAN throughout the process of the research allowed me to store my video files and to importantly link specific examples of learners' gestures and speech within the examples of negotiation to my ongoing analytic transcriptions of these within the tool (see section 3.8.1 of this chapter). In the final phase of the analysis, the stimulated recall data was stored and managed within ATLAS.ti 8 (see section 3.9 of this chapter). Coding the second data source entailed the management of a further set of written transcripts or 'texts' of the learners' spoken stimulated recall interviews. In the latter stages of the analysis, the

software enabled me to reduce the amount of interview material, to designate codes, and to also discern patterns and themes through visualisation of the data.

It was important for me to understand that the excerpts of negotiation and modes which I had chosen to focus on within the analysis entailed decisions as to which areas of data to afford attention to and which to dismiss. This was never an innocent decision but was instead driven by the framework of the negotiation of meaning and the multimodal approach adopted to extend this to encompass gesture and speech in order to answer the questions. For example, there may have been several other salient task-based incidents or deployment of other relevant modes which I could have included the analysis. Conversely, in constantly revisiting the video data from the tasks, I began to realise that in some instances the interactive lead up to the negotiation itself illustrated how this was crucially triggered and enabled by learners as they deployed mobile devices in settings beyond the classroom. Therefore, I chose to include some of this information in the transcripts and my descriptive accounts in order to elucidate how negotiation is initially stimulated and supported by peers within this particular research context. In another example, the learners ceased to gesture due to factors based around their use of technology and setting in ways which became relevant as to why they had chosen to continue their negotiation through speech alone.

In selecting these limited sections of data from the learners' tasks, I acknowledge that I am presenting a representation and interpretation of the interactions as I witnessed these based on my chosen framework and the focus of my research questions; and one which will also consistently reflect the interpretative research stance which I have chosen to adopt. My selective choices may also represent a form of modal foregrounding and bias, which were designed to permit me to explore the chosen modes of gesture and speech but also to understand these in a contextualised sense. I drew on the different sources of methods and data collected within my final interpretative analysis in order to answer the questions which I had posed in a manner which would ideally provide a more in-depth view of negotiation than had I relied on a single approach.

3.7.2 A methodological toolkit

It was necessary at the beginning of this study to understand that there was an existing methodological toolkit which could be adapted and further developed in order to support the goal of fine-grained multimodal analysis and transcription in relationship to gesture and

speech; and one which could potentially be integrated within the wider framework of the negotiation of meaning. The toolkit was made up of several components. Firstly, there was the requirement to roughly identify the learner's gesture type, determined through my studying the overall shape and trajectory of the movement, though also taking account of what the learner had said at the time. Secondly, I needed to break down the speaker's gestural phrase (McNeill, 1992; Norris, 2004) into its constituent parts in order to identify the segmented phases within this in terms of how the learners' use of combined modes was harnessed within their negotiation. I considered that a multimodal gesture-speech unit of analysis (McNeill, 1992; McNeill & Duncan, 2000), based on the notion of the GP, could take account of both modes operating together and help to explain the role of gesture for language learners. The final stages of the transcription phase involved a re-representation of the learners' use of gesture and speech with account taken of their use of technology and the physical settings in which they were situated in terms of how these had all influenced their use of gesture. In the following section, I will go into more detail about each of these components.

3.7.3 Gesture types

As mentioned previously, gesture-speech analysis and its established transcription procedures (McNeill, 1992; McNeill & Duncan, 2000) involves the researcher determining the gesture type (McNeill, 1992; Norris, 2004) through deconstructing the gestural phrase in relationship to the spoken language deployed at the same time. This method had been exploited in previous approaches to the analysis of L2 speaking tasks in face-to-face language learning contexts (see McCafferty & Stam, 2008 for an overview). The approach acts as a way for researchers to understand aspects of individual cognition and social communication in terms of understanding the role of gesture in relationship to how L2 learners combine both modes to construct their meanings.

I next describe the series of gestures as shown previously in Table 3 within chapter 2. This acted as a guide to the identification and coding of the gesture types of interest in the analysis within the present study.

- 1) **ICONIC**: “Possess a pictorial content, often mimicking what is conveyed verbally, describing specific objects or events, making them more vivid.” (Norris, 2004, p.28, based on McNeill, 1992).
- 2) **METAPHORIC**: “Possess a pictorial content, however, they present the invisible: an abstract idea or category” (Norris, 2004, p.28 based on McNeill, 1992).
- 3) **DEICTIC**: “Point to objects or people in the physical world or to abstract concepts and ideas as if they had a physical location” (Norris, 2004, p. 28 based on McNeill, 1992).
- 4) **PROSTHETIC DEICTIC**: Deictic gestures are usually associated with, but are not limited to, the use of the extended index finger. Pointing is achieved technologically or prosthetically by means of the camera (Jaworski & Thurlow, 2011).

Norris (2004) discusses the analysis of these first three cited gesture types (based on McNeill, 1992) in relationship to speech within her framework of multimodal analysis in ways which supported my own work. She notes that “hand and arm movements are often interdependent and concurrent with spoken language” and that whilst gestures may sometimes start slightly earlier than the language used to express the same idea, it is deemed less than useful to separate these types of gestures from the language with which they occur (pp. 28-29). Therefore, without understanding the language surrounding the gesture, it is difficult for the multimodal researcher to recognise and analyse the role of a gesture for a language learner. As “the two modes of gesture and language are [...] so closely linked” it is advised that in order for the researcher to reach a conclusion about the gesture type, it is necessary to refer to one mode in order to be able to comprehend the message conveyed in the other (Norris, 2004, p. 29).

For me, it became easier to understand the role of gesture and its interrelationship with speech within the wider framework of the negotiation of meaning when I had learned to recognise the learners’ gesture types in terms of what the learners had said during the same portion of negotiation. However, Krauss et al. (2000) discuss problems for the researcher in identification between iconic and metaphoric gestures. He notes that the iconicity of gestures “seems to exist primarily in the eyes of their beholders” in that different viewers have been shown to assign different meanings to gesture; even when observing them in relationship to their accompanying speech (p. 276). In the present study, the coding of

gesture types was approached in the second phase of the analysis, following my coding of the speech mode; and this was designed as a way to isolate and try to understand the meaning and role of gestures and the relevance of gesture type in relationship to this.

As discussed previously, two types of deictic gesture were categorised within the present study. These were identified as the learner pointing with their human hand and finger but also included prosthetic deictic gestures, formed through the learner pointing with the mobile device itself. Deictics constructed with the human finger were more easily analysable than the second type of deictic gesture which the learner formed with the mobile device and camera as they pointed at a feature of their setting. There had been prior identification of a deictic gesture type in terms of people pointing to aspects of their environment with cameras and videos (Jaworski & Thurlow, 2011), however, there had been a lack of proceduralization in publications at the time of writing this thesis. Given the scope of my research questions and the context of the research, I felt that these gesture types could not be ignored within the analysis. Norris (2004) argues that deictic gestures, in contrast to iconic and metaphoric ones, can make sense without speech but I felt that her observations on shared interaction were derived from research within face-to-face contexts where it was deemed possible for two participants to point at a referent within a context without the simultaneous requirement to speak. In contrast, it was considered that learners operating across two geographically separate contexts, in order to complete a pedagogic speaking task, could not point to something without also being prompted to explain their action through language in terms of indicating to an interlocutor which aspects of their setting they were attempting to show or present online.

3.7.4 The Gesture Phrase

One of the most useful elements of the methodological toolkit in the present study was my increased awareness as to how to analytically transcribe the learners' gestures via the breakdown of their segmented parts. This would allow the multimodal researcher to transcribe and understand a gesture's timed synchronicity with speech as a means to understand the semantic meaning of the gesture when this is linked back to the speech mode in a more holistic sense. I considered that the method would also be able to help me to understand where learners' speech and gesture were exploited in asynchronous ways or did not cohere in meaning for an online interlocutor. I also considered whether or not a

focus on the gesture phrase could reveal how a lack of coherence in meaning had contributed to issues of non-understanding for learners within their negotiation.

As discussed in Chapter 2, the gestural phrase is as follows:

- 1) *Preparation*: the limb moves away from the rest position to a position in gesture space where the stroke begins.
- 2) *Stroke*: the peak of the gesture where the meaning of the gesture is conveyed.
Post-stroke hold: the final position reached is held.
- 3) *Retraction*: the hand returns to its rest position.

(Norris, 2004, p.30 based on McNeill, 1992)

It also became apparent within the process of analysis in the present study that gestures via mobiles do not operate in the same manner as those executed in face-to-face contexts, for example, the preparation and retraction phases can involve a rapid movement of the arms and hands which are mostly lost via the mediating nature of the screen boundaries. Kress (2003) asks us to reconsider the affordances and materiality of mode in relationship to the mediating effects of technology; emphasising how resources such as gesture offer potentials for representation but can also create limitations in their usage (p. 45).

The following images in Figure 5 (shown next) are taken from the present study and depict an example of the gestural phrase in relationship to a learner’s iconic gesture which is formed with their hands in relationship to mediating aspects of the mobile tablet.




<p>Timing: 6:55</p> 	<p>6:57</p> 	<p>6:58</p> 
<p><i>Preparation</i>: hands and arms move upwards into a space beyond the boundaries of the mobile tablet screen.</p>	<p><i>Stroke</i>: hands construct meaning as the learner speaks the word ‘horn’ as the gesture stroke is formed.</p>	<p><i>Retraction</i>: hands and arms move downwards into a space beyond the boundaries of the mobile tablet screen.</p>

Figure 5: The three movement phases of an online iconic gesture via a tablet

In the case of gestures formed with the hands, it became apparent during the analysis that, in relationship to the boundaries of the mobile screen, the *preparation* phase usually

appeared as a fleeting, blurred image of the hands beginning to move upwards or forwards in preparation. Next, the meaning-based gesture *stroke* and *post-stroke hold* had to be identified in relationship to the positioning and boundaries of the mobile screen. The *retraction* was recognised as a distinguishable downward or backward movement which lasted for a period until the hands and arms were seen to disappear from my view. However, during the course of the analysis it became apparent that the online gestural *stroke* was the most visible but also the most important element of the gesture to both capture and transcribe in terms of the learner's meaning. In terms of the optional *post-stroke hold*, learners did not then tend to change the position of their hands. This made transcription of the *post-stroke hold* somewhat superfluous to my analysis. To carry out the analysis more efficiently and to simplify the rather dense final transcripts, it became more productive to focus on the three main phases of a *preparation, stroke, and retraction* in relationship to the learners' language use.

This is supported by the research into multimodal analysis, where it has been suggested that the meaning-based *stroke* is deemed to be obligatory but that all other phases of the gesture are optional but that “most iconic, metaphoric and deictic gestures consist of three movement phases” (Norris, 2004, p. 30). In their analysis designed to establish gesture's role in meaning-making for L2 learners, Negueruela and Lantolf (2008) chose to exploit McNeill (1992, 2005). Their analysis of participants' gestures and speech does not, however, take account of the gestural hold but instead focuses on the identification and transcription of the onset of the gesture, the meaning-based stroke; and the subsequent retraction phase in relationship to L2 language use.

The analytic concept of the gestural phrase then had to be reconceptualised and adapted in order to attempt to take account of the manner in which learners formed gestures with the mobile device as they pointed at aspects of their setting; that is using a: ‘prosthetic deictic gesture.’ The phases of this gesture were analytically determined by myself in relationship to observations of the learner exploiting their mobile device to attempt to highlight a particular referent for an interlocutor across Skype VC (see Figure 6).

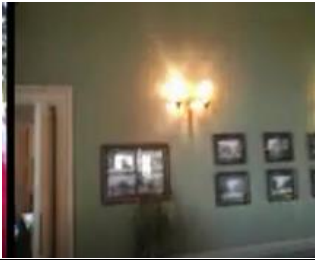
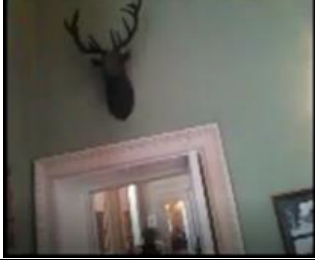

Timing 6:38 	6:40 	6:46 
<i>Preparation:</i> Learner begins to attempt to form the stroke with the device.	<i>Stroke:</i> Learner focuses on the designated artefact in her setting.	<i>Retraction:</i> learner removes device from its focus on the object.
Speech: ‘There is a beautiful	head	You see it?’

Figure 6: Andrea forms of deictic gesture with the device and camera within her setting

Figure 6 shows that the *preparation* was operationalized as the learner was beginning to move her device and camera towards the target referent in her setting as she moved into the vicinity where the artefact of interest was positioned. The *stroke* was identified as the moment of peak effort where the device and camera were directly and, ideally accurately, pointed at the referent in the learner’s setting to convey its meaning. A subsequent shift in the device’s position implied the *retraction* or ending of the gesture. This is how I am attempting to understand the role of this type of deictic gesture in terms of what learners were attempting to achieve with their deployment of Skype VC via mobiles in relationship to the conveyance of an artefact within their setting to an interlocutor.

3.7.5 A gesture-speech unit of analysis

The multimodal unit of analysis adopted for the purposes of the present study is derived from gesture-speech partnership theory (McNeill, 1992). This approach entails the researcher’s microanalysis and multimodal transcription of the timed co-occurrence of the modes of gesture and speech when viewed as operating in relationship to one another. I chose McNeill’s unit of analysis in that it potentially enabled the identification and deconstruction of the learners’ individual gestures with the possibility to then examine these gestures as they unfolded within exchanges which would possibly build in a sequential manner to assume a role within excerpts of negotiation. I considered that this meant that this gesture-speech unit of analysis could be integrated into the existing framework within SLA to extend this linguistic model into a multimodal analytic framework which was designed to examine L2 gesture and speech when operating together.

3.8 The multimodal analysis of the videoconferencing tasks

The rationale behind the use of transcripts within multimodal approaches to research implies that the researcher's creation of and interaction with these transcripts constitutes the analysis itself in relationship to their research aims and questions. Dörnyei, (2007) notes that written transcripts are 'seriously impoverished' in that they cannot capture eye movements, gestures and facial expressions (pp. 246-247). The approach to the multimodal design of the transcripts in the present study was that it would reflect the manner in which I had chosen to address my research questions and implement my chosen theoretical framework; and that it would constitute an interpretation of the learners' original VC interactions from beyond the classroom. The construction of a detailed series of multimodal transcripts was not seen as a preparatory phase within the present analysis but the transcripts were revisited and reworked on an ongoing basis and continually shaped through my consistent interaction with them as I analysed the data and considered my research questions.

Norris (2004) notes that language is a good place to begin multimodal transcription in that it has a "high information value;" also highlighting how we are educationally conditioned and, as a result, "more inclined to consciously make sense of what a speaker is *saying* than to notice what a speaker is expressing in other modes" (p. 66). The deployment of the SLA negotiation of meaning framework first enabled me to identify relevant sections in the video of the learners' tasks through the identification and coding of a set of verbal triggers and indicators before then locating the resolutions. However, as mentioned previously, negotiation did not always fall neatly into trigger, indicator, resolution and reaction to resolution with many excerpts across the data in the present study demonstrating much more complex and extended forms of negotiation which were often complicated to code.

3.8.1 The use of ELAN

Multimodal data analysis around gesture is complex and requires very detailed levels of transcription. ELAN was chosen as it is specifically designed to support fine-grained analysis of gesture and speech through the production of multimodal tiered transcripts. ELAN is categorised as Computer Assisted Qualitative Data Analysis Software (CAQDAS) <https://tla.mpi.nl/tools/tla-tools/elan>. The affordances of the tool allow for

video data to be imported into the software. Enabling the analysis of gesture with speech, for example, it allows the researcher to transcribe complex visual data containing split second movements through the re-playing video of gestures within an example of negotiation at, for example, half-speed, around one tenth of normal speed, frame by frame etc.

Once I had imported the video data from the speaking tasks, I was able to watch the learners' gestures in detail by slowing down their movements. This allowed me to locate the meaning-based portion of the gesture (the *stroke*); and to then gradually establish the other individualised phases which had occurred around this. This meant that I was able to first analyse the phases of the gesture in relationship to the coded speech turns and then how the learners had achieved negotiation multimodally, for example, I began to understand the Varonis and Gass coding categories through elucidation of learners' deployment of the two modes operating in interplay. I deployed ELAN to multimodally transcribe each identified negotiation excerpt within the VC sessions and whilst engaged in this process over many weeks and months I found that I was gradually able to more accurately transcribe and interpret what the learners appeared to be saying but to also understand what they were doing in order to construct their negotiation.

Other reasons that the software tool is valuable for this type of multimodal analysis and transcription is its ability to support the researcher by providing multiple tiers to depict different modes being exploited by the participants in co-orchestration as a way for the researcher to establish the semiotic relationships operating between them. The researcher ends up working with a form of musical score with the instruments represented as the modes and the necessity to read left to right but also downwards at the same time to view the interaction in its entirety. Effective use of this tool relied on my increasing ability to more accurately annotate the co-occurrence of modes whilst also building my own interpretative and analytic skills through a steep learning curve.

In general, I found that the first coding decisions from phase one of the analysis did not alter in the second phase as learners had seemingly integrated gesture into their speech turns. The Varonis and Gass (1985) framework was generally able to take account of both modes operating together from triggers through to forms of resolution within the negotiation. The exception to this rule was evidenced in the visual comprehension checks which could not be reliably coded until the data had been analysed multimodally. However, it became apparent in the second phase of analysis that language use had only

represented a portion of the wider interaction for learners in terms of their negotiation of meaning. Gestures made little sense without linking them to speech, and speech turns alone were unable to fully explain what learners were trying to achieve with their hands, devices; and interaction with the setting. In one example, from coding the speech turns in isolation, it initially appeared as if the learner was pre-occupied with indicating, through an indicator move, that she did not understand the specific Turkish word for a particular dish. Conversely, when the concept of the prosthetic deictic gesture was introduced and integrated into the analysis, it became clear that the verbal indicator had been exploited in order to draw attention to an incident of multimodal ‘non-understanding.’ When the gesture-speech unit of analysis was introduced, it revealed that the learner had queried the meaning and accuracy of an interlocutor’s gesture in relationship to what they had said. This incident of ‘non-understanding’ resulted in the necessity for both learners to collaborate in order to repair the manner in which the device and camera were pointed, rather than to focus on the clarification of a language item.

The transcript in Figure 8 is a representative example from an initial transcription phase enabled by ELAN. It should be noted that, even though I had transcribed the speech turns first, a five minute clip of video still took approximately five hours to transcribe within the second phase of the analysis.

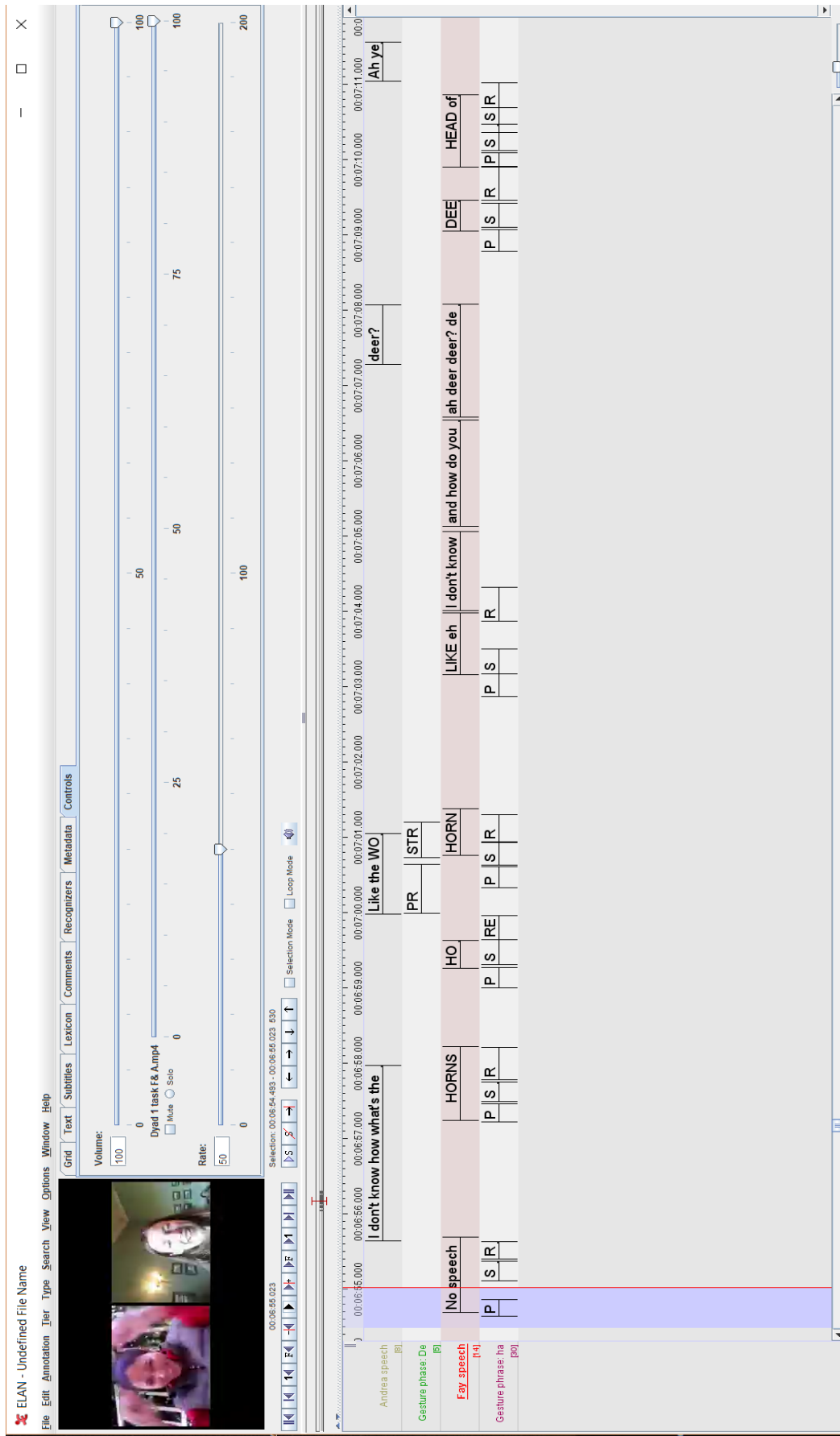


Figure 8: A section of multimodal transcription from ELAN


I designated and labelled sets of separate tiers for each learner around speech and gesture; and also allowed for reflexive notes recording my observations as I attempted to understand how the wider environment had both intertwined with and influenced the deployment of gesture and speech. I frequently used the loop mode in the software tool to keep replaying the relevant selection of video to ascertain the accuracy the gesture phrase. I then labelled the gesture as a new annotation whilst also editing and revising these decisions. The co-occurrence between gesture and speech were constantly checked, revised; and refined in ways which involved the multiple playing of microsegments of video and slowing them down to various speeds to more-accurately annotate the gestural phrase in relationship to the speech.

I also faced challenges when I had only collected data of the ‘picture-in-picture’ or ‘default’ view (one large and one small Skype window) from the internal capturing software Supertintin. When in doubt about a gesture type or when the gestural phrases could not be discerned because I could not see the movements from the learner in the smaller Skype window, I found that I could exploit the video viewer in ELAN. It was possible to zoom in on the video, however, the images captured this way could be quite blurred. I often enlarged a series of sequential screenshots to support me to study the unfolding sequence of the gestures and I used these to represent the gestures in the final set of transcripts. I was able to use ELAN to discern the gestural phrase within the multimodal unit of analysis as a systematic way to break down the learners’ gestures with their hands but also their deictic gestures with the device acting as prosthesis in order to attempt to understand their role in negotiation. The analysis of this latter type of deictic gesture was more complex than the analysis of the segmented phases where learners were gesturing with their hands alone. The *preparation* with the device would normally involve the learner moving their hands, device, and camera around the referent which they had wanted to share with an interlocutor. When captured via video internally, this phase was evidenced as the learner attempting to point in the vicinity of the artefact, for example, they would start to move the device and camera towards a facing wall or table where the target object had been positioned. However, the camera was often pointed inaccurately with learners often making second and third attempts to form the stroke accurately enough for an object to be seen by their interlocutor.

3.8.2 Challenges to the existing framework of the negotiation of meaning

Varonis and Gass (1985) depict various nuanced degrees of non-understanding within their examples of negotiation and I also observed that communication rarely broke down completely between dyads taking part in the present study. Negotiation offered them multimodal opportunities to reiterate, clarify, and repair information; and to indicate when unknown vocabulary items had emerged during their online exchanges. However, contrary to Varonis and Gass' (1985) study, the multimodal forms of analytic transcription of the data in this study illustrated that negotiation was often triggered through aspects of the setting itself. The deployment of hand gestures and speech were co-ordinated by learners in order to reference elements of these settings as they attempted, for example, to understand where their online peers were situated (see Excerpt 1). Indicators of non-understanding, and the subsequent attempts to resolve negotiation, could be achieved through an offer to 'show' or reveal a material space to another learner in virtual ways via pointing with the device and camera. It is important to note that the type of negotiation evidenced within the present study was a direct result of the context-sensitive task design in that learners had been encouraged to locate objects and share these across Skype, as shown in Excerpt 1:

Excerpt 1: Learners negotiate for meaning based on setting and artefacts

1. F: [Where are you] [sitting?] [It looks like] a library you know. Is it?	T
2. A: Sorry?	I
3. F: Ha ha ha [library]. It [looks like a library] where you're sitting. Eh [behind you the books]	R
4. A: Yeah	RR
5. F: [The books behind you]	R
6. A: Do [you want I show you? 	I

After introducing the gesture-speech unit of analysis and adopting a perspective on gesture in terms of how their use related directly to the setting of the interactions, it became clear that a written transcript alone, even a multimodal one, could not fully indicate how negotiation was achieved by learners. For example, images conveyed across mobile devices were exploited as a gateway for interlocutors to view gestures but to also glimpse

key artefacts within one another's settings. This prompted their curiosity as to where an interlocutor might be situated and this aspect of communication, rather than trigger moves based on solely on language use, contributed to the enabling of negotiation from beyond the classroom (see Figure 9 below).



Figure 9: A multimodal trigger for negotiation based on location

Within the process of transcription and coding of the speech turns, learners were also heard to form new types of comprehension checks such as: “Can you see?” I considered that this move represented a strategy for checking levels of understanding from multimodal perspectives, however, this type of comprehension check is not cited in the original framework by Varonis and Gass. I designated these categories as visual comprehension checks (VCC) within the present study. Excerpt 2 shows an example of a coded visual comprehension check which also serves as a trigger, followed by an indicator; and then a second comprehension check and response. This data has been taken from the spoken turns within an extended example of negotiation and is based around a problem with an item of vocabulary.

Excerpt 2: An example of visual comprehension checks

1. A: There is a beautiful head. I don't know if you see.	VCC/T
2. F: I can see it. It's a....	I
3. A: You see?	VCC/R

3.8.3 The final multimodal transcription phase

There are many transcript choices around multimodal methods which can range from deployment of tables, screenshots and drawings, and combinations of systems of writing with image (Austin et al., 2017; Baldry & Thibault, 2006; Flewitt et al., 2014; Streeck et al., 2011; Mavers, 2012). As Weber (2008, p. 47-49) notes, the last two decades have seen a considerable growth in interest in visual research methods for various reasons:

- Images can be produced by participants as data.
- Found or existing images can be used as data or springboards for theorizing.
- Images and objects are useful to elicit or provoke other data.
- Images can be used for feedback and documentation of the research process.
- Images are useful as a mode of interpretation and/or representation.

Much of my previously depicted analytic transcription in ELAN required interpretation and interpretative ‘sense-making’ derived from studying images of the learners’ gestures and this was approached with the same degree of rigour as the analysis of their speech turns. In consideration of the final sets of transcriptions in this study, the multimodal unit of analysis (McNeill, 1992; McNeill & Duncan, 2000) allowed me to exploit a written multimodal system to transcribe the learners’ gestures in relationship to their speech with bold typeface indicating the *stroke* and spoken language coming together; as within the theory of the GP. Conversely, McNeill’s approach to transcription could offer no indication of the visual appearance of gestures constructed with the hands across Skype or in terms of depicting the types of images which appeared for interlocutors when a learner pointed at an object within their environment. I considered that a multimodal written system (illustrating the gesture phrase in relationship to speech) would fail to illustrate how the gestures with the hands appeared in terms of shape etc; and the manner in which the mobile screen and device mediated gesture, for example, how gestures were enabled or constrained for learners due to aspects such as the size of the device’s screen. It was important for me to also consistently acknowledge that it is inherently difficult to translate a dynamic gesture, captured as a moving image from within a video clip, and then position this onto a fixed page. Flewitt et al. (2014) explain the process of ‘transduction’ as “dynamic situated speech or action” being transformed into a “slower-paced format” in ways which can also help the researcher to view their data from new perspectives (p. 52).

Dealing with large amounts of images also involves careful research consideration and recognition of the challenges involved in finding ways to systematically understand these types of data within the social sciences (Knight, 2002). When the learners' fleeting gestures were captured and frozen in time, it was possible to open them up to increased scrutiny and interpretation in ways which involved increased recognition of communicative patterns around gesture and speech which had only emerged when multiple images were placed into a final set of transcripts.

Mavers (2012) discusses the complexity of the transcription process within multimodal research:

Embodied expression and interaction are always multimodal. A transcript is multimodal when it contains more than one mode. Including image as well as writing [...] forces the transcriber to decide which meaning will reside where and how these modes relate [...] A series of photographic stills provides certain information at a glance, such as features of the setting, objects and what people look like, which may or may not be included in a written version. (p. 4)

As discussed previously, the learners' gestures had been analysed in consideration of their relationship to the wider environment in which they had occurred and, therefore, images captured from Skype which showed objects and glimpses of settings were viewed as integral to my description and interpretation of excerpts of negotiation as they unfolded in order to answer questions. Multimodal research aims can be addressed via inclusion of "the physical and material setting of communicative and representative activity" (Flewitt et al., 2014, p. 48). It was deemed impractical, due to space restrictions on the fixed page, to exploit both a detailed written transcription scheme and at the same time visually demonstrate all the phases of the learners' gestures through sequential screenshots which would represent each gesture from the *preparation* through to the *retraction*, as Norris (2004) suggests. In contrast, I considered that the area of gesture which most related to the learners' negotiation of meaning would be captured in a visual representation of their gestural *stroke* where the principal meaning of the gesture was concentrated.

In the final transcripts, I exploited a written system to transcribe the beginning and end of the gesture through placing left and right brackets in relationship to the learner's spoken turn. I then depicted the meaning-based *stroke* of the gesture in relationship to the corresponding word or words to demonstrate the GP, as illustrated in bold typeface within

the turn within the written portion of the transcript (see McNeill, 1992; McNeill & Duncan, 2000).

To try to begin to link the meaning of the written gesture-speech system to the meaning of the visual text of gestures, I used a basic timestamp system to support recognition of modal interconnections for a reader between the written and visual transcription system. I stamped the beginning of the speech turns and then labelled the timing of the gesture *stroke* in the individual screenshots to enable the reader to link both modes and ideally convert this into one holistic unit of meaning within the coded negotiation excerpts. I also had to consider how the reader would understand the gestures in relationship to the small portion of a setting which had been captured in the screenshot. The aim was to enable understanding that, in the case of a learner pointing with their device, the object shown in the screenshot was the visual representation of the written gesture-speech system. I used written and visual transcription strategies to illustrate how the learner had exploited the device to form prosthetic deictic gestures in relationship to how this action had occurred in relation to their use of spoken language.

Multimodal transcription around gesture, and other modes, always involves choices and levels of compromise within transcripts. It is also important for a range of interested researchers to continue to develop this emergent field. I consider this goal to be related to the search and proceduralization of increasingly systematic ways to accurately analyse and represent the multimodal complexity of learners' collaborative use of a range of ubiquitous technologies. Within qualitative work, the researcher's transcription ideally forms a link between the original interaction, the phases of analysis; and the descriptive and interpretative 'validity' of the findings for the future reader of the research.

3.9 The analysis of the stimulated recall data

In the previous section on analysis, I described the analytic transcription of the language learners' excerpts of negotiation. In the next phase of analysis, I transcribed the learners' stimulated recall interviews into a written text and spent time studying what the learners had said verbatim about their negotiation and compared this to the interpretative multimodal transcripts derived from their tasks. I next coded the interview data to further explore themes through the adoption of a qualitative content analysis approach (Mackey & Gass, 2012; Gass & Mackey, 2017; Schreier, 2013). Qualitative content analysis has been defined as "a research method for the subjective interpretation of the content of text data

through the systematic classification process of coding and identifying themes or patterns” (Hsieh & Shannon, 2005, p. 1278). I chose this method within a qualitative paradigm as it appeared flexible and constituted an interpretative analysis of “the underlying deeper meaning of the data” (Dörnyei, 2007, p. 246). Whilst qualitative content analysis is frequently associated with grounded theory approaches (Dörnyei, 2007), there are other forms of qualitative content analysis which can involve more directed, deductive approaches where the researcher may bring their own categories to the data based on theory, previous research findings, and research questions (Saldaño, 2013). As mentioned previously, I had wanted to gain a learner perspective on the data derived from the video-based tasks which had taken place from beyond the classroom. I was particularly interested in how but also why learners had exploited gesture in different ways and also considered that multimodal forms of transcription could not reveal in-depth levels of information about subjective learner experiences. I also considered that the triangulation of the stimulated recall data with observations would ideally mitigate limitations of both the methods of analytic transcription and stimulated recall as I aimed to explore a broader research picture.

I first created a full written transcription based on the learners’ interviews. Transcribing a one hour learner interview took me several hours but this was also a way to become familiar with the data in the early stages of the process. I then reviewed the multimodal transcripts of the Skype sessions and then began by reading through the transcripts of the interviews multiple times to reflect on these, write reflexive notes; and to gain an overall sense of the data in terms of the questions which I had wished to answer. The use of ATLAS.ti 8 enabled me to move to a phase where I could assign initial codes in relationship to segments of meaningful text within the learners’ comments (see Figure 10, shown next).

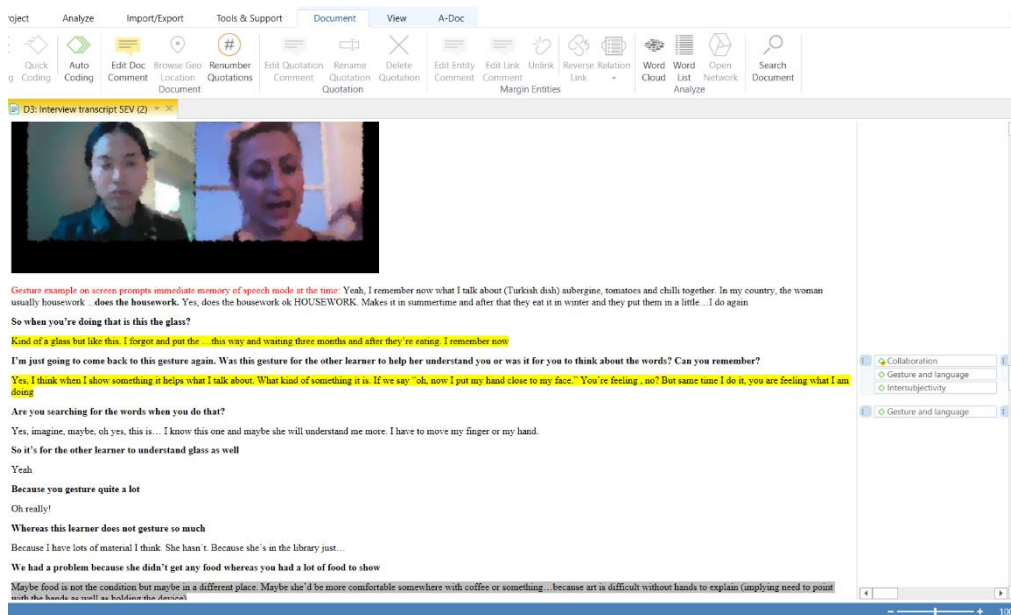


Figure 10: Example analysis of interview from Atlas.ti

This way, I began to reduce the amount of data through broad segmentation before moving on to code in a more fine-grained manner. I defined the code properties to allow for consistency across these but often changed and refined these further through an iterative process. Whilst I had previous experience in interviewing learners about language use within stimulated recall procedures, I was also learning in that I had never prompted participants to talk about their multimodal communication before. I felt that I gained initial insight into the condition of how this potentially operated during the process of the pilot study. In the case of the main study, the data analysis of the stimulated recall was approached from the perspective that this type of data within qualitative work can involve a richer understanding of a particular area.

For the purposes of this exploratory qualitative study, the flexibility of the approach allowed me to draw on my research questions and the previously analysed data from the tasks to further explore hunches and perceived themes which I had wanted to understand in greater depth. There were also other themes which emerged more inductively from the learners' comments in the interview data. As a result, whilst certain codes were arrived at in a deductive sense, other themes, relating to outcomes such as the enabling of multimodal metalinguistic discussion, emerged from the data itself. Miles and Huberman (1994) note the usefulness in exploring ways to arrive at categories both deductively (bringing codes to the data) and inductively, through working with the data in ways which can involve locating the codes from within the data. I was also able to use the data visualisation tools in

the software to better understand the linkages made between quotations and codes (see Figure 11).



Figure 11: Visualising linkages between quotations and codes in Atlas.ti8

I noticed that during the final phases of analysis, I began to explore the meaning of the data as a whole. I tested the validity of the coding categories by making comparisons between two separate rounds of coding within the space of a week and asked a colleague to then code some sections of the data exploiting my coding scheme. This was particularly useful as problems did become evident in my definitions regarding the deployment of codes such as gesture and lexis. Following this issue, I then made the decision that this category was only to be used when a learner mentioned a specific item of vocabulary or when they deployed the term ‘word.’ Following this input, I returned to the data and engaged in a final round of coding.

3.10 Triangulation of data sources

Triangulation around data collection sources (Mackey & Gass, 2012; Mackey, 2002; Lincoln & Guba, 1985) has been defined as:

[T]he use of two or more methods of data collection in the study of some aspect of human behaviour [...] Triangulation is a powerful way of demonstrating concurrent validity, especially within qualitative research [...] [E]xclusive reliance on one method, therefore, may bias or distort the researcher’s picture of the slice of reality she is investigating.

(Cohen et al., 2000, p. 112)

In the context of reviewing the literature findings, I had realised that the negotiation of meaning had been frequently reduced to an examination of language learners’ deployment

of interactionally-modified speech turns with little account taken of gesture or, perhaps more importantly, their own unique experiences of interacting. Using data examples from face-to-face scenarios, Mackey (2002) demonstrated how an examination of the researcher's interpretation of negotiation could be partially challenged when compared to the learners' own perceptions of their interactive opportunities. This discrepancy is rarely mentioned in studies on the negotiation of meaning with the researcher's coding decisions and interpretation frequently left unchallenged.

In the present study, excerpts of negotiation had not been coded at the point when the stimulated recall had taken place, however, they had been roughly identified to have been resolved when learners had made affirmative statements or moved on to another section of the task. When stimulated recall was used to approach these same excerpts of data through a different method of data collection and source, negotiation was subjected to a different form of scrutiny and interpretation from the learners themselves. Analysis of the interview data resulted in discovering that learners had sometimes expressed a difference in opinion to the coding decisions which I had taken in the transcription phase. An example of this involved a learner whom, whilst discussing her gestures and speech, told me that she was somewhat dissatisfied with the overall outcome of the negotiation. Conversely, the Varonis and Gass (1985) coding scheme prompts the researcher to interpret reaction to responses (RR) moves as indicative of an end to the negotiation with affirmative statements seen to signal mutual understanding between two learners.

As a result of triangulation of the transcript data with the stimulated recall interviews, new information began to come to light, for example, I had been unable to understand certain aspects of the mobile device and camera acting as prosthesis until participants explained more about this process within their interviews. When I combined the interview data and my interpretation of the transcripts, I began to more fully comprehend some of the difficulties which had been experienced by some of the learners. These included issues such as background noise (not heard on the video recording), the impact of features inherent in settings such as large staircases; and difficulty operating the camera on an unfamiliar device (the researcher's tablet). None of this information had been revealed in the analysis of the videoed tasks or via my transcription of the same portion of negotiation. Triangulation therefore involves making use of different sources but also methods of data collection to help to corroborate evidence and to provide validity to findings (Creswell, 2013, p. 251).

There were also times in the earlier phases of analysis where I had queried the type and purpose of certain gestures for a learner or learners within negotiation. For example, even when I had related the specific gesture back to the mode of speech, it was unclear to me in one instance whether the gesture had represented a concrete feature (i.e. was iconic) or whether the learner was perhaps trying to express a more abstract concept to an interlocutor in a concrete way (i.e. was metaphoric). This was especially evident when learners were gesturing but seemingly searching for language in order to express what they had wanted to say. In analysing the interview data, the learners had often explained how their gesture or gestures had related to their specific linguistic aims and had reflected aspects of their settings in ways which enabled me to more accurately confirm a gesture's type and role within a transcribed excerpt of negotiation.

Adopting a broader perspective on the data supported me to confirm aspects of reliability in terms of how I had approached and interpreted gesture-speech relationships in the initial multimodal transcription phase. Having drawn on dual methods and data sources, I took both of these into account within my overall analysis and presentation of findings in relationship to my research aims and the answering of my questions. I was enabled to present the multimodal transcriptions as one form of evidence and juxtapose these data with information about what participants themselves had said about the same excerpts of negotiation to include a learner-centred perspective. In my interpretation and descriptive forms of analysis, I drew on findings from the transcripts but also from wider themes which I had discerned from the interviews. This approach enabled me to more comprehensively describe and discuss the multimodal negotiation of meaning from a wider, and ideally richer perspective, than might have been possible had I either relied on one or other data source to address my research aims.

3.11 Choices regarding presentation of data

The principal multimodal transcript and interview excerpts which I am choosing to present in Chapter 4 (see Excerpts 3-10, pp. 148-197) are largely reflective of my wider data set. This is because they provide evidence of learners engaged in the use of gesture and speech in interplay within multimodal examples of the negotiation of meaning; and include data which demonstrates how learners both perceive and comment on their multimodal forms of communication. Each excerpt that I have selected to present, represents a particularly rich example of the negotiation of meaning in terms of the use of gesture. Whilst there were a multiplicity of examples within the wider data set which contained learners' use of gesture

and speech in ensemble (iconic, deictic, prosthetic deictic), the limitations of multimodal transcription procedures, and my aim to explicate the microgenesis of negotiation within a series of detailed accounts, ensured that I had to make challenging decisions as to which sections to include in this thesis. These decisions were guided by the rationale that the negotiation of meaning implies incidents of non-understanding and, therefore, a recognition that problem-solving is achieved by different learners in different ways. In line with Varonis and Gass (1985), I am choosing to present a combination of brief but also lengthy and highly complex examples of the negotiation of meaning. This is in order to illustrate the considerable variety in terms of how negotiation is approached and achieved by different learners. The excerpts also suggest that there are certain multimodal patterns across the data; particularly with regard to the types of issues experienced by learners around the use of prosthetic deictic gestures and the informal settings. In presenting both transcripts and short verbatim interview excerpts, the reader can view the data from two distinct perspectives in ways which are also designed to reflect my triangulation of the observations and stimulated recall data. Within Chapter 4, I exploit excerpts of data where learners comment in particularly meaningful ways about their multimodal communication. However, it is important to note that the majority of learners who took part in this study were able to talk about their multimodal interactions in ways which demonstrated both insight and considerable levels of understanding.

Excerpts 1 and 2 (see pp. 127-128 in chapter 4) are designed to illustrate examples of the earlier stages of transcription and coding; and to elucidate how features of negotiation such as comprehension checks (Varonis and Gass, 1985) are achieved in multimodal ways via mobile devices. Excerpt 3 (see p. 148 in chapter 4) is included for presentation because Fay and Andrea engage in complex forms of negotiation around the issue of vocabulary in ways which elucidate how processes in SLA are achieved multimodally. This particular negotiation excerpt involves a rich array of iconic and prosthetic deictic gestures which the dyad use in combination with aspects of their shifting language use. This particular choice of excerpt also demonstrates how differences can occur in terms of how a learner and a researcher decides to define and interpret an individual gesture type. For example, I consciously excluded Fay's thinking gestures (see p. 148, turns 4 and 5 in chapter 4) from my original analysis. One reason for this, was because I considered that Fay's movements did not represent 'true' gestures in that they appeared to represent examples of incidental and inadvertent movements, rather than being 'deliberate' or 'communicative' in their

intent (Kendon, 2004; McNeill, 2000). Moreover, Fay's thinking gestures could not easily be equated with any specific pictorial content as she is unable to name the object or to construct a visual representation of this with her hands at this stage in the negotiation process. In conjunction with her first thinking gesture (turn 4), Fay attempts to find a suitable word in order to describe what she can see on her mobile screen during this particular section of the task but she does not verbally describe a specific person, event or object (iconic gesture). The gesture is not metaphoric as Fay does not signal an iconic representation of an abstract concept. Furthermore, Fay does not use the gesture as a means to engage with a lexical search (Negueruela & Lantolf, 2008). This is because the learner fails to exploit her hands in order to recreate a concrete image of the word she is looking for as she tries to remember this: "It's a...". At the precise moment that Fay then deploys another thinking gesture (turn 5), her interlocutor simultaneously asks her if she can see a particular object on her mobile screen: "You see it?" Fay's second thinking gesture was also problematic to categorise for the same reasons that I have previously described. I could not categorise this particular gesture as 'co-speech' (McNeill, 2000) because Fay chooses to remain silent as she forms it. At this point in the negotiation excerpt, the semantic meaning between gesture and speech becomes severely disrupted. This is due to the fact that the construction of the multimodal ensemble does not even emanate from the same online speaker.

In contrast to my own opinion, Fay explained to me that the thinking gesture was designed as a signal which she hoped might be seen by her interlocutor across the mobile screen; whilst allowing her time to think in an individualistic sense. From this perspective, the movement represented a gesture for the learner in that she talked about it as a communicative move. In the presentation and the analysis of Fay's subsequent interview data (see p. 158 in chapter 4) my own interpretation is challenged and I, therefore, make a conscious point of highlighting the fact that the learner herself briefly talked about one of these problematic thinking gestures. Having interviewed this learner, I considered that I had overlooked the importance of these gestures because they had been problematic to categorise and analyse. This is because they failed to conform to my chosen definition of gesture or to the categories of gesture which I had chosen to focus on within this particular study (see Kendon, 2004; McNeill, 1992). It is important to note that in choosing a specific and narrow definition and category/ categories of gesture types to explore, the analysis may become more clearly focussed from a researcher's perspective but the approach may

also result in an exclusion of gesture types which learners themselves highlight as important within their interviews. From this perspective, the researcher and learner may assume divergent positions as to which particular movements qualify as gestures. If learners themselves talk about a particular movement, and they are able to explain its function, then the researcher should include these as gestures in their multimodal transcription and/or make a conscious point of highlighting them as a source of interest but also contention.

I am choosing to present Excerpt 5 (see p. 164 in chapter 4) because it illustrates how two learners (Nadia and Simone) briefly negotiate over an incident of non-understanding which is caused by the technology malfunctioning. This section of data challenges the notion that non-understanding within negotiation is inevitably prompted by difficulties which entail language use in isolation, as seen in examples of data from Varonis and Gass (1985). Excerpt 6 (see p. 169 in chapter 4) is included for presentation because it represents the lengthiest and most complex example of the negotiation of meaning across the entire data set. This interaction occurred between two learners called Paul and Lily. This excerpt is particularly interesting because it demonstrates one learner's deployment of an entire sequence of meaning-based gestures which trigger an extremely rich example of negotiation. The data illustrates the difficulties, but also the collaborative opportunities, which learners are exposed to when they attempt prosthetic deictic gestures in order to resolve incidents of non-understanding. The learners' subsequent interview data in Excerpt 7 and 8 (see pp. 184 and 186 in chapter 4) sheds further light on their negotiation in terms of what they themselves thought they were attempting to achieve through their use of gesture at the time of the task. These interview data were also chosen because the learners are able to clearly highlight the types of problems which they experienced within this technology-mediated environment in ways which I found both revealing and insightful. They also talk in personalised ways and provide information which I was not able to fully ascertain within my own transcription and analysis of the same section of negotiation. I include Excerpt 9 (see p. 189 in chapter 4) because one learner in particular struggles to execute a complex series of prosthetic deictic gestures in relationship to her interaction with the wider setting. This example further conveys the complex but critical role of these gesture types in indicating objects to an online addressee but it also elucidates how online peer support is crucial to support learners to resolve non-understanding in multimodal ways. Angela's interview excerpt (see p. 197 in chapter 4) is of particular interest because

she personalises her learning experience beyond a discussion of the technology for its own sake. The learner reveals information which contributed to my own understanding regarding the affordances of mobile devices and the learning potential inherent in these types of informal environments.

In chapter 4, I present and interpret a series of excerpts of negotiation which contain the particular sets of gestures under review. Conversely, learners also exploited a range of beat gestures throughout the data but, as discussed previously, I made the earlier research decision to exclude these from my analysis. I considered that these gesture types were more closely connected to aspects of discourse management, rather than to aspects of semantic meaning, within the negotiation of meaning.

3.12 Conclusion to the methodology

In this chapter, I have outlined the theoretical and methodological framework employed for the purpose of analysing language learners' gestures across Skype VC accessed via mobile devices from beyond the language classroom. The chosen approach involved an integration and adaptation of methods derived from gesture analysis and their multimodal transcription procedures (McNeill, 1992; McNeill & Duncan, 2000) which I introduced into the framework of the negotiation of meaning (Varonis & Gass, 1985). My research rationale is based on the conclusion that communication and language use is multimodal and that previous findings have demonstrated that gestures have been shown to support learners' L2 language acquisition in a number of potentially valuable and diverse ways within their face-to-face learning scenarios. To my knowledge, at the time of completing the research, there had been a lack of previous studies investigating gesture's role within the negotiation of meaning. Furthermore, there had been an absence of studies in terms of a focus on how negotiation operates for language learners via Skype VC across mobile devices from beyond the traditional language classroom. A final aim of the research was to explore the implications of supporting participants to raise their awareness as to their combined gesture and language use in ways which could enable forms of subjective reflection based on verbalization of their multimodal communication. In this chapter, I also outlined a content analysis approach to analyse the interview data (Mackey & Gass, 2012; Gass & Mackey, 2017; Schreier, 2013) with the aim of triangulation and justify the excerpts of data which I am choosing to present.

Chapter 4 The Findings

4.1 Introduction

Within chapter 4, I present the findings from the study (see Lee et al., 2019). This encompasses the presentation of findings from four negotiation excerpts derived from the learners' speaking tasks and I present these alongside the findings from four verbatim excerpts from the stimulated recall interviews, based around the learners' interpretation of the negotiation of meaning. Following the presentation of these findings, I illustrate the overall themes which I identified from the stimulated recall interview data through implementation of a qualitative content analysis approach (see section 3.9 in chapter 3). The negotiation excerpts which I illustrate in this chapter are derived from the data from eight language learners (four dyads) from an overall total of ten who took part in the study. The multimodal transcripts were coded using the Varonis and Gass (1985) framework but instead of an exclusive analysis of the learners' patterns of language use I exploit the format of the multimodal transcript to elucidate how dyads may harness and orchestrate combinations of gesture and speech within their coded negotiation excerpts. In order to understand how speakers forge a relationship between gesture and speech, the components of gesture need to be understood in terms of aspects of their synchrony and co-expressivity with speech (Gullberg, 1998; McNeill, 1992; McNeill & Duncan, 2000; McCafferty & Stam, 2008). As a result, my approach to the design of the multimodal transcripts reflects a methodological concern with gesture-speech relationships but also involves the challenge as to how to reflect aspects of the learners' settings in which their multimodal communication occurred. There have been previous arguments that to understand the meaning of gesture it is necessary to introduce the social and material setting in which the gesture is constructed (Goffman, 1964; Streeck et al., 2011). Moreover, the importance of exploiting and understanding the issue of context, in relationship to the manner in which L2 learners communicate and complete tasks from beyond the classroom, has also been more recently highlighted within studies in MALL (see section 2.9 of chapter 2). From my multimodal analysis of the data from the learners' negotiation excerpts, the unfolding nature of their deployment of combinations of gesture and speech were observed to be consistently shaped and influenced via peer interactions, the learners' interaction with the technology; and via their ongoing engagement with the world around them.

In this thesis, the purpose of the use of analytic multimodal transcripts is that they support my written accounts, consisting of thick description, and aim to enable the reader to observe examples of both written and visual forms of multimodal ‘evidence’ (Flewitt et al., 2014). Firstly, the multimodal design of the transcripts serves the purpose of illustrating the type, shape, and meaning of the learners’ gestures as they occurred within the wider context of the interaction. The written portion of the transcript is designed to illustrate the gestural phrase from its beginning to end in relationship to the learners’ speech (McNeill, 1992); including attempts to analyse prosthetic deictic gestures formed via pointing with the device and camera. Norris (2004) (who draws on McNeill, 1992) suggests that screenshots of interlocutors’ gestures within multimodal transcripts include an illustration of the *preparation*, *stroke*, *post-stroke hold*, and *retraction* (see section 3.7.4 of chapter 3). Conversely, her approach to capturing the entire gestural phrase via a sequence of unfolding visuals proved impractical in terms of the sheer amount of gestures which the learners demonstrated; also with regard to the inherent space restrictions of the paper-based transcript format itself. However, the meaning-based *stroke* is conveyed visually in that this can be linked by the reader to its co-occurring speech, which is shown in bold typeface, in order to illustrate interrelationships operating between the modes of interest. Conversely, due to mediating aspects of the technology, it often became difficult to capture clear screenshots of the learners’ gestural *preparations* and *retractions* and I found that I was more easily able to achieve these within the written portion of the transcripts. The timestamps in the transcripts, included in black typeface, indicate the beginning of the learners’ spoken part of the turn with red typeface used to show the timing of the learners’ gestural *strokes* in relationship to this.

The negotiation excerpts which I illustrate in this chapter consist of microsegments of data and, as a consequence, leave certain elements of the wider task-based interactions undiscussed. This constraint is due to the considerable levels of fine-grained detail required when coding and transcribing the modes of gesture and speech in interplay within a framework of the negotiation of meaning. Conversely, in certain cases, examples of data are presented to include the interactions between the learners which had led up to the negotiation. The rationale for this is due to my observation that it appeared to be important for certain dyads to address aspects of online rapport-building, combined with an ability to situate themselves in a physical and virtual sense, before they appeared prepared to launch into the task or any identifiable forms of negotiation. I also include an example of data

where gestures ceased to be used by both learners within their negotiation of meaning in order to explore why this might have happened in relationship to their use of technology and setting.

To gain a further insight into the data, and to illustrate how I enabled the learners to explore aspects of their own multimodal communication, sections from the stimulated recall interviews around the same set of coded negotiation excerpts are presented following the multimodal transcripts. However, in the case of one dyad (Dyad two: Nadia and Simone), I unfortunately did not have the opportunity to interview them due to one learner returning home unexpectedly and the other becoming ill and leaving the school. The interaction data is included as one learner (Nadia) is communicating via a smaller mobile phone, rather than a tablet device. An analysis of their task-based interaction revealed that problems with the quality of sound on Skype VC can prompt opportunities to negotiate for meaning.

The purpose of a dual approach to the analysis was to attempt to gain a more in-depth and learner-centred insight into the data which I had previously multimodally transcribed. This acted as a way to ideally enrich my analysis, for example, to possibly confirm or disconfirm my original interpretation of the negotiation excerpts. I therefore attempted to uncover further information which I felt could not be easily elucidated from examining the video data of the learners' negotiation in isolation. In analysing and transcribing the video from the tasks, I could usually distinguish what I thought that the learners were attempting to achieve, however, I could not tell what they were thinking or envisage how they themselves interpreted their own and one another's gestures within the negotiation of meaning. Finally, the wider themes or categories which were established during the analysis of the overall interview data are presented and discussed. Adopting a different research lens and overview of the stimulated recall interview data enabled me to identify trends which came to light; and enabled me to interpret and discuss the data as a whole in relationship to the research questions which I had posed.

I next introduce the first dyad and present the transcript of their negotiation excerpt followed by a section from a learner interview based around this same section of data.

4.2 Dyad one: Fay and Andrea

The first learner who took part in the study is called Fay. She is twenty-seven, from Kazakhstan and is a student of finance in her home country. Her interlocutor is a student named Andrea who is from the German-speaking part of Switzerland. She is twenty-three and has been mainly studying academic English in her formal classes but had requested opportunities to use language in a more communicative sense. They both represent confident speakers who had originally been placed at Upper Intermediate level for the purposes of learning within their formalised classes. At the time of the study, Fay had been at the school for four months and Andrea had more recently arrived in the UK (six weeks). They both knew the city well and had informed me that their favourite spaces had consisted of an English stately home and a public garden. I took note of this information, designed tasks around the different locations, and sent each learner to a setting which they had not previously visited.

For the purpose of undertaking the speaking task, Andrea is located in the stately home and Fay is situated in the nearby garden location. The dyad is initially encouraged to take a few minutes to relax, to order some food and drink, and to read through the task instructions. The dyad is left to introduce themselves and to begin the task only when they feel ready. It is notable that within this dyad my task instructions are left largely ignored. Both learners fail to consult the various task steps and instead decide to pursue their own communicative and social agenda. To my surprise, their more improvised and serendipitous approach did not appear to affect the creation of opportunities to negotiate for meaning. Andrea exploits a 2-in-1 device and Fay uses my tablet which had the video-capturing app Supertintin pre-installed in order to video record both the learners' screens. During the process of the Skype VC interaction, Fay and Andrea appear keen to become online friends with rapport quickly developed. They had first been briefly introduced to one another at the research meeting. At the beginning of the exchange, Fay waves at Andrea across Skype as she responds with warmth and enthusiasm to her simple social introduction. The co-presence established between the learners appears to break down any potential lack of social presence across the technology. This is achieved through multimodal moves such a friendly wave, the informality of Fay's verbal responses; and the sharing of smiles and laughter. The initial online rapport-building is possibly why these learners subsequently

felt able to engage in the negotiation of meaning in that trust appeared to be present from early on in the exchange (see Figure 12, shown next).



Figure 12: The building of online rapport before negotiation occurs

Instead of waiting for Fay to offer detailed descriptions of the series of relevant artefacts, which she is to locate from within the stately home, Andrea instead spontaneously moves around her space as she affords her interlocutor a virtual tour of the stately home. She pursues her own agenda via the deployment of the mobile device, at first in a rather hurried way, and later with more deliberate consideration as to the significance of each of the artefacts at which she points with the device and camera. As a consequence, Andrea has ignored the ‘splitting up’ of the relevant sources of information as she never describes the relevant items which Fay is to locate from her garden location across the entire process of the task. As a consequence, I observe that Fay sometimes assumes a less active role in terms of her use of the device in a portable manner, for example, she remains seated for quite long periods throughout the task in a similar manner to a desktop videoconferencing interaction.

Within the early phases of the interaction from the stately home, Andrea does not exhibit any clearly-definable gestural phrases with the device: there are no *preparations*, *strokes*, and *retractions*, as depicted in the unit of analysis and transcription procedures around gesture cited in the previous chapter (see section 3.7.4). Conversely, she chooses to swing her entire body as she pans her setting via the affordances of the mobile technology, for example, exploiting its portability and the enabling of video images across Skype VC (see Figure 13). Images are sometimes blurred because the device and camera are kept moving by the learner with no time taken to make the necessary gestural *preparation* and *stroke* in order to highlight a specific referent in her environment through prosthetic deictic gestures.



Figure 13: 'This is the place!'


However, Andrea does prioritise the sharing of her setting whilst attempting to simultaneously present herself in the corner of the Skype screen (see Figure 13). I considered that Andrea may have felt that it was important to ensure that her learner 'self' was to be maintained as an integral part of this online exchange. Furthermore, she is perhaps attempting to monitor the video stream in order to receive a reaction from Fay as she pans the setting of the stately home. It becomes apparent during the analysis of the 'lead up' to the negotiation that Andrea grasps the importance of situating her virtual interlocutor within her material world via her sharing key aspects of this space across the affordances inherent in the mobile technology. She remarks in an exclamatory tone: "This is the place!" but in adopting this interpretation of the task she has literally 'given the game away' and ensured that her entire setting is revealed to Fay via a sweep of the device and camera. My task rationale and design had foreseen artefacts and meanings gradually negotiated and visually shared between the dyad over the course of the task but this did not happen in this instance.






Andrea's spontaneous virtual tour, and her seeming lack of adherence to the task instructions, still resulted in an example of the negotiation of meaning (see Excerpt 3) which is presented next. I have chosen to present this particular negotiation excerpt as it demonstrates how prosthetic deictic gesturing through a learner's use of a mobile operates between learners; and illustrates how this gesture type can function as a trigger to engage learners in the negotiation of meaning online. The excerpt also includes a complex series of multimodal ensembles (Jewitt, 2011) which are achieved via combinations of the learners' hands and use of language. Therefore, the negotiation is both stimulated and

supported through the collaborative efforts of the learners; despite my observation that there is no complete breakdown in their communication. In this instance, the communicative issue is caused by the need for vocabulary (Wang, 2006). In this example, it is unknown language items which form the basis for the learners to negotiate for meaning in a range of quite complex ways. It may be important to note that in relation to Varonis and Gass (1985), Excerpt 3, which follows, would be classified as “*embeddings in negotiations of meaning*” in that the negotiation patterns are extended (pp. 78 - 79). This excerpt shows how aspects of gesture, speech, technology; and setting tend to overlap and converge as they contribute to the forward trajectory of multimodal negotiation.

Regarding Excerpt 3, there was general coding agreement between myself and the second coder regarding the decisions made. For example, taking account of the learners’ gestures and speech as one unit of meaning did not alter or challenge my original coding decisions based on the analysis of the learners’ speech turns alone. This may indicate that gestures are an integral part of learners’ speech and, therefore, the framework of negotiation allowed me to take account of both modes operating together for learners. Conversely, areas which did cause problems, in terms of how the Varonis and Gass framework functions from a multimodal perspective, included disagreement as to whether or not my category of a ‘comprehension check’ could be extended and coded as a visual phenomenon. I describe this phenomenon within my description of the coding process within the previous chapter (see section 3.8.2) and my realisation of the requirement to devise a new multimodal coding category to incorporate into the existing framework: Visual Comprehension Check (VCC). There are a few instances of learners exploiting gesture without accompanying language use across all the analysed excerpts within this study, however, I gradually realised these isolated gestures did relate to speech as they occurred either just before or after a spoken turn and so I made coding decisions around these. As the study is exploratory, and there were no previous examples to draw on or to compare with in the existing literature, I include my original coding decisions in the analysis; acknowledging any potential limitations. I next present a multimodal transcription of the negotiation of meaning (see Excerpt 3):

Excerpt 3: Negotiation of meaning via Skype VC on mobiles from stately home and garden

Turn	Gesture & speech	Coding	Visual transcription
1: A (00:06:21)	So there are many books there as well.		 (Andrea pans her setting with the device)
A (turn 1 continued) (00: 06:26) 2: F: (00: 06:32)	There is somewhere where you can light a fire. Oh that's lovely! Very very British. Charming!		 (Andrea appears in the corner of her screen)
3: A: (00:06:37)	There is  [a beautiful head . I don't know if you see?	VCC/T	 DEICTIC (PROSTHESIS): Stroke (00:06:40)
4: F: (00:06:43)	I can see it. It's a ...	I	
5: A. (00:06:45)	You see it?]	VCC/T	 Andrea has retracted her gesture from the object
6: F. (00:06:48) 7: A: (00:06:51)	[Um] [How do you call it?] [<i>Gesture no speech</i>] I don't know the name.	T R	 ICONIC: Strokes: (00:06:49) (00:06:50) (00:06:51)

8: F (00:06:54)	[<i>Gesture no speech</i>]	I	 ICONIC: hands beyond boundaries of screen (00:06:55)
9: A (00:06:56)	I don't know how what's the name.	R	
10: F (00:06:57)	[Horn] [horn]	I	 ICONIC: Fay's hands move closer within boundaries of screen (00:06:57) (00:06:59)
11: A (00:07:00)	[Like the wood]	R	 ICONIC (00:07.01)
12: F (00:07:00)	[Horn]	I	ICONIC: Fay repeats gesture from turn 10 (00:07:00)
13: A (00:07:01)	[Yeah]	R	ICONIC: Andrea repeats gesture from turn 11 (00:07:01)
14: F (00:07:03)	[Like eh] I don't know. And how do you say in Eng [ah deer] deer? deer.	T	 ICONIC: (00:07.03) DEICTIC: (00:07:06)
15: A (00:07:08)	Deer?	I	
16: F (00:07:09)	[Head] [Head of] [deer]	R	 ICONIC: Hand within boundaries of screen. (00:07.09) (00:07.10)
17: A (00:07:11)	Ah Yeah	RR	

Following a fleeting glimpse of the setting, through Andrea's camera panning various situated artefacts such as books and an historical fireplace, Andrea next triggers the previously shown negotiation excerpt (Excerpt 3). She achieves this move via harnessing and then co-ordinating a set of available multimodal resources (Hampel & Stickler, 2012). These are observed to include her use of language, her prosthetic deictic gesturing with the device; and an ongoing evaluation of the mediating aspects of Skype VC in relationship to Andrea understanding how best to convey her material setting to her interlocutor. In terms of the previously shown excerpt, Andrea generates a negotiation trigger through deliberately standing in front of the relevant hunting trophy which is positioned above a door facing her. She synchronises her prosthetic deictic gesture, formed with the camera and device, in direct relationship to her conveyance of this specific artefact (turn 3). Andrea's gestural action in this case is forged through her extending her hands and the device at a relevant aspect of her world: this action is not random but is both co-expressive and synchronised in timing with her use of the noun 'head.' Both the gestural *stroke* formed with the device and her co-occurring language use are well-organised in ways which were identified to represent her effective construction of the GP in that the key word was synchronised with the gestural *stroke*. This co-ordination of meaning was identified in terms of the learner's exploitation of gesture and speech. She orchestrates the two modes together in order to construct a wider holistic unit of meaning which triggered the subsequent negotiation. This particular deictic gesture formed with the device could be transcribed in accordance with the gestural phrase within my adapted chosen multimodal unit of analysis and transcription scheme (McNeill, 1992; McNeill & Duncan, 2000). This is because Andrea clearly extends her hands and the device in *preparation* and then, with the hands fully-extended, she forms the *stroke* with the tablet device and camera as these remain static enough in order to enable her to highlight the relevant object clearly enough to convey its meaning to her interlocutor, Fay.

Andrea indicates that the animal trophy is considered worth showing to an interlocutor through her choice of the adjective 'beautiful' which she uses to describe it. The learner also appears to understand that as she stands and consciously focuses the device and camera on this particular object, she must at the same time check her interlocutor's visual comprehension from the other location. The act of consciously pointing with a mobile device and Skype VC appears to entail the learner first interactionally positioning their entire body and next preparing and extending their hands and the mobile technology. This

act enables the possibility to draw attention to an artefact located in one setting in order to convey this virtually to another learner who is situated in a geographically-separated setting. Andrea introduces the co-referent via her gestural *stroke*; and this is skilfully achieved through her exploitation of the combined affordances of the device and camera as she draws attention to the hunting trophy whilst describing this.

The dyad is required to achieve joint attention as they ensure that they co-reference the same artefact with the shared aim to establish and maintain levels of intersubjectivity across VC to enable them to talk about this. As a result, Andrea needs to continually support her interlocutor to see and comprehend her material setting (the stately home) through her careful manipulation of the technology in relationship to a sharing of her setting. As a learner, Andrea appears to be adept at this in that the co-referent (the hunting trophy) immediately appears as a sharp image on Fay’s screen without indication of shaky hand movements or blurring of the video. Andrea also achieves the prosthetic deictic gesture with the device in her first attempt; rather than having to make several efforts to form the critical meaning-based gestural *stroke* (see Figure 14) to enable the image to be seen on Skype by an interlocutor. Andrea demonstrates a continued awareness of the co-presence of her interlocutor by also checking information multimodally as she points. This is indicated through her asking her peer for visual feedback via a visual comprehension check (VCC): “I don’t know if you see?”

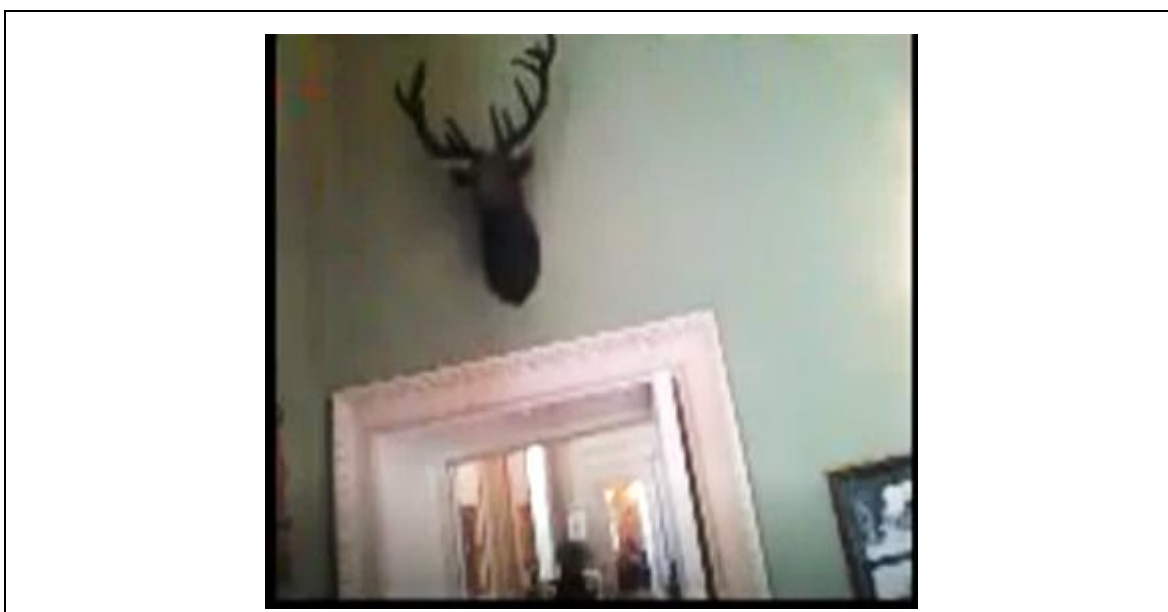


Figure 14: An accurate prosthetic deictic gesture formed with a tablet

Andrea's deictic gesture with the device is focussed on the trophy and is strategically sustained across her own but also her interlocutor's subsequent speech turns (see turns 3 - 5). She appears to be aware that her peer, Fay, is now engaged in attempts to talk about what she can now see on her mobile screen from the setting of the garden. In this instance of negotiation, Fay does not query a particular word in her indicator move (turn 4, see Excerpt 3) but instead considers what the object might be called. The multimodal nature of the learners' turns highlights the embodied, gestural, and visual aspects of an exchange via Skype VC on mobile devices when learners are pushed to establish a common frame of reference in order to begin to negotiate for meaning around vocabulary items. Conversely, from turn 5, the first levels of modal dissonance start to emerge within the online exchange. The communicative issue is driven by Andrea's subsequent lack of co-ordination between her speech mode in relationship to her pointing with the device as levels of intersubjectivity are no longer maintained between both learners. For example, whilst Andrea checks Fay's visual comprehension again in turn 5, at the same time she chooses to retract her prosthetic deictic gesture which had been previously directed at the co-present trophy. In contrast, Fay appears to want to pursue the negotiation of meaning in order to find a solution to her quest for missing language in order to express what she had seen on her mobile screen. Andrea physically turns her body to face her device and walks away from the trophy. This move could have put an end to the negotiation but the missing object is next re-represented via a series of meaningful iconic gestures formed through Fay's hands. This learner executes her gestures with both hands as she demonstrates an ability to pursue negotiation through exploiting carefully-crafted ensembles of gesture and speech. This multimodal orchestration is achieved through Fay's hands remaining free to move; and this ability to convey meaning is the direct result of her earlier decision to position the mobile device on a table rather than to choose to hold it in her hands. Fay's first set of gestures were identified to serve the role of forming a replacement for the virtual artefact which had just disappeared from her mobile screen (turn 6). Whilst Fay attempts to name this artefact, she is observed to co-ordinate her determined search for language with a sequence of co-occurring repeated gestures. Even before Fay defines her linguistic goal beyond an '[Um],' she has already created a gesture, however, it was not entirely clear what the gesture was designed to represent. Fay next continues to pursue the negotiation with a second gesture in which the *stroke* co-occurs with the relevant word '**call**' as she appears pre-occupied with searching for vocabulary items within the negotiation of meaning. She may also be attempting to hold the floor on Skype as she

appears to require further time to think about the language which she needs in order to describe the relevant artefact. This is illustrated in Figure 15, which follows.

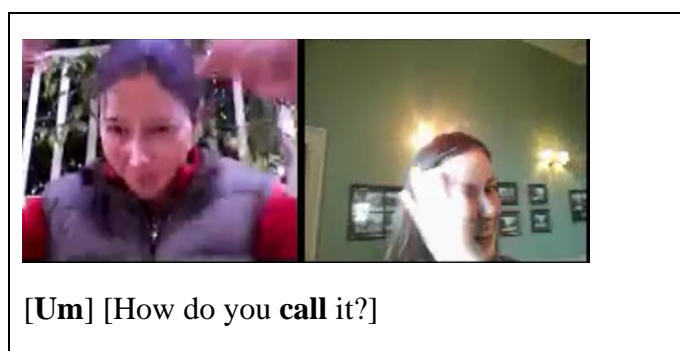


Figure 15: Fay searches for vocabulary through gesture and speech

Fay's repeated gestures in turns 6 and 8 caused me problems in the analysis as the meaning-based *strokes* could barely be seen within the boundaries of the mobile screen. I also concluded that the gestural movements in Figure 15 were somewhat vague, for example, I could not distinguish from the video whether Fay's hands were expressing parts of an animal's anatomy or alternatively the artefact's spatial positioning on the wall within her interlocutor's setting. Rather than referring to an abstract concept or idea, the gesture appeared to be reflective of the object in some way and for this reason I concluded that it is probably best described as iconic. However, Fay's use of language at this point is abstract in that the precise meaning of "it" is as yet unknown with gesture and speech operating independently rather than coming together in co-expressivity to form an entire unit of meaning for an interlocutor. From a communicative perspective, the gestural *preparations* and *retractions* are partially lost in that they are formed below, left, and right of the restricted boundaries of the mobile screen. I was not sure from my analysis of the video whether Fay's gestures were primarily designed to support her individualistic thinking or whether she was appealing for peer assistance across the technology.

Fay makes an attempt to gesture without speech in turn 8 but this time her hands are lost due to the constraints of the technology. In the analysis, Fay's next set of gestures were categorised as iconic (see Figure 16 shown next) as they represent a deer's antlers when analysed in combination with her co-occurring linguistic attempts: [**Horn**] [**horn**]. Her gestures reflect McNeill's GP in that this word occurs at the point where the learner forms the peak meaning of the gesture via the deployment of her gestural *strokes*. In turns 10-12 there are three iconic gestures from Fay, repeated with each reiteration of this key word.

The meaning-based *strokes* are perfectly co-timed to appear on Skype VC with each of the learner's verbal utterances and are visible, even though the hands are partially hidden from view during the *preparation* and *retraction* phases.

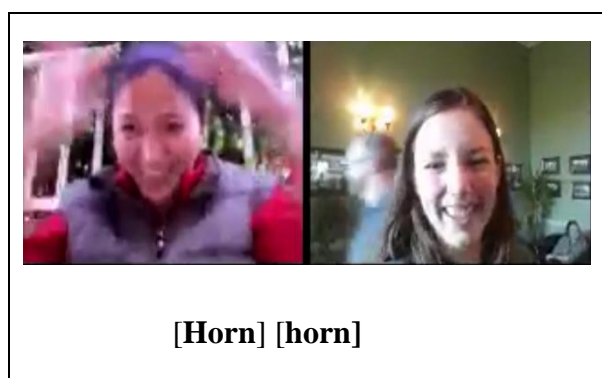


Figure 16: Fay gestures the animal's antlers

Although it appears as if Fay's multimodal efforts have been previously ignored, in Figure 17 (shown below), they do appear to result in a more engaged response from Andrea. For the first time, she mobilizes her own set of multimodal resources and responds with an enthusiastic left-handed antler gesture (turn 11) which is iconic in type. She continues to carefully balance her device in her right hand whilst walking, talking, gesturing and smiling as she demonstrates considerable levels of skill and co-ordination.



Figure 17: Andrea gesturally mirrors the concept of antlers

Andrea's deployment of gestures formed with the hands may be a signal to Fay that she now acknowledges the shared goal of negotiation and is indicating that she has assumed a more participatory role. The intersubjectivity established between the dyad, through the mirroring of their gestures with the hands, becomes a source of amusement but also represents an important strategy for the merging of shared worlds. Their gestures appear to

act as visual displays of the intentionality of the learners to continue to negotiate across technology and geographically-dispersed settings in order to solve the problem.

Andrea constructs her own unique representation of the artefact as she bends her fingers forward to form an iconic gesture; and this is designed to signal the jagged shape of the animal's antlers. She is perhaps more adept at ensuring that her gestures are always visible across the technology but her mobile screen is larger than Fay's; and she is consistently holding it in one hand which enables her to continuously manoeuvre its position in relationship to the deployment of her hand and body. At this point in the negotiation, it is the animal's horns or antlers which have been identified as the source of the communicative focus between the learners, however, Andrea chooses to focus on the material nature of the object's horns (turn 11) in noting that these resemble wood. From turn 14, Fay is now asking and answering her own questions as she negotiates with herself: 'I don't know. And how do you say in Eng ah deer deer? deer.' This may indicate that her gestures at this point could be serving a role in mediating her inner thought processes in ways which reflect forms of 'private speech' (Vygotsky, 1978; McCafferty, 2004). The meaning of Fay's iconic gestural movements transforms once again in keeping with her co-related shifts in linguistic foci and in terms of the forward trajectory of the negotiation (turn 16). As the language use moves away from the focus of the antlers and towards the issue of the animal's head (see Figure 18), the co-expressive use of gesture moves downwards towards the learner's own neck. Whilst Fay verbalises the idea that it is the head of the deer which is of interest, she executes cutting movement gestures in ways which may suggest the wider concept of hunting.

The iconic gestures appear to add a layer of meaning (McNeill & Duncan, 2000) which is not immediately made apparent from an analysis of her language use in isolation.



Figure 18: Fay gestures the head of the animal

To summarise, this negotiation excerpt was initially triggered by Andrea's prosthetic deictic gesture alongside her speech, however, it was sustained by Fay's hand gestures and speech which were mediated in relationship to her identification of the affordances but also the constraints of the mobile device and Skype VC. Andrea's orchestration of modes cannot be understood without acknowledging how the learner synchronises her behaviour with the device in relationship to her recognition of the possibilities which are inherent in her setting as she triggers and, therefore, enables the negotiation to occur. Fay is an active and autonomous learner in that she expresses considerable interest in discovering the appropriate language to describe the virtual image of the hunting trophy which has been relayed across her mobile screen by Andrea. She conceptualises the image of this highlighted object as an opportunity to explore its linguistic meaning; and she attempts to uncover the required vocabulary through her continued co-ordination of iconic gestures and speech as she progresses through a series of possibilities. Fay moves closer towards her linguistic goal via exploitation of the resource of the physical hands, operating with her speech, but also in her possible recognition as to how her gestures are mediated in various ways by aspects of the technology. The majority of her gestural *strokes* can be seen but, possibly due to the limited size of the tablet screen and where the device is positioned, her *preparations and retractions* are lost in a gestural space which is located beyond the boundaries of the screen.

Andrea demonstrates skill via her use of the mobile technology in a portable sense, for example, she uses the device and camera to accurately target salient aspects of her setting

in order to share them with Fay. Nevertheless, the urge to negotiate for meaning is not present to the same extent in this learner (although the formal learning background took place within the same institution). Andrea is honest in stating at the start of the negotiation that she simply does not know the word to describe the object, therefore, she may have seen little purpose in the continued pursuit of the negotiation of meaning. It is only through her interlocutor's persistence that she later adopts a more participatory role and this is evidenced through her harnessing the use of modal resources to offer further forms of linguistic but also gestural input for Fay.

In the following section, I present a verbatim extract from Fay's interview as she is asked to comment on Excerpt 3 (presented above in multimodal transcription format). I began the stimulated recall as an opportunity for both of us to review the excerpt. Each individual gesture was paused on by myself with an opportunity for the learner to comment following my questions. Fay was also encouraged to press pause on any of the gestures on which she would like to comment. The reason for presenting this example of interview data (see Excerpt 4) is that Fay is able to give a comprehensive account of the role and meaning of her gestures as she herself perceived them operating throughout Excerpt 3. The gestures she discusses in this part of the interview are constructed with the hands, rather than in her pointing with the device. As mentioned previously, the iconic gesture within turn 6 had been repeated by Fay but had caused me problems in the analysis. The language was abstract and the meaning of her gestures remained unclear. The negotiation had seemingly been resolved according to the Varonis and Gass (1985) framework, however, following the transcription phase I concluded that their coding scheme left certain gaps in the analysis and could only form a part of the wider picture. In this interview excerpt (Excerpt 4), the italics represent the learner speaking with the bold typeface and italics used to illustrate where the researcher spoke.

Excerpt 4:

Can I ask what you were gesturing here...what do you mean by that gesture?

When I was pointing about the antlers?

Yeah

Yeah, I was when I saw the object I couldn't find the name of the words so my first gesture was to find out what was that the name of the animal you know.

Was it for you?

It was for me to think yeah it was for me to think but at the same time as she saw my screen that I was thinking I tried to make a connection with her. What was the name what was the name of the animal. Eh so because it has antlers I couldn't find the word so I have to show it to her. You know what is it called what is it called. I was trying to eh kind of to share it together so we could find the word together.

...It's a good learning strategy to try to work out the words.

It would be easier for both of us to work out the words. You know sometimes you know the words but you have to search in the head you know to find it. Maybe someone would be quicker than you are and so it was the pointing so she would also remember and search at the same time with me about these antlers.

...So these are the antlers rather than the position on the wall. I wasn't sure whether you were indicating the position of the head on the wall but I see now that maybe yeah these were the antlers.

Yeah, I was pointing to the antlers with my head to the outside kind of with my hands.

So you started at your head and then moved out and then came back in again I think three times.

I did. I did actually because I was searching for words and I was hoping that Andrea would help me as well, you know. But we didn't succeed in that. That was OK.

So if we move on... Did you notice that?

Yeah, she ...she didn't stay for long. She didn't discuss with me about this word and if she did pay more attention and stayed on and listened properly then we would get the answer quicker and I wouldn't be looking for antlers in my vocabulary (laughs).

You got there. It's fine. You kept going with your gestures. So here we go again!

...Now I see that she did copy it you know so it was good that I got her attention to this word and tried to find it together. She mimicked it so it that was good. That means she understood me in a way kind of...

What did you mean by that gesture?

What I meant is that I didn't know the word again. How do I say in English. Later when I saw it it is hunting prize. When I pointed to my neck I was explaining that it was the head of the deer as hunting prize. Yeah?

Exactly.

But before I didn't know so I wanted to say it was the head of the deer. Instead of explaining that it was hunting prize...just the head not the full body... in a hunting prize.

Or hunting trophy.

Yes, that's another word. Thank you.

When did you find the word... was it just after the task?

Well it was interesting because eh at first I didn't know. I knew that there was another word for horns rather than horns and the second I forgot I know there should be some words explaining this exact hunting prize or hunter trophy as you just told me. I wanted to know what was the word and to find out what was the description of those words. So that was what I'm interested and so I had to look out for it and so I did.

OK. Yeah, it's really good you wanted to watch back the video. So you asked me that. You could use the video yourself or you could work with a teacher together and review the video and find more language. Is that how you would use it?

Yes, I would use it in that way especially um from a grammar point of view to see my mistakes to see how fast I can find those vocabulary. Now I'll never forget antlers and I'm sure those hunter trophies. In the night even I'll wake up and I will remember ANTLERS and so that helped me a lot and teachers would also explain the situation and the structure of the sentence. All this stuff it's very useful from a language point of view. I think it's extremely good...gesture adds more power into the language.

As a result of examining Fay's reflective verbalizations, I felt that she demonstrated a considerable degree of 'multimodal communicative competence' (Royce, 2013). This is evidenced through the awareness which she demonstrates in her understanding of the deployment of gesture with speech as she discusses the interplay operating between these modes in terms of her expression of meaning and intent. Awareness of multimodal forms of communication are also illustrated in Fay's ability to verbalize in detail, how, but importantly, why, she chose to use gesture during Excerpt 3. Fay's ability to make sense of aspects of her multimodal communication is indicated in her consistent recognition of the purpose with which she exploits gesture and speech to express holistic meaning which appear to principally relate to her self-identified learning goals. With minimal prompting, she is able to forge her own connections between her use of specific gestures and the ongoing requirement to find adequate language in order to describe the artefact of the hunting trophy. She elucidates and confirms to me that her use of multiple gestures throughout Excerpt 3 had assumed a dual role for her: internal (for herself as the individual speaker) and communicative (for the addressee) (Gullberg, 2010). Fay explains that her gestures are representative of her thought processes but that they assume a simultaneous role in supporting her attempts at connecting with a peer via an appeal for assistance across Skype. As a result, searching for missing words is acknowledged as a team effort or competitive race to the finish line: "Someone would be quicker than you." This is also indicated in the evaluation from Fay that "then we would get the answer quicker," suggesting that this particular learner feels a pressing need to achieve results and, therefore, expects to see a concerted effort from their online peer.

In terms of an identification of the affordances of the technology, Fay illustrates awareness of the presence of this in her mention of the word "screen," noting how Andrea would be able to see her thinking. She is referring to her thinking pose in turn 5, which did not strictly qualify as an iconic or metaphoric gesture but, for the learner, represented an important visual signal, designed to convey intent to an online interlocutor. Fay is able to comprehensively explain how her gestures are intimately linked to her linguistic meaning and goals. For example, it only becomes evident during the course of the interview that the previously unknown word "antlers" is now present in her lexicon. I observed that she talked extensively about her gestures based around this key word but that it had not been available to her at the time of the task. There was no deployment of this critical word

within the actual negotiation of meaning at the time of the task. Fay's interview also sheds light as to how gestures can be harnessed by a learner to enable an interlocutor to understand language goals which are conveyed in multimodal ways. Fay interprets her interlocutor's response gesture with the hands as a sign of attempts to build intersubjectivity but she considers that Andrea: "didn't stay for long." Her evaluation of their negotiation process highlights that Andrea chose to walk away with her device and retracted her prosthetic deictic gesture whilst the hunting trophy was still under negotiation. Some possible frustration is evidenced in Fay's evaluation of the quality of the negotiation within the conclusion: "we didn't succeed in that." Fay highlights the limitations of my chosen framework, according to Varonis and Gass (1985), in that the coding decision in the transcript of Excerpt 3 (see RR, turn 17) had signified that the non-understanding had been resolved between these two learners.

As a motivated language learner, Fay has assumed her own autonomy in both setting and achieving her personalised learning goals. She reveals that she actively remembered that a more precise description of the representation of language which she had seen on her mobile screen during the time of the task was later required. She also draws my attention to her newly acquired knowledge: "at first I didn't know." I knew that there was another word for horns rather than horns [...] Now I'll never forget antlers." Fay indicates language awareness in acknowledging that this particular vocabulary item was beyond her reach at the time of the task. It becomes apparent from the interview that she actively followed up the problem of her own volition. This reflective stimulated recall session allowed us, as learner and researcher/educator, to work through the multimodal negotiation together. Through our dialogue, regarding gestures which occurred within Excerpt 3, it became possible to talk further about specific language items which had emerged as a result of studying the captured video data together. For example, Fay confirms to me that the gestures based around her neck were representative of the previously absent language item: "hunting prize." This revelation from the learner creates a pedagogic opportunity for me to offer the further suggestion of: "hunting trophy." Fay is an enthusiastic and skilled language learner as she acknowledges the vocabulary suggestion and then attempts to use this newly acquired language within the interview itself. During the interview, I make suggestions as to how she might consider further exploiting the resource of the video in order to further address and support her learning needs. During the analysis and transcription of Excerpt 3, I realised that there was a general lack of attention to form

observed between this dyad within their negotiation. Conversely, Fay suggests that the video itself could act as a reminder of her grammatical inaccuracies and could be used as a means to further pinpoint where problems in recalling certain vocabulary items occurred during the negotiation of meaning. The speed of her ability to recall vocabulary items is once again highlighted as important to this particular learner. Furthermore, Fay offers possibilities as to the role of the teacher following this type of task from beyond the classroom: “teachers would also explain the situation and the structure of the sentence.” She emphasises the importance of a language learner understanding both the wider context and the structure of language. From this perspective, she has possibly been well-trained during her classroom lessons as to the importance of the learner forging connections between meaning and form. It is possible that the multimodal nature of the language use and the ongoing struggle to arrive at the word ‘antlers’ from beyond the classroom may have helped Fay to embed the language in her mind. For example, Fay comments on her potential acquisition in a humorous way: “I’ll wake up and I will remember ANTLERS.” This particular learner concludes that “gesture adds more power into the language;” and it appears that her primary concern in discussing gesture and speech is based around the notion that gestures could support her talk and acquisition of language.

4.3 Dyad two: Nadia and Simone

Nadia is twenty-six and Simone is twenty-two. Nadia is studying finance in her home city of Rome. Simone is studying hospitality in Spain. Both learners have been at the language school in the UK for approximately three months. Nadia is positioned outside a museum whilst Simone communicates from within a café space. To my surprise, Nadia chooses to remain outside the museum and never goes inside to share the objects; possibly because the signal was not very strong. She assumes a passive role in terms of learner mobility and the use of her mobile phone in terms of its potential for portability.

I chose to present data from this dyad as Nadia communicates on a mobile phone (the smallest and lightest screen deployed within this study) whilst Simone uses a tablet device (Excerpt 5). I have also chosen to present this section of data because it falls into the category of a routine without any embeddings or extension, according to Varonis and Gass. However, unlike the examples of data illustrated within chapter 2 (see sections 2.7 and 2.8) the trigger to this negotiation routine prompts the learners to adapt and clarify their language use for an atypical reason. This particular dyad experience problems with the sound quality on Skype and, as a result, I observed that they did not modify their language

use or employ gesture in relationship to factors which related directly to issues around learners' linguistic understanding; as is assumed in interpretations of the Varonis and Gass coding model.

At the beginning of the task, Nadia and Simone do not engage with the negotiation of meaning. Nadia asks questions as she assumes the role of a customer and Simone pretends to be the manager of the café. This idea was not in my original task design, but the learners adapt the exchange in ways which appeal to their sense of fun, as seen in Figure 19.

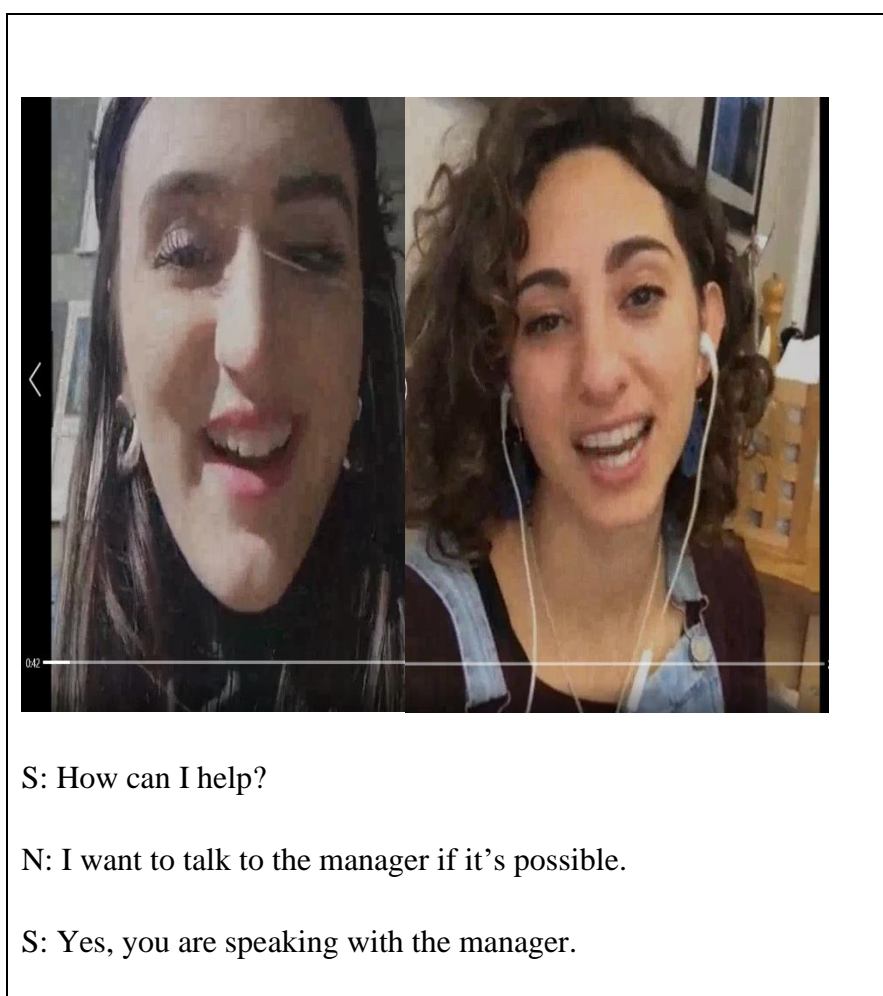




Figure 19: Nadia and Simone assume imaginary roles

Next, I present a brief and linear negotiation of meaning routine which occurred between Nadia and Simone within their speaking task (see Excerpt 5, which follows). The use of gesture and speech is not the result of a requirement to clarify and modify language use but is instead reflective of technical problems in the communication.

Excerpt 5: Negotiation of meaning via Skype VC on mobiles from outside museum and cafe

<p>1: N (25:54)</p>	<p>I [want to ask you about... (inaudible)]</p>	<p>T</p>	 <p>DEICTIC (FINGER) (25:54)</p>
<p>2: S (25:57)</p>	<p>Excuse me. Can you repeat...please [because the sound]. [The sound is not very good]</p>	<p>I</p>	 <p>DEICTIC (HAND) (25:58)</p>
<p>3: N (25:59)</p>	<p>I want to ask some questions about the menu.</p>	<p>R</p>	
<p>4: S (26:01)</p>	<p>Ah, OK.</p>	<p>RR</p>	

In the first turn, Nadia triggers a routine in that she asks a question which is partially inaudible for Simone due to the sound quality on Skype VC. Nadia is communicating via a mobile phone and, despite the smaller screen size, the use of her deictic gesture is enlarged and is clearly visible on the screen. The close-up appearance of her deictic gesture differs from the views of gestures when obtained through learners' use of tablets and 2-in-1 devices. Nadia's deictic gesture, when taken into account with her co-occurring language use, creates a trigger and indicates that she is asking something of her interlocutor: "you." In this case, both learners experience non-understanding in ways which can be attributed to the sound quality of the Skype technology; rather than to any specific difficulty with understanding particular items of language. As a result of the technological difficulties at this point in their exchange, Nadia's words disappear into cyber space (turn 1). In Simone's subsequent indicator move, she asks for clarification via a polite: "Excuse me. Can you repeat please because the sound. The sound is not very good." During her utterance, Simone moves her entire hand upwards in order to point towards her left ear. She forms two successive deictic gestures where the *strokes* co-occur with her use of the key word sound. She also moves her face and body close to the screen. Simone's deictic gestures are designed to compensate for the lack of ability for both learners to clearly hear one another; and to perhaps ensure that her visual message across Skype overrides the message conveyed via her spoken language use. However, Simone does not use gesture in a mimetic sense but instead synchronises her deictic gestures to co-occur with, and to emphasise, the key part of her verbal message. Following the indicator move, the sound problems on Skype disappear, as the connection improves, and the conversation is enabled to continue. The resultant resolution within the routine is achieved through language use in isolation. The original conception of the Varonis and Gass framework does not cater for overt indications of non-understanding being caused by a failure of the technology and the requirement to compensate for this difficulty in multimodal ways. In this instance, gesture and speech are not used to clarify a problem in understanding specific items of language but are instead driven by technical issues which can arise when learners negotiate for meaning across Skype.

4.4 Dyad three: Paul and Lily

Next, I present an excerpt of negotiation (Excerpt 6) from dyad three which is derived from the speaking task which occurred between Paul and Lily. Paul is twenty-two and Lily is twenty-six. Paul is from Columbia and is a law student back in his home country and is studying in the UK for the summer period. Lily is twenty-six and is from the Italian-speaking part of Switzerland where she is studying psychology. Paul and Lily had studied at the school for approximately two months and had been evaluated at Upper Intermediate level. For the purposes of the task, Paul is situated in a bar/café location and Lily in a well-known Bohemian café in the city. These choices of location are due to the dyad expressing an interest in and affection for these particular places. As with the other dyads, each learner was sent to a location which they had not previously visited. Lily decides to access Skype VC on her tablet device whilst Paul exploits my own tablet. Without prompting, he also carefully places his mobile phone on the table next to his tablet, and, in setting up the task, I wonder what his intentions are with regard to this second device.

The learners are encouraged to order what they would like to eat from the different locations and asked to sit at a suitable table. Lily asks me to explain some of the language in the task instructions before the task takes place but I instead encourage her to talk to Paul online about this, as I feel he will be a supportive peer. In contrast to Excerpt 3, shown previously, the task instructions are carefully adhered to by this dyad; and this is how I had envisaged the task operating. I notice that Paul actively scaffolds this next negotiation of meaning (see Excerpt 6) as he continually tries to support his peer in a variety of important ways. This support includes keeping his interlocutor ‘on task,’ the use of encouraging language to ensure that Lily continues to explore her location, and the learners’ deployment of shared forms of multimodal communication.

Before beginning the task, the dyad realise that they have forgotten to order some food and drink; and Lily abruptly leaves her chair but carries her device with her as she chats about the available menu options downstairs. She decides to order a muffin with some coffee which she then brings back to her table upstairs. Once again, some initial small talk and sharing of their respective menus across the two separate locations appear to act as an icebreaker and a way to build online rapport. This dyad converse and point with the device before the task begins and the illustrated negotiation is triggered. The small talk appears to engender a friendly and pleasant atmosphere in order to enable the necessary conditions for the task and

negotiation to occur. Moreover, these learners exploit this opening social exchange as a way to informally practise pointing with the device and camera in relationship to their respective settings before they officially begin; indicating that they understand what is required of them to complete the speaking task (see Figure 20, shown next).



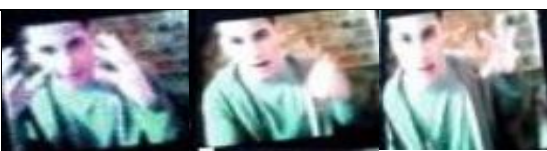



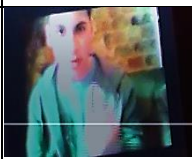
Figure 20: Lily practises pointing with the device before the task




The following negotiation excerpt (Excerpt 6) which I present next has been selected as it is the lengthiest and most complex across the entire data set (nearly four minutes long). It also demonstrates the learners' use of affordances whilst illustrating some of the challenges which a learner can experience in terms of interaction with their device in relationship to achieving an accurate prosthetic deictic gesture within the negotiation of meaning. The sharing of an image of an object across Skype also prompts one learner to exploit his second device and the search engine Google in order to continue to negotiate around facts based on the film which he associates with the object.


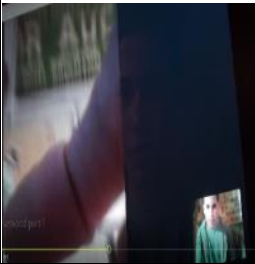




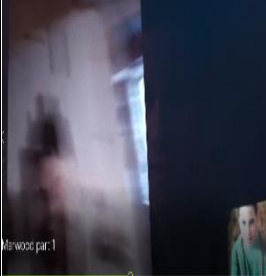
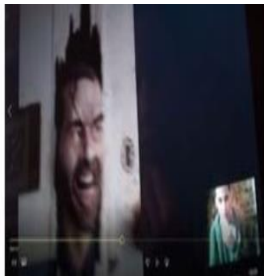
Coding disagreement with a second coder arose over which of the learners' moves could be categorised as indicators of non-understanding in this research context. For example, in this particular excerpt, unlike the original framework (Varonis & Gass, 1985), the indicators were not inevitably based around the speaker querying aspects of language but could instead consist of turns where one learner (Lily) asks their interlocutor (Paul) where to go with their device in terms of achieving the task goal of locating the correct artefact. From an analysis of the speech mode in isolation, it was sometimes difficult to comprehend the wider context of the interaction and could sometimes appear as if one learner was querying items of language, rather than attempting to understand their

interlocutor in order to negotiate the geography of their particular setting. After consideration, my original coding decisions were left in place with the caveat added that further research data is required to confirm more precisely how these coding categories operate from multimodal perspectives. The coding categories share commonality in that they were frequently generated by speech and gestures which, as in Excerpt 3, revolved around how to convey certain features of the original object. Due to the task design, one learner had been required to locate and share this object across Skype from their café setting. However, following a prosthetic deictic gesture, the learners notably abandon their use of gesture and speech and instead choose to continue to negotiate for meaning via language use alone. A multimodal analysis, beyond the mode of language, revealed an interpretation and explanation as to why stretches of talk in isolation had suddenly occurred in the latter portion of the negotiation. Excerpt 6 is next illustrated in multimodal transcription form.

Excerpt 6: The negotiation of meaning via Skype VC on mobiles from two café locations

<p>1:P (00:09:34)</p>	<p>It's a place. There is a [like a head of a man] [In the wall] attached in the wall. Like eh [he's like eh getting through the wall you know] [like a door and there's a door is] [broken] [and his face is like eh getting through the door]</p>	<p>T</p>	 <p>(00:09:37) ICONIC (00:09:39) DEICTIC (00:09:44) ICONIC <i>Sequence of gesture strokes</i></p>  <p>(00:09:47) ICONIC (00:09:49) ICONIC (00:09:50) ICONIC</p>
<p>2: L (00:09:53)</p>	<p>[The face?]</p>	<p>I</p>	 <p>(00:09:53) DEICTIC (Partially off-screen)</p>
<p>3: P (00:09:54)</p>	<p>Yeah [like it's a]</p>	<p>R</p>	 <p>(00:09:54) ICONIC</p>
<p>4: L (00:09:55)</p>	<p>It's a man?</p>	<p>I</p>	
<p>5:P (00:09:57)</p>	<p>Yeah, it's a man] yeah. [It's a place around you I think]</p>	<p>R</p>	 <p>(00:09:57) ICONIC</p>
<p>6: L (00:09:59)</p>	<p>Could be downstairs. So ok. Just ask. It's a face of a man. He goes outside of the wall. It's near to the door?</p>	<p>I</p>	
<p>7:P (00:10:00)</p>	<p>Maybe downstairs</p>	<p>R</p>	
<p>8:L (00:10:01)</p>	<p>Could be downstairs.</p>	<p>I</p>	
<p>9: P (00:10:02)</p>	<p>Yeah.</p>	<p>R</p>	
<p>10:L (00:10:03)</p>	<p>So ok. Just ask. It's a face of a man.</p>	<p>I</p>	
<p>11:P (00:10:06)</p>	<p>Yeah that's...</p>	<p>R</p>	
<p>12:L (00:10:07)</p>	<p>He goes outside of the wall.</p>	<p>I</p>	

13:P (00:10:08)	Yeah	R	
14:L (00:10:09)	It's near to the door?	I	
15.P (00:10:10)	Yeah. It's a scene of a movie actually. I don't know if you know this movie of Stanley Kubrick. I don't I don't think so but it's really easy to find I think. [It's a man] [that is like eh getting through the door you know].[The face]	T	 (00:10:24) (00:10:25) (00:10:26) ICONIC ICONIC ICONIC
16:L (00:10:27)	Ok. So I have to walk and to find the place?	I	 (Lily stands up and walks with device)
17: P (00:10:30)	Yeah. Maybe you're going to find it downstairs.	R	
18: L (00:10:33)	Ok. So you told me it's near to the to one door or not?	I	
19: P (00:10:44)	I think it's eh [in the wall]	R	 ICONIC (00:10:45)
20: L (00:10:45)	Ah ah I think I've. It's ok it's ok. I have it. I got it.	RR	
21: P (00:10:47)	You can see?	VCC/R	
22: L (00:10:48)	I got it. Yes. I have to show you	I	

	maybe this one. I don't know if your picture?		
23:P (00:10:52)	Yes. [Try to show me] 	R	 DEICTIC (PROSTHESIS) (00:10:56)
24: L (00:10:57)	[Ah wait] 	RR	 DEICTIC (PROSTHESIS) (00:10:57)
25: P (00:10:59)	Yeah. [I can see]   [I can see it]. Do you know who is the actor?	R/T	  DEICTIC (PROSTHESIS) (00:11:00) (00:11:01)
26: L (00:11:08)	I don't know the name of the actor. I don't know] You know the name?	I	
27: P (00:11:15)	No.. do you know which film is that?	R	
28: L (00:11:17)	No no. I don't have a good knowledge about films or ...I really don't know, sorry.	RR	
29: P (00:11:25)	I know there is. I don't know the name of the film but I know that it's a film of	R	

	Stanley Kubrick. Director,		
30: L (00:11:32)	OK OK.	RR	
31: P (00:11:35)	But it's really famous. I can google it maybe.	R	
32: L (00:11:39)	OK. But what kind of film?	RR/I	
33: P (00:11:42)	I think it's a horror movie. I think. Not a horror movie but about a psychopath I think.	R	
34: L (00:11:50)	Yes, you can help me.	RR	
35: P (00:11:55)	I will try to google it. Just a second,	R	
36: L (00:11:56)	Ok ok. S*** I forgot to ask for a spoon.	RR	
37: P (00:12:28)	Just a second. I'm finding it. The movie is The Shining.	T	
38: L (00:12:40)	Shining...is the title?	I	
39: P (00:12:43)	The title of the movie.	R	
40: L (00:12:44)	OK	RR	
41: P (00:12:45)	And uh. It's a film 1980.	R	
42: L (00:12:54)	OK. Excuse me. Can I have a spoon?	RR	
43: P (00:13:07)	I will try to find who the actor is.	R	
44: L (00:13:09)	Ok. Sorry, I don't have a good knowledge about movies.	RR	
45: P (00:13:15)	I think I've never seen this movie but eh....	R	
46: L (00:13:18)	I think it's an old movie.	R	

47: P (00:13:20)	More than thirty years ago. But it's really famous. It's Jack Nicolson. Ah yeah Jack Nicolson. The actor. There's another place....	RR	
---------------------	---	----	--

Paul initiates both the task and the negotiation within Excerpt 6 as he is observed to create a multimodal trigger via his deployment of six, mainly iconic gestures, within a single speech turn. This first turn, and the multiple gestures within it, refer directly to the relevant artefact which Lily is required to locate from within her setting of the café. Paul creates the possibility to negotiate for meaning via his co-orchestration of a meaningful series of multimodal ensembles. These appear to be performed for the communicative benefit of an interlocutor through the affordances of the setting and his use of technology (see Figure 21). Paul accurately co-ordinates his gestures with the essential aspects of his linguistic message, centring his hands accurately within the boundaries of his mobile screen where they can be seen. I interpreted that he was recreating a cinematic scene via his hands but also through the dynamic use of his head and body; which are all exploited in conjunction with his language use. Paul's meaning-based gestures are 'larger than life' with the hands functioning as a resource from which to convey a narrative. Unlike Lily, who continually holds her device in one hand, Paul's hands remain consistently free to move position in direct relationship to his shifting use of language. Figure 21, which follows, shows six screenshots of the gestures which occurred across Skype VC in turn one. Gestures 1, 3, 4, 5 and 6 are iconic with one deictic gesture formed with the thumb seen in gesture 2.

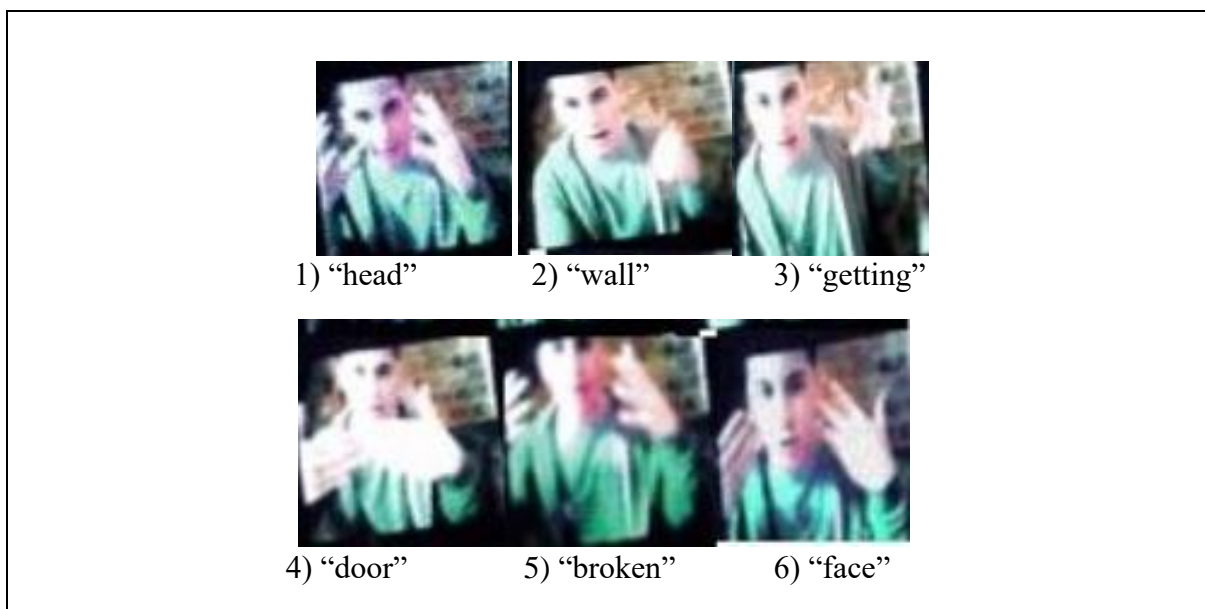


Figure 21: Paul triggers negotiation via a sequence of gestures and speech

Analysed in close combination with Paul's use of language, the previously illustrated gestures appear to forge a meaningful sequence in that these movements are as much a part of the learner's communicative efforts as his deployment of the spoken word. During the process of analysis, it became clear that Paul's gestures could be positioned side-by-side to form a coherent narrative: the meanings could thus be read from left to right and row by row as in a multimodal text. Without understanding their spoken context, an accurate identification of their meaning becomes problematic. Therefore, the gestures, whilst sequentially constructed, could not be understood without first linking them back to the learner's deployment of specific examples of speech. As Norris (2004) states, in viewing the gesture alone it is usually not possible to come to a conclusion because "the analyst needs to refer to one mode to be able to understand the message in the other" (p.29).

Multimodal links between Paul's choice of language and his creation of gestures are evidenced in the re-appearance of the original gestures (turn 1). These recycled iconic gestures are derived from the initial trigger as he tries to support Lily in further clarifying meaning for her (turn 15). During the process of analysis, I realised that the same gesture *strokes* co-occurred in conjunction with his recycling of items of language but that these were sometimes paired with different key words when they reappeared for a second time within the negotiation. For example, the iconic gesture which occurred in conjunction with Paul's initial use of the verb "**getting**" (an open hand coming forward in the first turn) is transformed into a different gestural representation of this same verb (the learner's face emerging through his hands) by turn 15. Paul also points to the wall behind him within his setting with the use of his thumb (turn 1), however, when he tries to re-represent the object once again (turn 1) he next chooses to move his hand forward towards the mobile screen positioned in front of him, as he transforms the screen into a wall.

The meaning-based *strokes* of Paul's gestures are always synchronised with his use of key words and ensure that combinations of gesture and speech reflect incidents of the GP in that his gestures occur within close synchrony with language use. However, the sequence of gestures within the first turn appears to be supporting Paul's use of language, rather than contributing new meanings which are not already present in his L2 speech. From this perspective, Paul's use of gesture as a second language speaker is probably serving the function of 'concretizing' the verbal channel (McCafferty, 2004). In the example of the first trigger (turn 1) there is gestural redundancy (Negueruela & Lantolf, 2008) in that Paul talks about a "**head**" as he gestures towards his head and then conveys the concept of a

closed door. His representation of a door is enabled via both of his hands brought together in full view of his interlocutor's screen. The fourth successive gesture is designed to represent a door which is at first closed but which then shifts to signify to Lily that the relevant door is: "**broken.**" At this point in the trigger, both the learner's hands move apart in a dramatic fashion. Paul's gestures may also be valuable to the trajectory of the negotiation in that, as in examples of his language use, these can be recycled to enable him to attempt to reiterate and clarify sources of information in order to continue to negotiate for meaning with an interlocutor. Incidents of non-understanding appear to support this dyad to work together as a team in order to solve the problems which they encounter.

I observed that Paul did not employ meaning-based gestural *strokes* to occur on his use of incidental words, for example, they do not coincide with his use of a range of discourse fillers such as "like." He has probably acquired this colloquial language during his stay in the UK. I consider that these fillers afforded him time in order to think before he formed the critical, meaning-based *strokes*. His gestures may be supporting his own talk whilst also being exploited as a way to scaffold his interlocutor in her understanding and achievement of the task goal. The fillers allowed Paul to begin the gestural *preparation* phase with his hands in order to go on to convey the principal message via his deployment of the key gestural *stroke*: "that is like eh **getting.**" The gestures also move beyond the use of the hands in that Paul draws on a multiplicity of other affordances. These include his body, the wider setting; and his recognition of the possibilities made available within video-based forms of communication which support him to enact his performance. Paul's gestures and language use are always reflective of the wider context (Streeck et al., 2011) in that they are interdependent on his ability to identify and co-ordinate a range of available affordances which operate in social, material, and virtual ways. The synchronicity and complexity of this multimodal orchestration is illustrated next in Figure 22.

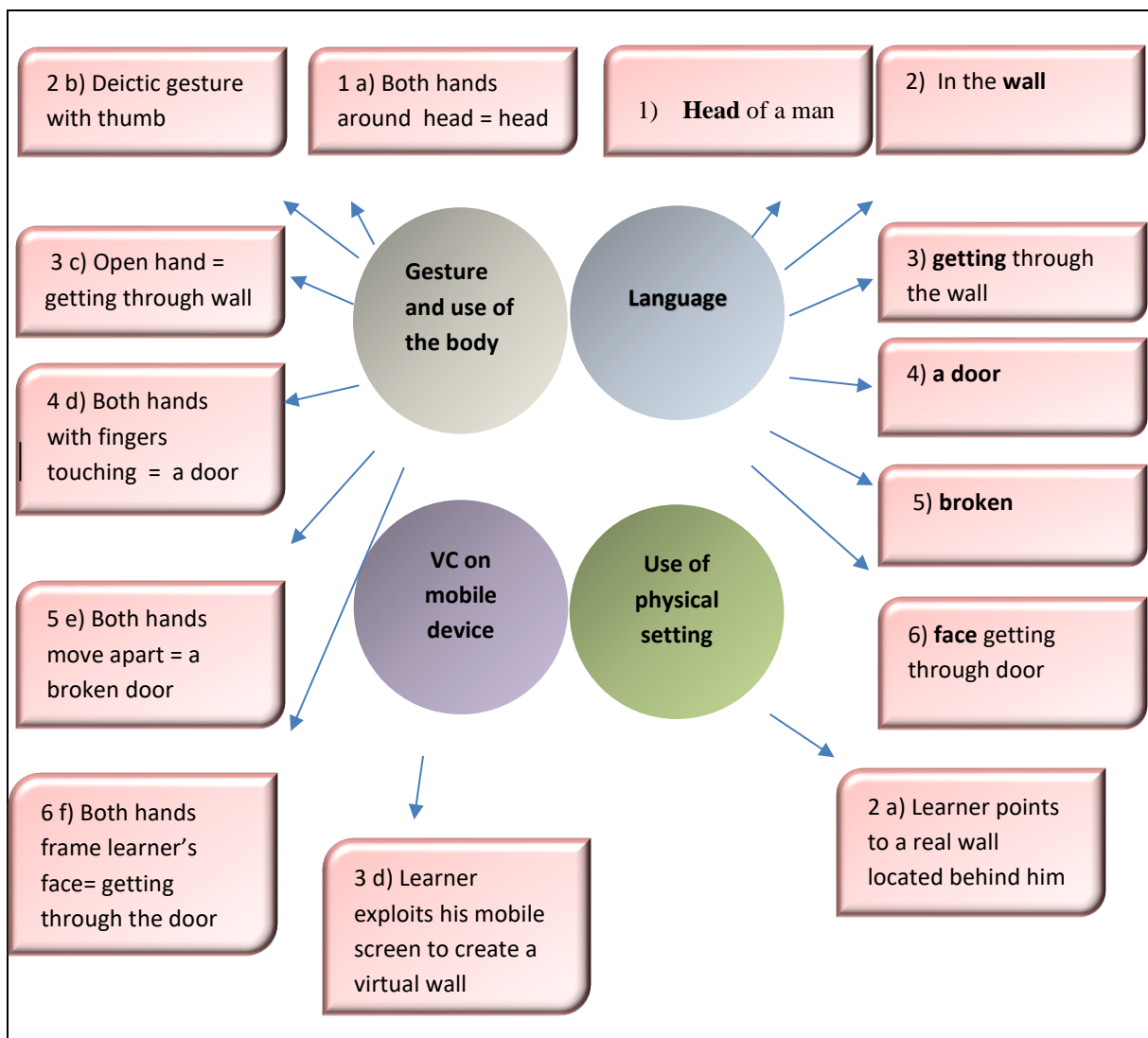


Figure 22: Paul's simultaneous orchestration of gesture and speech

There is further complexity within this interaction in that the Paul's use of gesture and speech is reliant on his recognition of the affordances made available to him via his own physical setting and through the virtual world which he creates via his use of the technology. For example, Paul's deictic gesture is achieved with his thumb as he executes this movement to co-occur with his use of the key word: "**wall**" (see Figure 22, 2 and 2b). This deictic gesture cannot be detached from the learner's understanding that there is a physical wall located behind him which he recognises can be seen by his interlocutor on Skype. However, a gesture towards the wall behind Paul is designed to instead represent the presence of a wall in Lily's own context in ways which appear to merge both settings into a single communicative space. When Paul next explains that the man is "**getting**" through the wall" he decides to structure his deictic gesture in relationship to the affordances of the technology as he moves his open hand towards his mobile screen (see

Figure 22, 3c). He has constructed a ‘virtual wall’ as he gestures towards his own screen with his hand in the hope that this will be understood by an interlocutor as the conveyance of a wall which the Jack Nicolson character is “**getting** through.”

This type of complex multimodal orchestration by one learner may potentially support understanding but I also considered that the diversity and amount of multimodal information conveyed could equally confuse aspects of communication from an interlocutor’s perspective. In Excerpt 6, the requirement to constantly clarify sources of information in order to complete this section of the task is evidenced in the period of time which the negotiation continues; and the considerable amount of embedded indicator and response moves observed throughout this excerpt. During this excerpt, Paul uses the mode of gesture as an opportunity to initially convey meaning but also to reiterate meaning. However, it is clear that there is a lack of understanding for Lily in that elements of Paul’s message are potentially confusing. For example, it is difficult to discern from his combined use of gesture and language whether the designated object is emerging from a model of a wall or a door; or whether the object is embedded within a real wall or door which is located within Lily’s café setting. Lily confuses Paul’s description of the object: “getting through the door” (turn 15) with the requirement for her to locate a physical door within her own setting which she feels may enable her to locate the object: “It’s near to the to one door or not?” (turn 18). As a result, Lily expresses non-understanding as she continues to work hard in order to clarify the accuracy of the information conveyed within Paul’s trigger and response moves. Unlike the traditional notions of the Varonis and Gass framework, she is not exploiting coded indicators as a way to question aspects of his language use but to instead understand which object she is looking for and where she is required to go with her mobile device.

Due to task design, the issue of non-understanding experienced between this dyad is based on their multimodal attempts to clarify and explain sources of information contained within a description of a particular object (see Figure 23, shown next).

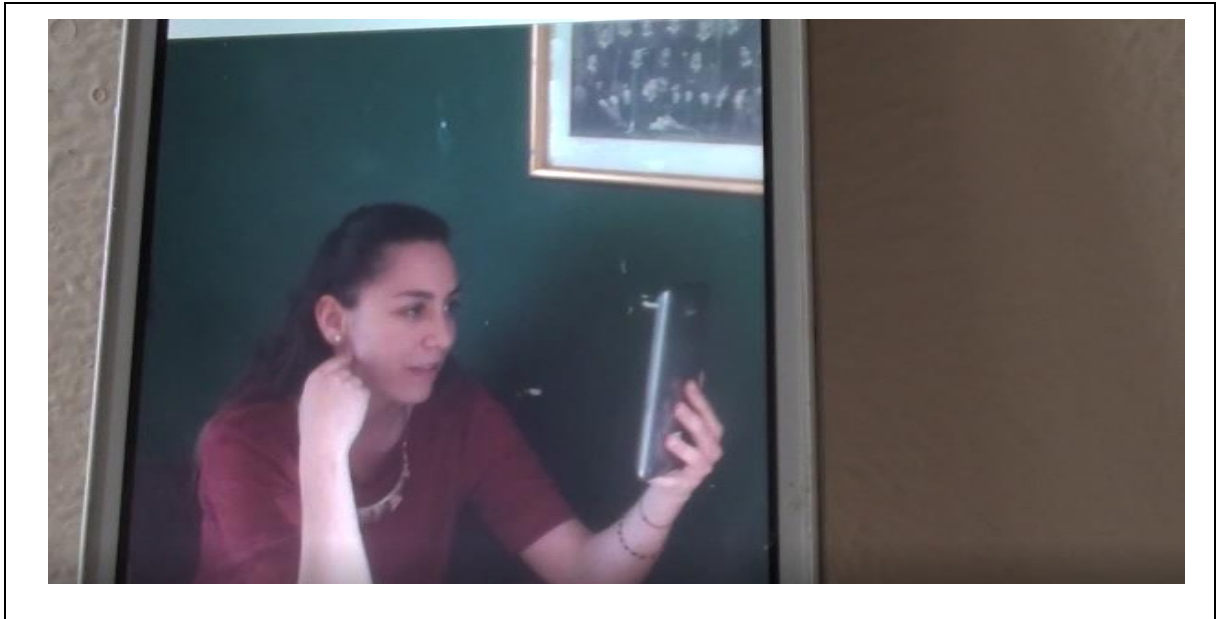


Figure 23: Lily exploits a deictic gesture as an indicator: “the face?”

I observed that Lily harnesses and modifies gestures in her own way in that she does not inevitably copy gestural movements observed within her interlocutor’s previous turn. The indicator move is enabled through an echo strategy, according to the original Varonis and Gass framework, however, in this example gesture and speech are brought together in order to achieve this move in multimodal ways (see turn 2). Conversely, Lily’s deictic gesture towards her face can barely be seen as it appears as a fleeting movement which mainly occurs beyond the boundaries of her mobile device. At this stage in the task, Lily expresses little recognition of the requirement to adapt her gestures in order that these can be seen by an interlocutor, within the spatially constrained boundaries of the mobile screen, as she continues to hold her tablet device in her hand in a fixed manner.

There are no deliberate prosthetic deictic gestures observed with the device in the early stages of the negotiation. Conversely, Lily walks around her space whilst she inadvertently conveys random images of parts of her body (see turn 16, as an example). These random images are created by one learner’s seemingly unconscious use of the visual affordances of the technology in ways which differ from conscious prosthetic deictic gestures used to point at objects. From turn sixteen, Lily appears to fail to engage with the video in any meaningful sense as she walks around her space to search for the object whilst listening to Paul. Therefore, I considered that there was no deliberate communicative intent conveyed via her visual use of the device within turn sixteen (see Figure 24).

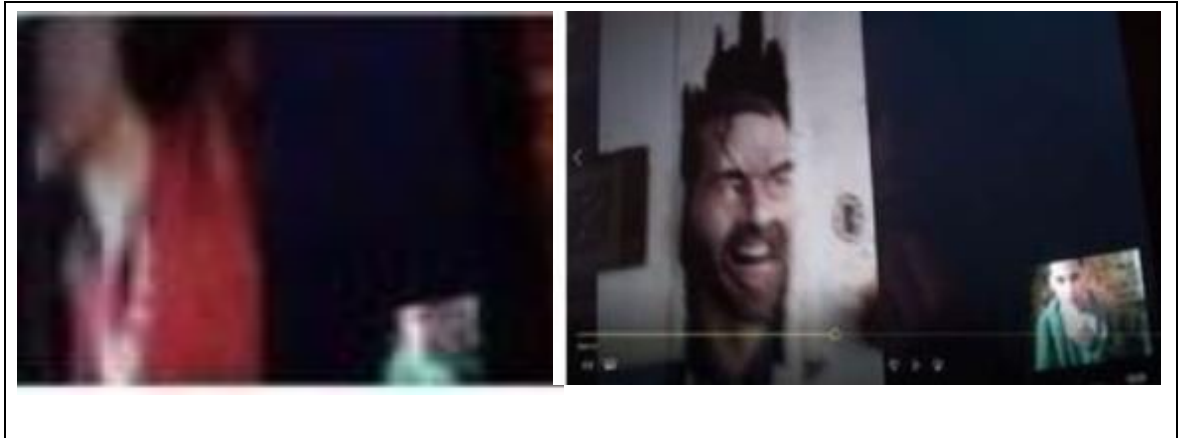


Figure 24: A random image on Skype versus a deliberate prosthetic deictic gesture

In turn twenty-three, Lily locates the correct artefact from within her setting. She is next faced with the challenge of having to point the device and camera accurately enough in order to share the model of Jack Nicolson via her conveyance of a clear image of this object. Whilst she extends the device and moves towards the object in preparation, she fails to point in an accurate fashion until, after three attempts, she eventually creates a successful prosthetic deictic gesture. As a result, her pointing does not automatically lead to the correct object being indicated to an interlocutor. The pointing gesture does not produce a coherent space within which the gestures may be jointly understood (Haviland, 2000). Whilst Lily is engaged in pointing, the language use involves collaborative dialogue focussed on whether this is the correct object and how next to ensure that the object appears on Skype for Paul (turns 23-25). Paul can indicate to Lily critical information about the accuracy of her prosthetic deictic gesturing in real time: “I can see it.” Lily's final prosthetic deictic gesture conveys a clear image of Jack Nicolson's face across Skype VC (see Figure 25, shown next which includes various attempts to form the *stroke*).

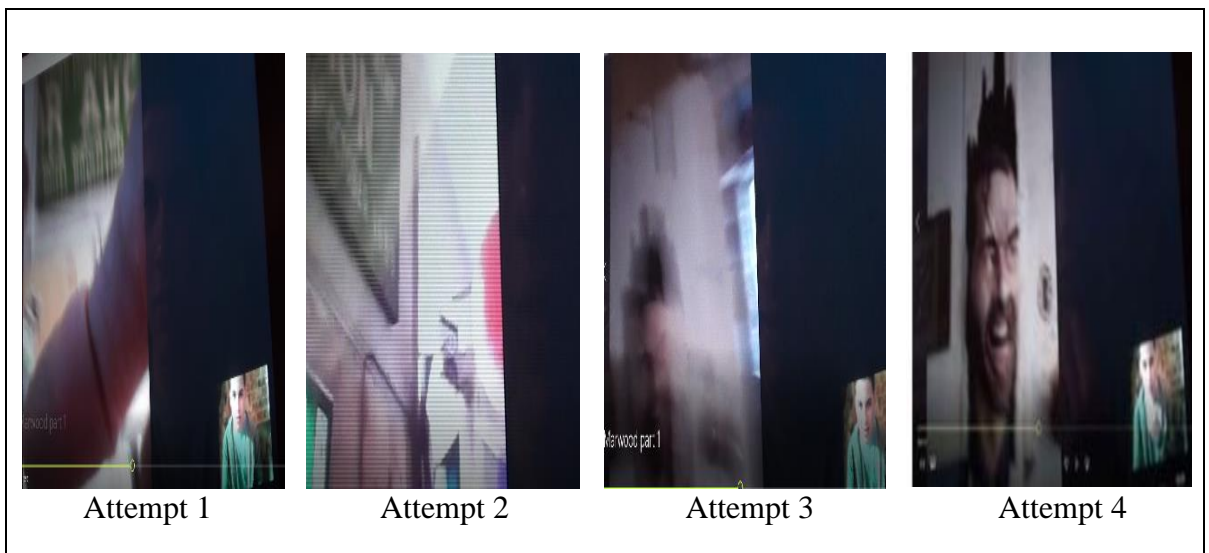


Figure 25: A series of attempts to form an accurate prosthetic deictic gesture stroke

As seen in the images in Figure 25, Lily re-engages with the Skype video stream through her attempts to deictically gesture with the device and camera after she has located the object which now needs to be visually conveyed to Paul. Lily first stands in front of this and extends one hand with the device. Conversely, at first, she targets the wrong object, for example, a poster and the learner's extended arm appear on Paul's screen (see Figure 25, attempt 1). Lily next captures the poster from a different angle (Figure 25, attempt 2) and continues to shift the device and camera in the hope that she will be able to locate the correct object and achieve a more accurate gestural *stroke* (Figure 25, attempt 3). In the analysis, I concluded that these three earlier attempts represented a series of false moves as they failed to convey an accurate visual representation of the object to Paul. The analysis showed that the only accurate prosthetic deictic gestural *stroke* (Figure 25, attempt 4) was achieved due to peer feedback regarding the accuracy of the visuals which Paul had simultaneously received on his mobile tablet screen. Paul consistently scaffolds Lily in order to enable the dyad to jointly locate and share the designated object in a collaborative manner. For example, he first checks if she can see the object from within her own location. This was coded to represent another example of a VCC (Visual Comprehension Check) (see turn 21). Paul next responds with encouragement without placing Lily under any undue pressure: "Yes. Try to show me" (turn 23). Lily engages with a pointing action with the device and camera and Paul confirms when he receives an accurate image of the face of Jack Nicholson conveyed to his mobile screen (turn 25). It is the learners' collaborative efforts which have succeeded in the establishment of a shared interactive space which is capable of bridging the gap between their physical and virtual worlds. However, Lily next chooses to retract her prosthetic deictic gesture with the device

and subsequently removes the image of the Jack Nicolson model from Paul's screen in turn twenty-six. This is the same pattern of behaviour around prosthetic deictic gesturing as demonstrated in Excerpt 3 by Andrea. The reason why Lily chose to retract the prosthetic deictic gesture from the object under negotiation was not clear to me from my analysis of the video footage.

Instead of attempting to re-represent the object via his physical hands, Paul continues to negotiate for meaning through the exploitation of his second device, a mobile phone, which he uses to search for key facts about the original film. This action leads to a brief embedded coded routine, negotiated via language use alone, with an opportunity for Lily to possibly query and correct the grammatical dissonance of the film title in turn thirty-eight: "Shining is the title?" It is left unclear as to whether this represents a self-generated repair move around inaccurate aspects of grammatical form or instead demonstrates a need to establish meaning in terms of confirmation as to whether or not this is definitely the title of the movie. The negotiation appears to have been resolved in that the object has been located and the details of the film confirmed with the dyad free to move on to negotiate over other objects. However, these learners extended the task as Paul used the image on his mobile device as a springboard to go on to find out additional information about the film. Lily is clearly unfamiliar with the film and the learners then begin to share information in ways which support their further negotiation of meaning through the use of language, operating in isolation. Gestures are notably absent within the learners' talk from the point where the object has been shared on Skype (from turn 26 onwards). The dyad cease to gesture in the latter stages of the negotiation. I observed that this is because Paul shifts his gaze to another mobile device whilst Lily continues to walk around her setting, creating random images, without looking at her mobile screen or pointing at anything specific. Paul is now engaged in an ongoing internet search but continues to talk to Lily to ensure that she remains co-present. He finds the information out about the film and then shares this with Lily. The final exchange consists of a series of response and reaction to response moves until Paul, having established who the actor is, suggests to Lily that they locate another place within her café.

To summarise, there are considerable amounts of multimodal negotiation of meaning which occur within this particular excerpt. Both learners appear to be motivated and therefore collaborate effectively with one another in order to sustain their negotiation around the artefact until this is located. The negotiation also continues for a period after the

object had been identified and shared across Skype VC through Paul's exploitation of further affordances via his second device. Paul first scaffolds Lily via gesture and speech in order to enable her to locate the correct object and Lily then perseveres with her prosthetic deictic gestures, despite encountering difficulties. The sheer complexity of the task and wider learning environment appears to drive the trajectory of the negotiation. Multiple factors come together to push this dyad to clarify, respond, and modify some potentially confusing elements of their interaction. The negotiation of meaning contains examples of gesture and speech operating together but also large amounts of language use in isolation. The Varonis and Gass (1985) coding categories proved valuable within the analysis in order to elucidate how triggers, indicators, and response moves are constructed by two learners through their deployment of gesture and speech but also speech alone. This excerpt of negotiation further illustrates that the act of pointing is as much about the actors involved as the object which is to be pointed at: the requirement to "pin-point" a specific place within a location (Jaworski & Thurlow, 2011, p. 257). In this instance, the communicative issues with the device could be resolved through effective and supportive forms of scaffolding and peer collaboration. Forms of dialogue ensured that, whilst learners were operating from separate geographical spaces, they were never left isolated or unsupported when they encountered challenges.

In the following two interview excerpts (Excerpt 7 and Excerpt 8), I interview Paul and Lily separately about the negotiation excerpt (Excerpt 6). I chose these excerpts from the interviews because they illustrate how two different learners appear to show some overlap in their verbalization on the perceived role of gestures, however, they also associate two different modes (visuals versus language) with the use of gesture as they reflect on Excerpt 6. It becomes clear from the start of the interview that Paul initially appears to have no memory of his gestures from the time of the task. I feel that through the stimulated recall video clip, replayed in order to support his recall, he is able to bring the gestures into his more immediate consciousness as he begins to reflect on his thoughts at the time of the task. My first question in the stimulated recall that relates to the negotiation in Excerpt 6 is based around Paul's sequence of gestures which had occurred in his first turn which I had coded as a trigger. This was a source of interest, given the amount of gesturing which had occurred at the start of the negotiation of meaning. In my original analysis, based on a transcription of the video from the task, I had considered that Paul's deployment of gesture was entirely conscious. His interview data challenges my assumption in that it indicates

that speech-associated gestures may pass learners by without a more structured, reflective opportunity in order to review them. Lily appears to be able to more easily recall both Paul's and her own gestures from the time of the task itself; and I ask her about these as well as asking her to describe her behaviour with the mobile device as she tried to convey the correct object. Excerpt 7, shown next, presents a section of Paul's stimulated recall interview based around the negotiation in Excerpt 6.

Excerpt 7:

So we're going to first talk about your gestures. Can you remember what you were gesturing with your hands here?

No.

Can I ask what you were gesturing here?

I don't think that this is conscious. I'm talking about the head and making a gesture around my head. Ah yeah but the truth this is maybe conscious because I want her to understand that it is getting through the door. It's an action a movement so I tried to be clear. It was for her. I'm recreating the scene. The door and getting through it. Creating the image in her head. I'm trying to make the image in her head so she got the image. To inject the image into her head.

Did you notice her gesture at the time?

No. Ah yeah she's pointing to her face. To make sure that she understood the task I think. It shows agreement which is good but I couldn't see this at the time.

Can you remember what Lily was doing with the mobile device here?

No. Ah yeah she took the camera away because she thinks that I've got the image now so she goes upstairs because of the noise I think. As she saw that I saw the image she pointed at it and then she saw and I saw so I think there was no point in keeping the image on the screen.

I consider, from analysis of this section of the stimulated recall interview, that Paul did not initially remember his use of spontaneous gestures from the time of the task. However, via the continued stimulation of the event and questions, the original interaction is reactivated for him. During the process of our conversation, this learner appears to shift his perspective from first considering that his use of gesture was unconscious but then seemingly changes his mind as he begins to explain his communicative intent, for example, he suggests that his gestures are supporting his use of the verbal channel: "I'm talking about the head and making a gesture around my head." He next connects his use of a gesture to his action-

based verbal description which had been designed to enable Lily to “understand that it is getting through the door.”

The revisited gestures on video were seemingly designed for Lily’s benefit; although Paul interprets the use of gesture as the ability to literally transfer a series of visual images across technology from one learner’s mind to another. His interview suggests that gesture can be related to the imagistic aspects of thought (McNeill, 1992) but can also be used as a way to transfer individualised cognition to another person. It does not appear to be important to Paul that Lily’s mind is located in another geographical location as it is seemingly possible to “make the image in her head.” It appears that the visual nature of gestures can enable shared forms of cognition and Paul uses graphic language to inform me as to how exactly this process works: “To inject the image into her head.” This particular learner appears to have a sophisticated understanding of the meaning-based potential of images relayed across technology. The images which he creates are made available through resources such as gesture but also through recognition of the affordances of video-based forms of communication and the world around him. Paul next informs me that once the designated object has been visually relayed, and seen by the interlocutor, the *retraction* of the gesture (and the resultant image) is deemed to become superfluous to the negotiation process. Therefore, Paul appears to consider that it is sufficient to simply glimpse an image across Skype in order to acquire its meaning: “I’ve got the image now.” Lily’s prosthetic deictic gesture and the image she conveyed of the object enabled Paul to continue to explore facts about the film via his use of Google on a second device.

Aspects of this learner’s ability to autonomously pursue and then share knowledge is evidenced in his identification as to the power and relevance of images within multimodal forms of communication across digital technology but also his deployment of a search engine as another available learning resource. He also appears to have discerned some information about Lily’s setting at the time of the task as he suggests that background noise caused her to retract her device and to move upstairs. This information about the technology use in relationship to the setting had not been made apparent to me from studying the captured video of the negotiation in isolation.

Next, I present a section from Lily’s interview as she discusses her experiences and interpretation of the same video clip as Paul (Excerpt 6) (see Excerpt 8, which follows).

Excerpt 8:

So here we're first going to talk about your partner's gestures. Can you remember what Paul was doing with his hands?

Yes. I see his gestures on my tablet in the café but it's also quite difficult you know because I have to see listen see listen all of the time... Like it is a face of a man and he is getting through the door (learner repeats gestures from video) but later I check to understand if I'm correct or not you know if it's a face and it's near to the wall or door or getting through the door.

Could you tell me what you were doing with your device?

Yes, I remember this. Downstairs was very difficult to understand because the music was very loud. I had to think about the light and the position of my hands. If it is high or low... maybe it's not in front of it [the object] but I was on the stairs. At this moment I was in the middle. It was the same the music again because all the time I tried to control the screen with his mouth moving and speaking and so I had to move.

Lily seemingly expresses a more conscious memory of the gestures which occurred throughout this negotiation (see Excerpt 6); and within this interview she first elucidates some issues which she experienced around multimodal overload. She also depicts some of the issues which can arise for learners when they are using their device to complete tasks from beyond the classroom. Unlike Paul, Lily does not talk overtly about the power of images in relationship to gesture but she does reflect on her confusion as to where the object was situated within her location in relationship to her understanding of the information which her interlocutor had conveyed.

Lily next explains the difficulties which she encountered around pointing with the device in relationship to her interaction with the wider setting. For example, from studying and transcribing the video, I could not discern any background music playing so I had little idea as to why Lily had suddenly retracted her deictic gesture and had chosen to move to a different section of the café. In this case, Lily confirms the information which Paul had provided but was able to explain something more about the physical and multimodal challenges which she had personally experienced. Lily's interview highlights that there are complex levels of multitasking required by a learner in relationship to their use of Skype VC to achieve goals in settings beyond the classroom; and her reflections are reminiscent of a professional photographer attempting to take a good picture. Lily explains that she is









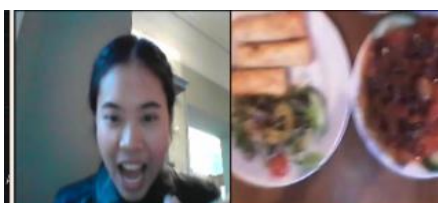
required to understand how and where to position her hands as she reflects on her attempts to hold the device ‘high’ and ‘low;’ and to take aspects such as light source into consideration in relationship to an accurate conveyance of the object as she balances on some stairs. Prosthetic deictic gesturing with the device and camera was clearly not an easy task for her. Lily reminds me that there is background music and distraction which can occur in an uncontrolled manner when learners are asked to communicate from an informal setting such as a lively café. As a result, multimodal elements of the dyad’s communication were operating in synchrony with unpredictable aspects of café life. These settings offered possibilities but also constraints which could not be easily controlled by learners or a researcher. However, Lily solved her issues in an autonomous, practical manner in that she ‘became mobile’ and took her device back to a quieter part of the café where she had been previously seated in order to continue the negotiation.

4.5 Dyad four: Angela and Bobbi

Next, I present a final negotiation excerpt (Excerpt 9). This excerpt is derived from the data of the speaking task which occurred between Angela and Bobbi from dyad four. Angela is twenty-seven and is a marketing student in the city of Istanbul. Bobbi is aged twenty-three, from Thailand, and is a fine art student back in her home country. Angela had been studying at the school for one month and Bobbi had quite recently arrived in the UK and had only attended the school for around two weeks. Angela exploits my tablet device and Bobbi decides to use a 2-in-1 device as she says that this is larger than her mobile phone. For the purposes of the implementation of the speaking task, Angela is situated in a Turkish restaurant and Bobbi in the local museum and gallery (one location). These choices of settings for the task are not random but the result of Angela specifically asking me if she could complete the task from a space which was more reflective of her own Turkish culture. Angela is unusual as a learner participating in this study in that she approached me directly about a specific work-based need. She is responsible for taking international clients out to lunch and has to present items of Turkish food in English within this role within her job in Turkey. This is the specific skill which she wishes to practise in English. The learner is thus directing the design and goals of the task herself in ways which demonstrate important aspects of language learning such as self-identified needs, motivation, and autonomy. I tried to accommodate this learner’s needs within my task design by instructing her to go ahead and order some Turkish dishes in order to exploit these as a way to negotiate for meaning with her interlocutor.

Excerpt 9 has been selected as this dyad were observed to employ prosthetic deictic gesturing as a way to repair the identified communicative problem which had occurred between them. Contrary to my previous excerpts of negotiation, the learner uses the device in an atypical way as she is required to point at objects directly positioned on a table below her. Angela demonstrates how challenging this action can be for a learner whilst they simultaneously try to convey verbal information to an interlocutor. Furthermore, the concept of speech and gesture inevitably operating as a holistic unit of meaning is disrupted as the gesture-speech interplay is initially incoherent and confused. Within my discussion of Angela's use of the device, I also include some visual data from dyad two (Nadia and Simone) and dyad three (Lily and Paul) to demonstrate how prosthetic deictic gestures are achieved by learners from settings beyond the classroom in a variety of different ways.

Excerpt 9: Negotiation of meaning via Skype VC on mobiles from restaurant and gallery

Turn	Gesture and speech	Coding	Visual transcription
1.A (00:03:05)	 [We put some vegetables sometimes. Sometimes cheese sometimes meat and near normal salad. And this name is a little difficult but Shakshuka they call it.	T	 DEICTIC (PROSTHESIS) (00:03:05)
2.B (00:03:19)	Shak-shu-ka ?	I	
3.A (00:03:20)	Yeah]. [This is made of  Aubergine and Chilli and Tomato we made. Firstly all vegetables we put in fried and after we cut it and with tomato with salad. Sometimes we put near Yoghurt]	R	 DEICTIC (PROSTHESIS) (00:03:23) DEICTIC (FINGER) (00:03:23)
4.B (00:03:46)	 Yeah [can can you move your camera? I can't] ..	I	 DEICTIC PROSTHESIS (00:03:48) DEICTIC WITH HAND (00:03:49)
5.A (00:03:50)	 [Now?	VCC/R	 DEICTIC (PROSTHESIS) (00:03:50)
6.B (00:03:51)	Ah Yes. Good]	RR	

In this negotiation excerpt (Excerpt 9), Angela describes a well-known Turkish dish to her interlocutor, Bobbi, who is situated close by in a combined museum and gallery space. Whilst Angela uses considerable amounts of language to convey her meaning, it becomes evident that her attempts at forming a set of accurate prosthetic deictic gestures act in combination with her language use to trigger a communicative issue between this dyad. The issue of non-understanding is caused by Angela's problematic attempts to gesture with the device which, when analysed in combination with her use of the speech mode, helped me to elucidate aspects of asynchronicity and incoherence operating between the learner's co-ordination of these two modes from a Turkish restaurant. Angela's series of gesture and speech moves result in considerable confusion for Bobbi. Conversely, they generate rich technology-motivated negotiation opportunities, based around the necessity to cohere the meaning of gestures in relationship to the meaning of language use, in order to make oneself understood. Angela's gestures, formed with the device and camera, result in a series of random images which simultaneously appear on Bobbi's Skype screen and which lead her to query the information. This is because the images bear no relationship to the meanings which Angela wishes to convey via her use of co-occurring language. It was difficult to discern from the internally-captured video exactly how Angela was pointing with the device and camera. She makes subtle adjustments to the manner in which she is pointing in ways which made a clear identification of the gestural phrase and GP challenging. It was observed that the negotiation of meaning between this dyad was triggered by the requirement to ensure that combinations of visuals and speech cohered in meaning in ways which could not be explained by examining the use of language in isolation.

Before examining the gestural mode, I had initially considered that the issue of non-understanding between these two learners had been caused by Bobbi's query of a specific Turkish word which she echoes: "Shak-shu-ka?" (turn 2). Via the process of multimodal analysis, it became clear that Bobbi was querying the visual absence of this particular dish because it had failed to appear on her mobile screen when Angela had talked about it. Therefore, whilst the excerpt is seemingly prompted by an initial indicator, based on the clarification of a Turkish word, the wider communicative problem between the learners is actually prompted by an ongoing set of incoherent images. As a consequence of a lack of co-expressivity between modes, there is an ongoing disconnect in the communication as Angela's use of language depicts various items of food which remain consistently absent

from Skype VC. Conversely, Angela's detailed verbal description of the Turkish dish is accompanied by a series of meaningless images of coffee cups and an array of irrelevant condiments as she points with her device and camera.

In her first attempt to form a prosthetic deictic gesture with the device (turn 1), Angela does not extend her hands but instead clutches the tablet in one hand and moves it closer towards her body. As she points with the device, the camera shakes a little but the principal problem appears to be that Angela considers that she is pointing accurately at the relevant dish on the table below her. Angela does not initially engage in the necessary visual comprehension checks which would have enabled her to clarify that Bobbi could definitely see the relevant dish via her mobile screen located in the museum. Angela instead tries out a variety of strategies to emphasise what she wants to draw joint attention to, for example, she co-ordinates her next prosthetic deictic gesture with secondary deictic gestures formed with her physical finger (turn 3). The result of the overall inaccuracy of the deictic gestures is a lack of joint attention forged between the learners and an inability for them to form a coherent interactional space in which to enable mutual understanding. For instance, the deployment of indexicals becomes disrupted in that Angela's use of words such as "this" make no sense to Bobbi because she prosthetically points at the wrong objects on the table.

As a language learner, it appears to be important for Bobbi that her interlocutor's deictic gesturing with the device leads to the correct set of images appearing via Skype VC. In turn four, Bobbi intervenes in an overt manner through speaking and pointing with a left hand (barely visible at the bottom of her mobile screen) as she draws attention to the ongoing communicative issue. Bobbi understands exactly how and where to direct Angela's gesture with the device and camera to ensure that the correct images are enabled to appear on Skype VC. To ensure that she is pointing correctly this time, Angela checks with Bobbi via a Visual Comprehension Check (VCC) in that she confirms that Bobbi's non-understanding has been rectified through checking the accuracy of her own pointing: "Now?" The use of this key word co-occurs with an accurate prosthetic deictic gesture *stroke* as the relevant Turkish dish finally appears on Bobbi's screen. Gesture and speech remained incoherent in meaning throughout this excerpt because Angela's description of the dish called Shakshuka failed to appear on Bobbi's screen until she had finished describing how it is made. From this perspective, it was difficult for Bobbi to comprehend Angela's meaning in any holistic sense until the negotiation had been resolved.

To summarise, this excerpt is triggered by Angela's prosthetic deictic gesturing as she simultaneously describes ingredients and how to make a specific Turkish dish. Bobbi then queries the information from Angela's trigger through an indicator move. If the Varonis and Gass (1985) framework had been exploited to exclusively focus on the learners' use of language, the multimodal nature of Bobbi's query around the dish could not have been fully explained. In this example of negotiation, Angela is seemingly learning a new digital skill and is able to reconfigure her gesturing more accurately through the support of peer feedback. Angela eventually points down at the correct dish positioned on the table below. This accurate gesture is rewarded with a smile and praise from Bobbi as she implies that an accurate pointing gesture formed with the device and camera represents a satisfactory resolution to the communicative problem: "Good." In the transcript, this resolution was coded as a reaction to resolution (RR) with the conclusion drawn that multimodal understanding had been achieved between the dyad only when holistic meaning was successfully conveyed across Skype VC through the holistic interplay between gesture and speech.

In the case of the excerpt shown previously, I was able to briefly observe and film Angela with an external video camera from the location itself. Within this section of data, I therefore had access to two visual perspectives (or angles) on which to draw. This was useful as it enabled me to observe external aspects of the prosthetic deictic gestures formed with the device within the negotiation as it unfolded. This is shown next in Figure 26.

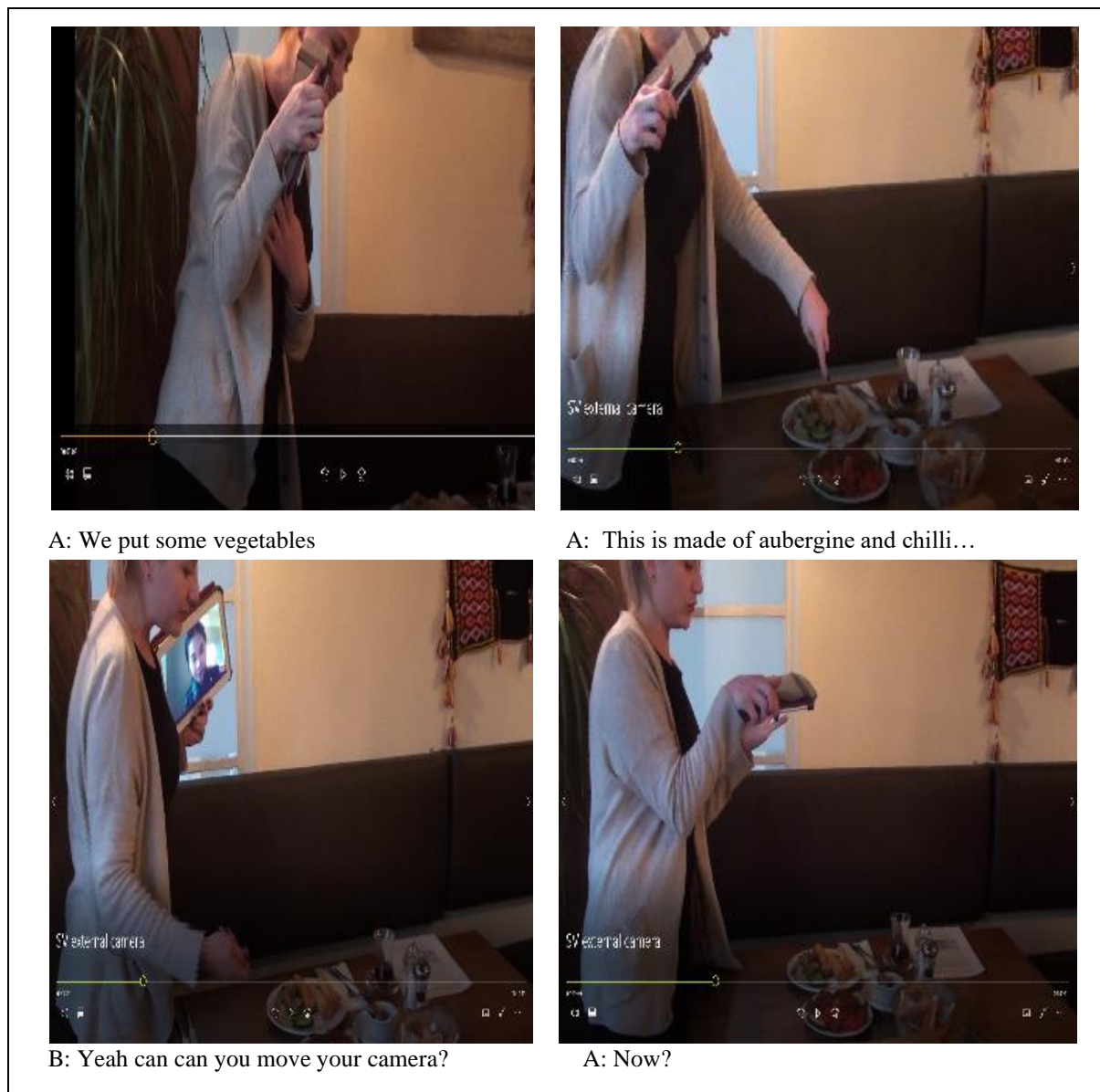


Figure 26: Angela attempts a series of prosthetic deictic gestures with a tablet

As shown in Figure 26, I was able to study Angela's ongoing relationship the technology but also with her material setting. This acted as a means to try to understand more comprehensively why she had experienced difficulties in relaying an accurate set of images across Skype VC. The learner is pushed to co-ordinate her multimodal resources to establish joint attention with an interlocutor in order to orientate them within a meaningful and coherent space. Pointing with a device and camera does not appear to be easy for learners and Angela makes several attempts to realign her gestural *strokes* via assuming a variety of positions with her body and hands. Angela forms her gestures with the hands and arms as she moves parts of herself in relationship to her manipulation of the device and camera; and in relationship to the constraints of the material setting. Just as in Excerpt 6,

the learner is dependent on the active scaffolding of a peer in order to make the necessary gestural adjustments with the device in order to resolve incidents of non-understanding which are caused by a lack of coherence between gesture and speech.

The size and weight of the tablet device, and the manner in which Angela positions herself between a wall, plant, and a table in the restaurant, may have all contributed to making it difficult for her to point accurately. Her use of an additional deictic gesture (with her finger) as she says: “This is made of” would have solved the communicative issue between a dyad within a face-to-face learning context. Conversely, Angela’s use of a deictic gesture with the human finger can barely be seen across Skype VC; and her dual deictic gestural strategies (use of the finger combined with the prosthetic deictic gesture) do not succeed in directing Bobbi’s attention to the relevant Turkish dish.

Learners appear to engage with and interpret the act of pointing in a variety of different ways which include dual use of human fingers and prosthetic deictic gesturing; and the idea that the device itself can be used to provide a gestural substitute for the requirement to point with the human finger or hand (see Figure 27, shown next).



a) Angela uses her finger to point (blurred finger in the left corner of screen) as she simultaneously points with the device



b) Simone uses a finger to point as she points with the device

(Image redacted)

c) Lily dispenses with the human finger and uses the device in isolation

Figure 27: Learners exploit deictic gestures in different ways

I observed that learners adopt a range of strategies in order to draw attention to specific objects within their settings in order to share them across Skype. For example, Nadia and Simone (from Dyad 2, Figure 27 b) developed an effective method with which to share artefacts online. The end result is the learner's ability to more-accurately highlight co-present objects whilst they talk about them with the interlocutor left in no doubt as to where their attention is to be directed. As Simone and Nadia had both been unable to attend the interviews, I was not able to ask them more about this phenomenon but have included a visual example of data on their use of dual deictics (using a finger and the device) (see Figure 27 b). Lily (Figure 27c) instead chose to dispense with the human finger and relied instead on the device acting as replacement for this. This is a successful strategy if the image appears clearly on an interlocutor's screen; and provided there are no images conveyed of other incidental objects located in the vicinity of the specific object under negotiation of meaning.

I next present the interview data (Excerpt 10) from Angela based on the previously shown negotiation (Excerpt 9). I am particularly interested in her opinion on what she thought about what she was doing with the mobile device as she pointed at the food and talked. She recognises the affordances inherent in smaller forms of technology which can be manoeuvred. However, for this learner, the technology is of little interest for its own sake as she begins to connect the possibilities inherent in the technology with an opportunity to virtually enter a new space in order to learn about this. Angela also explains to me how she is excited to have the opportunity to share aspects of her own culture with an interlocutor out in the world. Her exploitation of the technology and setting allows her to engage with language in a manner which she clearly enjoys. Angela also explains some of the difficulties which she experienced in pointing with the technology in relationship to sharing the world around her during the previously shown negotiation (Excerpt 9).

Excerpt 10:

Can you tell me about communicating on Skype when you have a tablet. What was that like?

Eh great because...a different place somebody could explain me some place and if I used some bigger communication equipment it's difficult. But tablets um phone or tablets it's amazing because we saw something different ... I learnt about a place. I've never been there and I learn. This is good for me.

Can you remember what you were doing with the mobile device here?

*I remember that I'm exciting, I'm exciting... I'm very excited. Actually, I would love to share some of my culture, food or places with somebody. I miss some of my old culture. I feel she really wants to come here because I pointed and she asked lots of questions. I talked English and I speak spoke English and I used my culture with English...actually it's a really good method because eh if someone's talking from a different place and eh and told me something and after I remember because I'm living it at the same time. **You're living the place?** I live in the place ... I saw and my I 'm thinking this is this and this is this. When we study English, I easily forget...outside if somebody tells me I live it at the same time and I learn quickly.*

Do you remember having any problems with the mobile device here?

Yes. I want to explain what is the food and how they make it. I remember I took the tablet but now I see that it didn't work but I'm thinking I'm pointing at the correct food... it's difficult because it's difficult to talk and watch on the screen and the food is all over the table. I moved the camera a lot a lot and I did not understand. So she can tell me to move the camera and then I understand why but I say said the dish before so maybe it's too late...

In response to my first question, Angela highlights the nature of learning within a 'place' with her identification of the opportunities enabled through the use of a portable tablet device. She refers to the mobile tablet as "communication equipment" in ways which illustrate that she perceives the mobile device as a means to communicate. She also comments on it being small enough to use for the task. Angela appears to be less interested in talking about her use of the technology but wishes instead to convey the ways in which the overall experience was meaningful and to explain to me why it personally appealed to her. This learner's interview reminds me that, in the design and implementation of language learning tasks (including those implemented via the deployment of technology), certain participants may become tired of the manner in which generic course materials and programmes fail to recognise their individual identity and learning preferences. Angela

misses her own culture and wants to reconnect with it in a manner where she is using the L2 target language to convey artefacts which are reminiscent of her life back home. She ideally gains something practical from the task in that this process is practice for her job role in Turkey. Therefore, in exploiting Skype VC and a mobile device beyond the language classroom, the experience itself is “exciting” because the technology and setting offer an opportunity to share: “my culture, food or places.”

Despite the fact that the learners are located in different physical locations, the requirement for them to forge an effective interactional space is implied in Angela’s interview comments. She notes that she felt that her fellow peer wanted to enter the cultural space of the Turkish restaurant; even though Bobbi was physically situated in the local museum and this would have been impossible. Angela’s comment about remembering information is linked to ideas about “living it at the same time.” The learner sees a purpose in communicating from beyond the classroom and in experiencing the world beyond its walls in order to learn a language. This can be contrasted with her complaint that she has difficulty remembering language when it is perhaps disembodied or disconnected from a real-world context: “When we study English, I easily forget.” The learner implies that the experience is useful. To my surprise, she considers it a ‘method’ and recognises it as an opportunity to speak in English whilst retaining her cultural identity. She displays important aspects of learner autonomy through the realization that she can glean ideas from the research experience itself. I sense that she represents an adult language learner who has experienced some alienation or lack of success through more conventional approaches to language learning. Angela seems to have the motivation to implement her own version of this type of cross-cultural exchange when she returns home.

Angela’s later comments about the levels of multitasking involved in talking, monitoring a screen, and concentrating on the objects of food further confirm that this is a challenging task for learners. This learner also demonstrates a level of self-awareness and understanding regarding her reflections about multimodal communication. For example, in hindsight, she recognises that her co-ordination of language and prosthetic deictic gesturing needed to correspond in terms of the simultaneous images which she had created for her interlocutor's benefit. Her awareness of aspects of multimodal asynchronicity operating between the modes of gesture and speech are indicated in the following comment which she makes: “so maybe it’s too late.” It seems that she has been made aware that the conveyance of accurate visuals on Skype appeared following her spoken description of the

relevant dish. She also elucidates some of the difficulties which she experienced with the device itself and acknowledges that the collaborative solution to non-understanding was made available through peer assistance.

My presentation and discussion of the multimodal transcripts and sections of learner interviews have been designed to elucidate how participants exploit multimodal forms of communication within this research context. The negotiation excerpts are principally based around learners' construction of multimodal combinations of gesture and speech which are evidenced to form triggers, indicators, and resolutions with gestures assuming a broad variety of roles. Aspects such as use of the technology and how the learner interacts with the setting itself all appear to influence how gestures and speech operate within the negotiation of meaning via mobiles from beyond the language classroom. The presentation of sections of learner interviews were designed to help to confirm, challenge, and create further insights regarding my own interpretation of the negotiation excerpts; and to illustrate how learners themselves perceived their multimodal communication.

Furthermore, I had wished to explore how learners might be supported to talk about gesture and speech in ways which could ideally help them to learn through an increased awareness as to how these modes operate in conjunction with one another and in a wider, contextualised sense.

These sections have presented the findings from a multimodal analysis of the video data and include sections of interview data. The previous presentation of these findings focussed on micro-analytic approaches to analysis via representations of excerpts of the negotiation of meaning derived from learners' tasks in multimodal transcription format. I also presented some sections of verbatim interviews, as I had wanted to analyse what the learners themselves had said. This approach afforded an opportunity to view the data from differing perspectives in ways which I had hoped would also enrich the analysis from a qualitative stance. However, I had also wanted to examine the interview data in a more formal manner and to understand these data from a broader perspective. Adopting a broader view of the data supported me to explore themes which were generated via my original questions and the transcripts; and via a more inductive approach where themes were also allowed to emerge from the interview data.

In the following section, I next discuss and present the findings from the content analysis approach to the data analysis. This approach formed the latter phases of my analysis before I triangulated the data and findings from both the multimodal transcripts and the content

analysis. I then considered these sources together in order to attempt to answer the research questions and to draw conclusions.

4.6 The content analysis findings from the interviews

Using content analysis allowed me to adopt a different research lens in order to view the interview data and to establish a series of themes. The latter stages of the analysis moved away from micro-analytic procedures of the recorded data and instead suggested that there are some trends which can be linked to the use of gesture and speech when learners negotiate for meaning via mobile devices as they access Skype VC from beyond the classroom. Qualitative content analysis is ideally flexible but also systematic as an approach. It shares features with other qualitative methods such as concern for meaning, interpretation, and the importance of context in determining meaning (Schreier, 2013).

In this next section of the findings, I will describe these themes before going on to answer the research questions based on my triangulation of the analytic transcripts and interviews.

4.6.1 Making meaning

Whilst I had personally been preoccupied with the theory and framework of the negotiation of meaning from a multimodal perspective, learners themselves never overtly mentioned this term within the data collected. Conversely, they did talk about ‘meaning’ in a more general sense; and the broader theme of making meaning was therefore evidenced across the interview data in terms of how this is conveyed, debated, and understood by language learners via their deployment of combinations of gesture and speech. Learners perceived the ability to gesture with their hands as a meaning-based resource which they used to explain and share information but to also search for the meaning of language items in an individualistic and collaborative manner. Learners linked their gestures to the meaning of individual words (iconic gestures) but also used hands (via iconic gestures) to convey and to repeat sources of information and to resolve communicative difficulties in order to complete the tasks. Learners associated gesture with a variety of different modes in terms of how they thought meaning could be achieved: gesture and the use of specific language items, gesture and the conveyance of image, and gesture and the voice.

The deployment of gestures with the hands could also frustrate and confuse learners; particularly when the meaning of an interlocutor’s gesture and speech failed to make sense or when gestures could not easily be discerned due to the mediating effects of the mobile

screen. Pointing with the device was especially complex to achieve because learners felt that they had needed to co-ordinate the meaning of their language, gesture, body; and to also take account of the structure of the environment in which they were pointing. The successful conveyance of images to interlocutors entailed the adoption of practical strategies such as the necessity to check whether or not their peer could actually see the correct object via Skype VC. For example, individual learners said that they were unable to understand exactly which images they had relayed to an interlocutor at the time of the task, causing incoherence in terms of the meaning of images for interlocutors. This was due to factors which included an inability for learners to monitor the images on their screen, pressures to multitask; and the need to sustain their communication in real time. The ability to make meaning within this communicative environment involved considerable levels of multitasking, for example, learners explained in their interviews that they had been expected to read a task sheet whilst walking, talking, pointing, and monitoring their interlocutor; and that this was demanding for them. Meaning was also identified as inherent in the objects themselves in that learners often associated these with key aspects of both their own and the target language culture. For example, learners described how they enjoyed sharing tea and scones in virtual ways; and also remarked that the technology allowed them to explore the meaning of previously unknown artefacts such as a set of Toby jugs, situated within a local museum.

4.6.2 Gesture and language use

Iconic gestures were identified by learners as a way for them to illustrate specific lexical items within their online tasks. The use of these particular gesture types was related to verbal descriptions of objects, clarification of information, association with language items; and searches for vocabulary which had either been forgotten or was previously unknown. The deployment of gesture and speech in combination could provide a strategy to support learners to remember and think about language in an individualistic sense, however, these multimodal ensembles were also designed to visually represent specific words to other learners across the technology.

The use of prosthetic deictic gestures were seen as “difficult” in that co-ordinating these with the relevant language use was deemed to be challenging. Learners noted how interlocutors could easily point towards the wrong object or create a blurred image when they first began to verbally describe a relevant feature of their setting to an interlocutor.

Gestures formed with the device and camera, and their resultant images, could also prompt learners to talk about a specific topic: “If I didn't see the picture, maybe I don't talk about this situation.”

Recording video-based communication via mobiles allows for opportunities for language learners to reflect on problematic areas of communication from multimodal perspectives. Via stimulated recall sessions and lines of questioning, most learners began to notice where and why their communication within the original tasks had been compromised. Through the researcher prompting them, some learners were able to interconnect their use of a particular gesture to their deployment of a specific linguistic item; and they went on to demonstrate an ability to repair their language use from multimodal perspectives during the process of the stimulated recall interview. This is evidenced through examples of metalinguistic discussion with the researcher where a learner is able to notice an issue and to arrive at their own solution:

(ICONIC gesture, co-occurring with a learner's use of the word ‘plate’)

So I'm confused. For you a plate is just the receptacle? So not a plate in this case. I should say instead a traditional dish. It's a false friend !

(Learner from study)

During the stimulated recall sessions, learners could also be prompted to engage in forms of metalinguistic discussion via the researcher asking them relevant questions. This could act as a way to draw the learner's attention to a particular gesture and its co-occurring language use in order to analyse what happened at the time of the task. The example which follows was generated by replaying a section of video where a learner had deployed a prosthetic deictic gesture as they discussed recent political events in their own country. In reviewing the negotiation of meaning excerpt at the time of the interview, I noticed that the learner had exploited her L1 to ask a nearby Turkish waiter to clarify a word in English. This serendipitous moment, which had been captured on film for the purposes of reflection, led to an opportunity for the researcher and learner to collaborate in order to discuss the difference between the words strike and riot.

What did you ask him in Turkish?

Ah, like strike.

And what English word does he use?

He says, riot?

Yes.

What is a riot? I don't know...

(Learner from study)

4.6.3 Affordances and constraints of technology and setting

The affordances of the device, camera; and Skype VC were identified by learners as serving the purpose of allowing them to engage with one another and to experience a range of spaces in a material and virtual sense. The use of Skype VC on mobile devices allows dyads to communicate and exchange knowledge in ways which involves them communicating across “differently organised spatial and representational domains” (Keating & Sunakawa, 2011, p. 194). Most learners said that they had enjoyed the experience of being located in an informal setting beyond the classroom with an opportunity to virtually connect to an interlocutor in order to communicate. This approach was considered “new,” “exciting,” but also “confusing” and “difficult.” Some learners fully exploited the affordances of the technology and their surroundings to complete the tasks whilst other learners overlooked the task instructions and chose to remain more static during the online exchange. They explained: “easier to stay in one place.” The use of the device and camera, in prosthetic deictic gesturing, ensured that the possibilities offered by the technology (portability, mobility of learner, and real-time video) were interdependent on the preferences but also the skills of the learner in terms of their ability to communicate successfully across the technology.

The settings in which the learners were situated offered an array of artefacts for them to explore, negotiate around, locate, and share online. These artefacts carried a series of powerful associations which could prompt learners to form sequences of iconic gestures with the hands in order to discuss and negotiate for meaning in interesting and creative ways. However, when learners communicate on mobiles from settings beyond the classroom, they are also exposed to features inherent in these types of informal settings. Learners explained that these can include background music and people talking nearby. Some learners felt that the presence of background noise and issues with the technology itself meant that gesturing formed a useful strategy in order to compensate for the difficulties encountered across Skype. Whilst the learners were required to communicate

with one another, they noted that they had sometimes been able to take advantage of serendipitous moments, for example, when they encountered staff in cafes and restaurants with whom they could converse and ask questions.

4.6.4 Intersubjectivity

The dyads taking part in this study were pushed to engage with “temporarily shared worlds” (Rommetviet, 1974, p. 29) as a way to enable them to negotiate and to complete the online speaking tasks. For example, learners’ comments during the stimulated recall sessions reflect the need for them to create a mutual space between the material and virtual world: “I feel she really wants to come here.” Relationships between gesture and space within this context also operate in unpredictable and challenging ways; and this could lead to problems for learners in terms of their ability to make sense of the online exchange. When learners glimpsed features of settings, such as books across Skype VC, they explained that they had exploited deictic gestures with their hands and fingers in order to indicate that they were interested in learning more about where an interlocutor had been situated. On the other hand, it was sometimes unclear to interlocutors at the time of the task whether a peer’s deictic gestures had signified: “here” in my setting or “there” in your setting. Aspects of indexicality based on gesture and language use become disrupted in that the listener needs to comprehend the world of the speaker in order to understand them (Scollon & Scollon, 2003).

Pointing accurately with the device and camera acted as a means for the learners to attempt to integrate two different settings as a way to establish common ground across Skype. For example, learners demonstrated that they could recognise when they had failed to establish a mutually-comprehensible space: “I don’t understand which picture she’s talking about because I showed her a different picture.” Prosthetic deictic gestures could confuse learners and cause them to misunderstand where an interlocutor had been situated at the time of the task. For example, learners might come to the conclusion that their interlocutor had been situated in a library setting whereas the learner had actually been situated within a museum/gallery space: “I think eh she’s been in a library and she showed me some art.”

Learners exploited iconic gestures with their hands as a way to establish intersubjectivity between them. It appeared to be important for learners that their interlocutors would both see and respond to gestures as a way to demonstrate that their multimodal signals had been noticed and understood. Mirror gestures formed with the hands were interpreted by

learners as a strategy designed to demonstrate that mutual understanding had been achieved between them. However, these response gestures were often missed at the time of task; leaving learners confused as to whether or not their communicative intentions and messages had been understood.

4.6.5 Peer collaboration

The importance of support and input from peers within the negotiation of meaning was evidenced in comments such as: “she can tell me to move the camera.” Learners explained that peers were able to clarify information, guide gestures with the device, give feedback on language and the images; and work with them in order to complete the tasks. When peers did not sufficiently engage or were perceived to have not worked hard enough to support the online interaction, this was viewed in negative terms. Learners enjoyed sharing objects derived from the world around them within their online collaborations. Despite the difficulties encountered, pointing to objects with the device and camera could imply: “learning together” and “helping each other.”

Learners said that they worked collaboratively, through using their hands as a resource, as a way to learn but to also ascertain where to go in order to find a specific object which had been described. They highlighted that gestures formed with the device and camera relied on help from interlocutors in order to support learners to achieve these successfully. A seeming lack of collaboration was sometimes found to be the result of factors which lay beyond the learner’s control at the time of the task. For example, unpredictable factors within an environment could affect a learner’s ability to stay focussed on a particular object: “she goes upstairs because of the noise I think.”

4.7 Conclusion to presentation of findings

This chapter has presented the findings from the learners’ interactions in transcription form and from an analysis of the interview data, focussing on the manner in which language learners interpret their use of gesture and speech within examples of their negotiation of meaning. The findings illustrate how the negotiation of meaning is enabled and resolved via learners’ deployment of multimodal resources based on their use of gesture and speech; and how learners may be potentially supported to become more aware of their multimodal forms of communication. The overall data was analysed from a qualitative perspective and, as a result, this has been presented using multimodal transcripts consisting of screenshots and written methods of transcription. I have also presented sections of learner interviews

(verbatim) and illustrated the wider themes which emerged from an analysis of the interview data as a whole. Having presented my findings from the study in detail, I now describe these in terms of how they relate to and helped to answer the research questions which I had originally posed.

Chapter 5 Discussion

5.1 Research Question 1

1) What is the role of gesture in the negotiation of meaning when learners are communicating via Skype VC accessed on mobiles from beyond the language classroom?

a) How are gestures exploited and responded to between learners during their negotiation of meaning?

The present study illustrates that the negotiation of meaning can be supported through dyads' deployment of Skype VC on mobiles in relationship to their undertaking of a collaborative speaking task. The affordances of the technology and the setting mean that speaking tasks, designed to support the negotiation of meaning, can be designed around an array of material artefacts situated in everyday public spaces such as restaurants and historical buildings. The harnessing of gesture and speech in interplay plays an important role in enabling language learners' negotiation of meaning across mobile technologies from beyond the classroom. For example, gesture forms a multimodal resource which can support aspects of SLA from individualistic and collaborative perspectives; and in ways which learners themselves highlight as important in terms of their meaning-making processes. Gestures are also valuable within this technology-mediated environment in that they act as means to support learners to engage with and gain control of a spoken language task implemented from beyond the classroom.

Language learners exploit gestures across their use of mobile technology in ways which reflect "intrapersonal problem solving" (McCafferty, 2004) whilst also enabling forms of social interaction and collaboration. From this perspective, gestures support internal cognition whilst simultaneously contributing to aspects of social communication for addressees (Gullberg, 2010). Iconic and deictic gestures (mainly prosthetic deictic gestures) appear to dominate exchanges beyond the classroom via Skype VC on mobiles. These gesture types play a role in triggers, indicators, and resolution to incidents of non-understanding in relationship to aspects of technology use and learners' interaction with a material setting. Learners exploit them with speech in order to collaborate, scaffold one another; and to establish intersubjectivity across hybrid worlds.

The present study supports previous findings within technology-mediated learning environments in that the dyads taking part were observed to engage in ‘the negotiation of lexis’ (Levy & Stockwell, 2008; Smith, 2005; Yanguas, 2010), rather than choosing to negotiate over form. Conversely, learners exploit gestures with speech in order to negotiate lexis in their video-based communication on mobile devices. The use of gesture appears to assist learners to gain control of an online exchange in terms of how they internalize language, achieve learning goals; and exploit the affordances around them. It is important to note that the type of language and gestures generated by the learners will have been influenced by the meaning-based nature of the task design. This had required the dyads to locate and discuss a series of situated objects, rather than prompting them to produce a pre-designated set of grammatical structures within their exchanges. Conversely, the findings indicate that learners can be supported post-task to engage with ‘multimodal repair moves’ (Olsher, 2008) and forms of metalinguistic discussion.

When learners negotiate for meaning under these conditions, their lexical searches (Gullberg, 1998; Negueruela & Lantolf, 2008) appear to serve a dual purpose. This is because iconic gestures with the hands are exploited as an online resource to support attempts for the individual to recall lexis whilst simultaneously functioning as a signal across Skype that peer assistance is required. As McNeill (2000) notes, “the opposition of ‘inside’ and ‘outside’ is superficial” in terms of the role of speech-associated gesture (p. 11). Gestures in this technology-mediated environment serve multifunctional purposes which, when studied, can help to elucidate how learners both think about and share language together in multimodal and meaning-based ways. Learners themselves associate their gestures with both cognitive (i.e. ‘inside’) and communicative (i.e. ‘outside’) roles in ways which can make it difficult to discern where a gesture for the ‘self’ ends and where a gesture for ‘the other’ begins.

The fine-grained analysis of gesture provides a rich insight into the unstable and continuously evolving nature of L2 language acquisition when learners negotiate for meaning across mobile technology from beyond the classroom. In terms of an analysis of the GP (McNeill, 1992; McNeill & Duncan, 2000) within the present study, learners principally used their iconic gestures as a way to “concretize the verbal channel” (McCafferty, 2004, p. 155), rather than using these gesture types to add information which was not already present within the meaning of their spoken language. The learners’ deployment of iconic gestures across mobile technology reflects McNeill’s (1992) GP in

that the *strokes* are normally co-timed to coincide with the learner's deployment of relevant, meaning-based language items during their negotiation of meaning. This is exemplified in their attempts to convey the meaning of particular objects or concepts across Skype. For example, learners synchronise their hand gestures to coincide with their use of key words which need to be seen and heard across Skype in order for an interlocutor to fully understand them. However, learners may also use gesture as a strategy to compensate for problems such as background noise or issues with the technology itself in ways which can also contribute to opportunities to negotiate for meaning.

In terms of use of technology, language learners' use of prosthetic deictic gestures challenges the notion that speakers invariably 'bring together' gesture and speech in order to create a holistic unit of meaning. For example, learners in the present study aimed to coordinate their use of the device and camera with their spoken language in order to enable an online interlocutor to comprehend their meaning. Conversely, the use of this type of deictic gesture nearly always caused them difficulty and confusion. These incidents of non-understanding also led to opportunities to negotiate for meaning based on the requirement to resolve problems based around visual communication. Prosthetic deictic gestures are difficult to achieve but can result in learners scaffolding one another and offering support. Learners themselves highlight the importance of gesture in terms of their language use but they also talk about the power of visuals within this communicative environment. Learners are also able to identify incidents in a video-captured task where their prosthetic deictic gestures and language use operated asynchronously and in dissonant ways for interlocutors.

Learners identify the importance of their deployment of gesture in relationship to meaning and the setting of autonomous learning goals; and Skype VC importantly allows for the possibility that an online peer will respond to an interlocutor's gesture and help learners to establish intersubjectivity. Learners mainly respond to iconic gestures by mirroring them across Skype; indicating that these gesture types are treated by interlocutors as a communicative 'move' or 'contribution' (Kendon, 2004). In this technology-mediated environment, learners rarely copied one another's gestures but instead adapted their interlocutor's gestures in their own unique way. Gestures are used to establish mutual understanding, however, individuals can often be unaware as to how they are gesturing, or how their gestures are received by interlocutors, at the time of a spoken task.

Unlike previous findings from face-to-face and online negotiation of meaning contexts (see Varonis & Gass, 1985; Wang, 2006; Yanguas, 2010), problems associated with language use are not the sole reason why learners choose to engage in the negotiation process beyond the classroom. Issues such as the function and use of technology, the requirement to locate objects within a geographical space; and serendipitous aspects of the settings themselves can all contribute to the creation of opportunities for the negotiation of meaning. Within the present study, the learners' online interactions entailed forms of clarification and modification, considered beneficial to SLA, however, these were frequently achieved in unconventional ways, not cited in the previous literature. Examples included learners clarifying and modifying their gesture and speech in order to help a peer to negotiate their setting or to resolve communicative difficulties engendered through the non-understanding of visuals across Skype. The multifunctional nature of gestures within this mobile learning environment dovetails with some of the previous research but also raises new questions which would require further exploration to be more comprehensively understood.

b) What are the affordances and constraints of mobile technologies, and the settings they enable, in relationship to learners' deployment of gesture during their negotiation of meaning?

This study shows that use of Skype VC on mobiles within settings beyond the traditional language classroom results in opportunities for learners to engage in the negotiation of meaning as they socially interact on Skype and simultaneously engage with the settings around them. Learners' use combinations of digital tools and their immediate surroundings allows peers to connect with one another without the presence of a teacher at the time of a task. Conversely, these types of learner-centred scenarios beyond the classroom, enabled by mobiles, also adhere to existing principles identified within communicative approaches to language teaching and learning. For example, they encompass learner-centred communication, opportunities for speaking and listening, reciprocity of interaction, negotiation, synchronicity; and unpredictability (see Beatty, 2013, p. 7). Operating beyond the traditional language classroom allows learners taking part to assume autonomy and to become actively involved in exploiting the everyday world around them in ways which can be used to support their interactions, according to principles within the negotiation of meaning (Long, 1996; Varonis & Gass, 1985). As Kukulska-Hulme (2012) notes, the

portability of the device offers distinct advantages and enables learners to relate their language learning to a specific physical context.

In the present study, learners used the affordances of Skype VC to show objects and to create authenticity (Austin et al., 2017; Wang, 2006). However, they achieved this in different ways when compared to earlier findings, derived from the analysis of learners' interactions across desktop communication. For example, the capacity to convey meaning through the camera across Skype, the portability of the device, the mobility of the learner; and the use of affordances such as situated artefacts extend opportunities for language learners to interact with the world around them.

In the present study, the learners' engagement with aspects of the context itself supports previous findings in MALL which have demonstrated that mobile devices can be used to enable L2 learners to interact and communicate with the context in which the mobile learning task is occurring (Al-Shehri, 2011). Learners can be prompted to avail of learning resources which are already present in everyday life but which are too often overlooked by teachers and learners (Wong et al., 2012). The affordances inherent in the use of Skype video on mobiles means that the learning 'experience' itself has the potential to 'become mobile.' Learners have the opportunity to oscillate between different contexts in ways which will influence the way in which their learning unfolds (Pegrum, 2014). In the present study, the type of lexis which learners chose to negotiate over was dependent on the particular context/s (physical or virtual) in which they found themselves. However, in order to forge an effective interactional space, learners are required to work together to establish sufficient levels of intersubjectivity as they trigger, sustain, and attempt to resolve incidents of non-understanding. In engaging with tasks from beyond the classroom, learners also face unique challenges such as possible exposure to background noise and difficulties in manoeuvring the technology in relationship to aspects of their settings. However, peers in the present study were frequently encouraging and supportive of one another.

In the present study, learners co-orchestrated a range of affordances which included their use of gesture and speech but which also highlighted the importance of them understanding how modes such as visuals and spatial layout operate. It would appear that the materiality of mobile screens and the deployment of Skype VC beyond the classroom can both enable and constrain language learners' use of gesture. The findings support the notion that

technological devices will inevitably mediate aspects of an interaction and, from this perspective, mobile technologies cannot be said to represent ‘neutral conduits’ (Kern, 2014). In the present study, the act of pointing is also disrupted because there are no conventional ‘cardinal directions’ (Haviland, 2000) for learners to follow as an object is being indicated. Whilst a learner may choose to direct an interlocutor’s attention to a particular referent within their environment, there is no guarantee that the same referent will appear on Skype VC for an interlocutor.

In terms of the deployment of iconic gestures in the present study, learners adopted a variety of strategies which encompassed gesturing with both hands or holding the device in one hand and gesturing with the other. Some learners remained seated during tasks and did not choose to explore their surroundings but wished to watch an interlocutor do this instead. The findings reflect Hampel and Stickler’s (2012) earlier conclusions that learners’ use of technology results in their adaptation of multimodal tools. The adaptation of technology in the present study could result in a learner choosing to exploit a portable device in a non-portable manner, for example, through remaining seated or standing still outside a museum as they watched an interlocutor explore a space without sharing the affordances within their own environment.

Within the negotiation of meaning, the boundaries of the mobile screen normally allow for the gestural *stroke*, with the hands, to be viewed by an interlocutor. However, the materiality of the device itself (screen, size, weight) influences how gestures are both executed and received by online peers in this communicative environment. For example, the face and hands appear enlarged on a smaller mobile phone screen whereas tablet and 2-in-1 devices tend to frame the learner’s wider body; and they encompass a view of the head and shoulders with some limited information revealed about the wider context to interlocutors. Views of interactants on Skype VC on mobiles, and the visibility of gestures, are also dependent on the proximity of the device in relationship to the position of the individual learner in relationship to their interaction with an environment. The visual appearance of a gesture with the hands is affected by factors such as whether the learner has chosen to place the device on a table, leaving both hands free, or has instead decided to hold their device in one hand and gesture with the other as they walk and talk. Some learners commented on the difficulties which they experienced in terms of the demands to multi-task within this type of exchange from beyond the classroom.

The noticing and comprehension of the meaning-based content of iconic and deictic gestures is dependent on how accurately a learner positions their hands in relationship to their ongoing awareness of the constraining boundaries of the mobile screen. From this perspective, gestures need to reflect the speaker's aims and to demonstrate their ability to use these expressive resources; although they are always constrained or enabled by aspects of context (Kendon, 2004). In the case of prosthetic deictic gestures, the device and camera act as an artificial device which can be used to entirely replace the pointing finger or which can be deployed in conjunction with it in order to highlight an object for an interlocutor. Pointing involves "the entire bodies of the co-participants [...] (dis) aligned orientations, and the (re) arrangements of their postures and mobility within the interactional space" (Mondada, 2014, p. 121).

Learners in this study were pro-actively encouraged to explore the settings in which they were situated. They used the particular sets of circumstances within the different settings to forge social relationships with one another and with a series of objects (Benson, 2011). However, learners also directed their own interactions, for example, they might choose to explore a space more spontaneously whilst still generating opportunities to negotiate for meaning. Learners deploy gesture and speech together in order to negotiate around objects which they associate with everything from characters in American films to memories of events within their own country. In the beginning, these objects are seemingly inanimate or 'frozen' (Norris, 2004), however, they contain meaning-potential which is then brought to life through peer collaboration as learners interact with them out in the world both in physical and virtual ways. The constraints of the settings encompass issues such as the requirement for learners to negotiate the setting itself, for example, they talk about the need to perch on stairs and to comprehend sources of light as they point at an object. Within these spaces, there will always be random people coming and going with background noise a feature of informal settings at busy times. However, in communicating from beyond the classroom, learners also gain possibilities to avail themselves of nearby staff in cafés and restaurants, for example, practising their English with them and asking them relevant questions.

5.2 Research Question 2

To what extent can a supported focus on gesture and language use help learners to understand aspects of their multimodal communication during the negotiation of meaning?

The aim of the supported focus on gesture and language use through stimulated recall in this study was designed to gain a learner perspective on the interactions. The sessions were also based on exploring ways to support learners to better understand aspects of their multimodal communication. Given that gestures have been shown to assume a role in second language acquisition, I was interested in exploring the extent to which learners could be helped to ‘notice’ (Schmidt, 1990) and talk about aspects of their gesture and language use in interplay. It has been suggested that most gestures are typically performed with a lack of awareness but that in order to facilitate SLA, they require understanding (Gullberg, 1993). Furthermore, the notion that learners require extended skills, based around multimodal literacies, has been argued within both digital and non-digital communicative environments (see chapter 2, section 2.9.3).

Combinations of video, questions, and forms of discussion with the researcher/educator can be combined to enable individual learners to notice and comment on areas of their multimodal communication on Skype VC via mobiles within the negotiation of meaning. Watching replays of sections of the task via video, from a gesture-speech perspective, helped to raise the learners’ consciousness in terms of enabling their understanding as to how and why they had deployed gestures with speech. As a researcher and educator, I quickly learned that forms of flexible questioning and discussion are critical in supporting learners to talk about how they use gesture and speech. I asked them about gestures, and within this process, drew attention to issues which had affected their communication at the time of the task.

The reflective approach adopted in the present study, gave learners an opportunity to assess and reconsider their communication in ways which would not have been possible at the time of the task due to the pressures of real-time communication across technology. However, the learners who took part in the present study demonstrated that they were capable of more than reflection and analysis of their use of language in isolation. Stimulated recall can support

learners to better understand aspects of their multimodal communication within the negotiation of meaning, however, learners also need various degrees of structuring and input in order to achieve this.

I had initially considered that my principal role was to bring up points of attention in the video and to ask semi-structured questions but I increasingly found myself engaged in dialogue with the learners. Analysis of the video from multimodal perspectives highlighted relationships between gesture and speech in ways which prompted examples of ‘meta-linguistic awareness’ (Thomas, 1988) from the learners. Within the interviews, learners demonstrated an interest in the appropriacy and accuracy of their use of language at the time of the task, however, due to pointing at objects with devices and camera, opportunities also arose to discuss the interplay between language and their conveyance of image. The stimulated recall process became collaborative as we jointly explored excerpts of the learners’ negotiation of meaning through the affordances of the captured video. It has been previously argued that there should be a separate discussion level which is distinguished from the level of action-based tasks to encourage language learners to evaluate their communication and to draw conclusions (Cumming, 1993; Skehan, 1998). More contemporary pedagogic frameworks within MALL, have suggested that a ‘mobile-captured task’ can afford an ‘authentic artefact’ for the purposes of reflection, as teachers and learners collaborate together in order to review this (Kukulska-Hulme et al., 2017).

Theories of ‘multimodal communicative competence’ in second language learning contexts have highlighted the importance of learners understanding how modes operate together in order to create a more holistic expression of meaning (Royce, 2013). Analysis of the interplay between modes is similarly reflected in McNeill’s (1992, 2000) earlier theoretical argument that gesture and speech ‘come together’ to form a wider unit of meaning for speakers. However, coherence and the creation of a holistic unit of meaning are not inevitable features of gesture when learners deploy these across mobile technologies from beyond the classroom. Conversely, a reflection on gesture and speech can prompt learners to begin to analyse the interplay which occurs between the two modes in order to better understand how these need to form a coherent unit of meaning to be understood by an online interlocutor. The use of stimulated recall supports learners to observe how they use gesture individually, but also enables understanding as to how an interlocutor has experienced and interpreted the results of their gestures at the time of a spoken learning task via mobiles: “Now I see that it didn’t work.”

Learners taking part in the interviews had been keen to explain the reason why their inaccurate prosthetic deictic gestures had ensured that the meaning of their language use had sometimes become incoherent for an interlocutor. Whilst gesture with the hands has been traditionally linked to the use of language, in the case of prosthetic deictic gestures, learners also attribute meaning to their use of visuals in relation to language use across Skype. From this perspective, learners' interpretation of the meaning of gesture varies in terms of the different aspects or functions which they choose to attribute to them (Sime, 2008).

Rather than focus on linguistic structure or form in the traditional sense, I found that a reflection on gesture could act as an entry point to enable multimodal repair work and forms of meta-linguistic discussion. The ability for learners to identify, analyse, and manipulate language forms is considered to be important within TEFL paradigms, however, this can be approached from a contextualised, multimodal stance. As Gullberg (2010) remarks in her extension of paradigms within SLA, gestures can illuminate the nature of representations of knowledge at any given moment in time; as well as reflect the changes in representations characteristic of acquisition (p. 79). A supported focus on gesture and language use via the use of video-captured tasks from mobiles is a valuable means to help learners to begin to notice and comprehend aspects of their multimodal communication, resulting in a variety of different learning outcomes. In the context of the current study, learners themselves used the sessions to ask questions about language and to explain the advantages, but also the difficulties, which they experienced in exploiting mobile technologies from beyond the classroom.

The opportunity for learners to engage in reflection appears to enable them to gain the necessary time and space in order to explore examples of the dynamic interplay between gesture and language use, derived from real-time tasks which have taken place in authentic settings. Mobile technologies, and the possibilities to capture TBLL interactions beyond the classroom, can be used to support exploratory pedagogic goals based around raising learners' awareness of multimodal forms of communication across Skype VC. The current lack of an established pedagogic meta-language with which to talk with learners about the complexities of gesture-speech communication within this learning environment is also fully acknowledged.

5.3 The role of gesture

The present exploratory research study has examined the multiplicity of roles which gesture assumes when learners engage in the negotiation of meaning through online communication via mobiles, used beyond the classroom. Gesture analysis can contribute to our understanding of theoretical SLA in relationship to elucidating how learners choose to deal with communicative issues multimodally. Confusion about the meaning of gestures, geographical location; and problems with technology such as the sound failing on Skype VC, can all contribute to opportunities to negotiate for meaning.

Iconic gestures assume a range of multifunctional roles which seem to indicate that they serve both individual and communicative purposes for learners. The first part of question 1 (1a) illustrated how learners harness gesture and speech in ways which are designed for the purpose of supporting their individual cognition and shared communication. Gestures with the hands and prosthetic deictics encourage and support the negotiation of meaning via the use of mobile devices from beyond the classroom within examples of triggers, indicators, and resolutions. However, prosthetic deictic gestures play a particularly critical role within this communicative environment; although these gesture types can be difficult to achieve for individual learners. Prosthetic deictic gestures rely heavily on collaboration in order to help peers to establish whether or not an interlocutor has accurately conveyed the correct object across Skype.

The second part of question 1 (1b) has shown how the affordances inherent within the technology and the settings enable learners to share objects and collaborate with one another in ways which support their negotiation of meaning. Operating beyond the classroom in this way importantly allows learners to assume autonomy and become actively involved in exploiting the everyday world. But the use of technology also constrains gestures as there may be problems in terms of their visibility across the restricted boundaries of the mobile screen; and use of technology can result in combinations of image and language use becoming incoherent in terms of creating holistic meaning for interlocutors. It was found that the settings themselves are rich in affordances, due to their authenticity and the linguistic and cultural associations which learners derive from the situated objects within these settings. Peers may scaffold learners through gestures, based around enabling them to locate and talk about objects in a particular space. Interactions within this environment can also result in learners being exposed to random

events such as background music playing and problems inherent in the functioning of the technology. However, learners can use these incidents of non-understanding as an opportunity to work together in order to rectify problems and to establish greater levels of intersubjectivity. There are also valuable people such as waiters and bar staff, located in the settings, who can act as resources when learners need to translate a word.

Question 2 highlights the value in working with learners to help them to understand aspects of their multimodal communication, based around their use of gesture and speech. In this study, the learners were offered time to reflect and talk about what they felt that they had been attempting to achieve through their deployment of combinations of gesture and language. Through an exploitation of stimulated recall, it was possible for me to better comprehend how the learners themselves had experienced the interactions at the time of the task; and to support them to understand their multimodal communication through forms of post-task discussion based on us reviewing the video together. Through questioning and dialogues, based on the highlighting of examples of gesture and language use within the video clips, learners were enabled to notice aspects of their multimodal communication and this involved some of them engaging in multimodal forms of repair work and meta-linguistic discussion. Supporting learners through the use of tools such as stimulated recall can help them to talk through their linguistic and wider communicative goals from multimodal perspectives; and the approach enables them to better understand how gesture and spoken language use is required to operate in concert in order for interlocutors to comprehend holistic meaning. Stimulated recall could be said to represent much more than a methodological research tool. It has wider reaching implications in terms of its potential use within exploratory forms of pedagogy that accommodate a review of tasks implemented from beyond the classroom via mobile devices. The extent to which the approach works may also depend on variables such as learners' existing levels of 'multimodal communicative competence,' the attributes of gesture which they choose to focus on; and the degree to which the educator working with them understands how multimodal communication might operate for language learners in technology-mediated learning contexts. In the present study, the researcher worked with the learners and engaged them in forms of discussion in order to try to help them to learn.

In the final chapter of this thesis, I will first discuss the pedagogic and methodological implications of the research. Following this, I will evaluate the impact and contribution of this work, including its limitations and suggested future directions for research in the field.

Chapter 6 Conclusions

6.1 Introduction

The findings from the current study were derived from synchronous online exchanges which occurred between adult language learners across Skype VC and the subsequent use of stimulated recall interviews, based around these exchanges. The Skype platform was accessed on mobile devices which included learners' use of tablets, 2-in-1 devices, and mobile phones from beyond the classroom. The principal aim of the investigation was to examine the role of gesture within the negotiation of meaning within an environment which unusually involved the dyads communicating from settings which included cafes, restaurants, museums/ galleries, gardens, and historical buildings in order to complete speaking tasks. However, dyads varied in terms of whether or not they chose to adhere to the outline of the task. One principal outcome of the research has been to argue for a reconsideration of the theory of the negotiation of meaning within SLA from wider, multimodal perspectives. It would appear that gestures play a crucial role in areas considered to be important within interactionist and emergent accounts of language acquisition such as input, output, and noticing (Gullberg, 2010). There are examples in the multimodal transcripts of learners attempting to clarify and modify their input and output through the use of iconic and deictic gestures in ways which challenge the notion that interactional 'moves' considered beneficial to acquisition are achieved through language use alone. Gestures also supported learners to notice multimodal features of their input and output, particularly during the interview phase when learners were offered an opportunity to review the interactions. Having interviewed the learners who took part, it would appear valuable to include them as key stakeholders in a project due to the insights which they bring to this type of exploratory multimodal research.

In adopting a multimodal research lens, it was possible to reveal the multi-faceted role of gesture in relationship to learners' use of language via mobiles from beyond the classroom. By adopting the perspective that gestures occur with, or in relationship to, L2 language use, it was possible to observe the dynamic interplay that technology affords but to also understand when relationships between gesture and speech begin to break down for learners. Another research outcome is to have revealed the potential value, but also the sheer complexity and variety of roles, which gesture assumes when executed via mobiles from beyond the classroom. In adopting a gesture-speech unit of analysis (McNeill, 1992;

McNeill & Duncan, 2000) within the framework of the negotiation of meaning, it was possible to reveal the multimodal manner in which learners employ and engage with gesture and speech in order to achieve their negotiation. By expanding a view of the negotiation of meaning beyond the researcher's interpretation, it was also possible to gain a learner-centred perspective and to try to support learners to better understand their multimodal forms of communication in ways which I considered were important if pedagogical conclusions were to be drawn. In combining forms of micro-analysis of the multimodal interaction data with an exploration of the broader picture of themes in the learner interviews, I was enabled to triangulate the data and see different perspectives in my attempts to answer the research questions.

In hindsight, I realise that there were also difficulties and limitations experienced with regard to the implementation of this type of research study, especially in terms of technology. However, since the data collection phase in 2016, the technology has advanced considerably in ways which today would make the collection of Skype data from mobiles beyond the classroom considerably easier. Furthermore, recent technological developments inherent in Skype software has implications in terms of the ease with which teachers and learners would be able to access the online platform 'on the move' to enable communicative exchanges beyond the classroom and for learners to record these in order to engage in forms of reflection.

6.2 The pedagogical implications

Teachers considering the use of Skype VC on mobiles need to recognise that an emphasis on linguistic skills in isolation does not reflect the multimodal nature of learner communication from beyond the language classroom. The pedagogic perspective adopted within the present study, views language use as multimodal with a suggestion that teachers pay attention to learners' use of gesture within online tasks/exchanges with the same attention normally afforded to language use. The settings in the present research study provided resources and social opportunities in order for learners to interact and learn from their surroundings in multimodal ways. Skype 'on the move' importantly enables communicative skills to be practised in contextualised ways with teachers potentially playing a role in the promotion of multimodal viewpoints on SLA, task design for beyond the classroom; and encouragement for learners to reflect from multimodal perspectives:

“[t]he contexts of language teaching, like the more social contexts within which they are located, are continually changing, continually changing habitual ways of thinking” (Widdowson, 1992, p. 2). It is perhaps important to note that during the period of gathering the data for the study, there were some problems experienced around capturing the tasks. This was mainly due to a lack of available capturing tools which had been specifically designed for mobiles and the requirement to rely exclusively on the third-party software tool, Supertintin (created for use with Windows software), in ways which could be unreliable. At the time of the research, the technology for recording Skype VC on mobile devices was extremely limited, and, unfortunately, incompatible in terms of downloading this to certain ubiquitous operating systems such as *Apple* and *android*. This limitation would have severely restricted the possibilities for teachers to implement and record Skype-based tasks in MALL. It also appeared to be difficult to save recordings and then share them across different devices. Fortunately, these problems have since been addressed within the latest developments in Skype 8 for mobiles. The latest developments in Skype (see <https://www.skype.com/en/get-skype/skype-for-mobile/>) enable learners to connect much more easily and contain inbuilt video-capturing software with which they can record L2 tasks and interactions from beyond the classroom. This updated software now enables an exchange from a mobile device to be video-recorded to a better standard (in terms of visuals and audio) in ways which may help teachers to more clearly evaluate examples of speech and gesture used by their language learners. Furthermore, the current camera ‘view’ now available via the inbuilt Skype capturing software is limited to a ‘side-by-side’ for those interested in harnessing gesture in L2 learning tasks. Having experimented with different ‘views’ during the course of the present study, I came to the conclusion that the ‘side-by-side view’ affords the most valuable visual perspective in terms of enabling teachers and researchers to both observe and analyse a dyad in terms of their collaborative use of gesture and speech across Skype VC.

From a pedagogical perspective, software developments ensure that Skype video calls can now be recorded by learners and saved with a right click, creating an MP4 file, which can then be shared with teachers or between peers via email or through the Skype site itself. There is now the option for video-recorded tasks to be reviewed by teachers and learners alike within the Skype site itself with opportunities to send comments via chat modes or to share documents such as feedback on tasks before meeting with learners. I consider that teachers would also need to address any possible issues of trust and confidentiality which

might arise in terms of video recording interactions and in allowing others to access this data. It would be important to explain the possible learning outcomes of implementing tasks from settings beyond the classroom; highlighting the benefits of a reflection on tasks such as raising awareness of linguistic issues multimodally through forms of pedagogic support.

A consideration of the role of gesture is relevant to communicative approaches to L2 language teaching and learning because the mode critically depends on “a social other” whilst being “a dynamic element of one’s individual cognition” (McNeill, 2005, p. 53). From this perspective, a study of the interface between gesture and speech within negotiated interaction can shed light on features of peer collaboration whilst providing multimodal information in terms of which items of language an individual learner may be trying to recall or express at the time of a task. Having reviewed my research findings, it becomes evident that some learners can struggle with multi-tasking demands involved in following a set of task instructions as they walk around a space to communicate with one another learner. I consider that it would have been more productive to have explained the task goals in advance and to have uploaded a task document to Skype itself; rather than expect learners to carry around and consult a separate sheet of paper. It is probably necessary to give learners at least twenty minutes before a task is due to begin to allow them time to process the information and to discuss with a teacher or peer what they think they are required to achieve. Teachers might also consider when a specific setting might be particularly crowded or noisy, for example, at lunchtime, and, as a result, choose to implement the task at a quieter period in the day. It is essential that teachers visit the locations in advance and speak to those in charge in order to confirm that learners will definitely be made to feel welcome and not asked to stop talking or recording mid-task. Taking gesture into consideration in relationship to spoken language appears to offer a multimodal route for teachers to more fully understand oral L2 within frameworks such as negotiation of meaning. This is in terms of elucidating how processes within L2 development unfold within particular learning sequences such as the negotiation of meaning (Faraco & Kida, 2008) whilst enabling an understanding of the broader manner in which learners experience communicative difficulties from beyond the classroom. Whilst there have been arguments that multimodal language learning with technology should concern itself with how to avoid “distortion” and “breakdown” (Chanier & Lamy, 2017, p. 431), this stance deprives learners of opportunities to modify and clarify aspects of their

input and output in ways which indicate benefits for their acquisition of language.

Moreover, problems in communication can also enable learners to better understand how they construct their multimodal communication in terms of their deployment of modes such as use of visuals across Skype.

It would appear that learners' use of iconic gestures in particular have a direct bearing on the meanings which they attach to specific lexical items as they engage in negotiation. A review of these gesture types within a recorded exchange may also help teachers to understand how learners work through different linguistic options and the points where they need might require further support through pedagogic input. Some ability to recognise the appearance of specific gesture types, and recognition of the meaning-based gestural *stroke*, may provide teachers and learners with a structured and tangible meta-language with which to begin to discuss the holistic interplay between gesture and speech in online learning contexts. It is evident from the findings that learners will require degrees of support in order to start to talk about their gestures; and that teachers would first need to acquire these skills in order to help to support their learners in a variety of ways.

The findings from the present study also have ramifications in terms of the teacher's role in the development of learners' multimodal and digital literacies. For example, Stickler and Hampel (2015) discuss new skills for language teachers and suggest that learners should be encouraged to develop the capability to communicate in a variety of interactional spaces. The study of learners' gestures with the hands could also become a focal point for teachers and learners working via desktop forms of communication in that the use of iconic and deictic gestures will have a bearing on how learners negotiate for meaning in these learning environments as well. However, Kukulska-Hulme (2013) has argued that the nature of language learning with mobiles raises new questions in terms of how we 're-skill' learners to help foster skills such as autonomy in the twenty-first century. The act of prosthetic deictic gesturing could arguably be seen to represent a 'new skill' for language learners operating from beyond the classroom via mobile devices. This gesture type is unique to the affordances of mobile technology and reflects the notion that digitization has "reshaped the communicative landscape" in ways which call for new forms of L2 literacy but also multimodal teaching practices (Lotherington & Jenson, 2011, p.226) The execution of this particular gesture type appeared to be an unfamiliar practice for most learners, however, it was considered important by them in that its accuracy ensured whether or not relevant objects could be seen and shared across Skype. This gesture type impacted how language

use was received and understood by online interlocutors from both linguistic and visual perspectives in ways which influenced the trajectory of their negotiation. The technology itself has better evolved to meet the requirement for people to point at objects with an artificial device through use of a front and back camera since the research took place.

Spoken exchanges via Skype on mobiles from beyond the classroom operate multimodally and in ways which can support interactions considered beneficial to SLA according to the negotiation of meaning. However, learners will probably struggle to communicate accurately and to remember their use of gesture and speech from the time of the task; due to the spontaneous nature of gestures and the significant pressures of multi-tasking involved in communicating across Skype VC from beyond the classroom. For this reason, a pedagogical task-based cycle for the deployment of Skype VC on mobiles from beyond the classroom could fall into two task-based cycles (see Figure 28).



Image redacted

Figure 28: Task-based cycles in MALL

In terms of support, teachers may choose to exploit gesture in the ways which they feel will most benefit their particular learners, however, I suggest the following areas as potentially valuable:

- Identification of multimodal examples of input and output.
- Identification of problematic aspects of L2 communication.
- Identification of examples of multimodal coherence and incoherence.
- Identification of lexical searches.
- Support for learners to understand the interplay between gesture and L2 speech.

- Support for learners to forge connections between semantic representations of vocabulary and co-occurring language use.
- Support for learners to comprehend visual messages conveyed across technology in relationship to language use.
- Opportunities for multimodal repair work and meta-linguistic discussion.

In the following table (Table 10), I propose a multimodal pedagogic framework for teachers interested in working with learners through attention to gesture and speech within online learning environments.

Table 10: A multimodal pedagogic framework based on gesture and speech within technology-mediated communication

Multimodal forms of input and output	Review gesture types used to achieve this: iconic, deictic, prosthetic deictic gestures
Vocabulary items which arise multimodally	<p>Review gesture types: iconic (online) and prosthetic deictic gestures (mobiles only)</p> <ul style="list-style-type: none"> • Ask the learner to examine how they/ their interlocutor has expressed the meaning of their language use with their hands. • Ask the learner what language they/ their interlocutor was looking for during lexical searches with their hands. • Ask the learner to explain what they/ their interlocutor was attempting to focus on as they pointed with the device and what language they used (mobiles only).
Interplay and coherence of meaning between modes	<p>Review gesture types: iconic, deictic (online) and prosthetic deictic gestures (mobiles only)</p> <ul style="list-style-type: none"> • Ask the learner to examine whether or not modes made sense in relationship to one another. • Ask the learner to explain the relationship between the use of visuals and language use.
Repair work and meta-linguistic discussion	<p>Review gesture types: iconic, deictics (online and mobiles)</p> <ul style="list-style-type: none"> • Ask the learner to look at the gesture and talk about the accuracy of their language use. • Encourage the learner to ask questions. • Offer further linguistic suggestions and give contextualised examples.

6.3 Methodological implications

The methodological tools employed in the present study encompassed use of the Varonis and Gass (1985) framework of the negotiation of meaning and a gesture-speech unit of analysis (McNeill, 1992; McNeill & Duncan, 2000), alongside unfolding images (see Norris, 2004). It is important to acknowledge that the Varonis and Gass coding scheme is normally deployed to examine spoken language in isolation; rather than to take account of examples of gesture and speech operating together within a framework of negotiation. However, the close connections which speakers forge between gesture and speech (Kendon, 2004; McCafferty & Stam, 2008; McNeill, 1992, 2000; McNeill & Duncan, 2000; Norris, 2004), and the findings of this study, suggest that the Varonis and Gass framework is capable of supporting the multimodal analysis of both modes operating together. The taxonomy of gesture types and procedures such as transcription of the gestural phrase (McNeill, 1992; Norris, 2004) were not originally designed to analyse features of SLA but have since then been deployed to examine SLA in face-to-face contexts (McCafferty & Stam, 2008). To my knowledge, at the time of writing, none of these methodological procedures around gesture had ever been deployed in order to take account of interactions across Skype VC via the use of desktops or mobiles from beyond the classroom. However, in exploiting a recognized framework in SLA and adapting a standardized transcription system (McNeill, 1992; McNeill & Duncan, 2000), I hope that it will be possible to build on this work and allow for some comparability across studies in the future.

Despite being several decades old, the Varonis and Gass framework still represents a valuable methodological tool for researchers interested in making sense of language learners' online talk in relationship to an interactionist account of acquisition. Conversely, when combined with a gesture-speech unit of analysis (McNeill, 1992; McNeill & Duncan, 2000), the model can be methodologically adapted and extended to support the researcher to analyse the multimodal manner in which learners deploy individual gestures, and then sequences, in relationship to their stretches of negotiated talk. Exploiting an SLA framework such as the negotiation of meaning, when implemented as a multimodal hybrid, enables the researcher to take account of gesture and speech operating in modal interplay and highlight the multi-functional ways that language learners "move the hands [...] in conjunction with speech" (McNeill, 2000, p. 1). Conversely, it can also be used to determine how language learners point with devices, talk, and convey visuals across Skype

to interlocutors. However, the learners' use of prosthetic deictic gestures in this study frequently lacked a clearly-definable *preparation* and *retraction* phrase with learners' focus on the object and the ensuing image deemed to represent the gestural *stroke*. As in gestures with the hands, the aim of pointing with the device is to seemingly to create a holistic unit of meaning for an interlocutor or to support the learner themselves to talk about what they are showing. From this perspective, the use of a mobile device and camera can support the negotiation of meaning and, as result, pointing with 'an artificial device' (Clark, 2003) under these sets of conditions has relevance to aspects of SLA. Conversely, the manner in which prosthetic deictic gestures operate in relationship to L2 language use is quite different to the manner in which learners' co-ordinate gestures with their hands as they speak. Unlike the inevitable GP, constructed with the hands, learners appear to co-construct prosthetic gesture types in much more chaotic ways which can mean that the learner's manipulation of the device and camera might involve several attempts. This gesture type requires levels of scaffolding from interlocutors as they try to ensure the correct visuals appear on Skype. The nature of prosthetic deictic gesturing ensures that establishing their relationship to aspects of L2 talk is more challenging than the analysis of gestures with the hands alone. This is because the gestural *strokes* formed with the device and camera will not inevitably fall on key words, with a meaning-based content, but may instead involve the researcher analysing a set of images which may bear no relationship to what the learner has actually said. In hindsight, it would also have been more productive to have exclusively focussed on analysis of the gestural *strokes* and language use in that learners exploit their mobile devices to point in too diverse a manner to enable the researcher to easily discern the *preparation* and *retraction* with the device.

In terms of a focus on acquisition, I suggest that it is also productive for the multimodal researcher interested in language acquisition to transcribe sections of data where learners' gestures and speech become disrupted or where the interplay between modes appears to make little sense for interlocutors. This is evidenced in the researcher's recognition of lexical searches with the hands (Negueruela & Lantolf, 2008) but also in terms of paying attention to how learners attempt to co-ordinate language with achieving images via deictic gestures with a mobile device. Within online learning environments, recognition of gesture types and the gestural phrase need to be considered in relationship to the specific affordances and limitations of the technology, for example, mediating screen size and aspects of the settings need to be considered in relationship to the use of gesture and talk. If

learners are operating beyond the classroom, gestures can be viewed from the perspective that their gestures act as meaning-based resources which they co-construct as a way to talk about aspects of the world around them.

The findings from the study suggest that integrating methods derived from gesture studies within a framework of the negotiation of meaning is valuable in that it enables multimodal forms of transcription whilst challenging traditional notions of SLA. An interdisciplinary approach to the methodology served the purpose of elucidating the multi-faceted ways in which language learners deploy gesture in relationship to their language use and helped to explicate its role in the negotiation of meaning. The highly complex multimodal manner in which processes within SLA operate for learners across their use of mobile technologies from beyond the classroom is also made apparent via the methodological approach.

However, the findings also indicate that in applying an interdisciplinary methodology, in an exploratory sense, there is a lack of refinement in terms of how the gesture-speech unit of analysis operates within this mobile learning environment. I suggest that approaches to multimodal transcription of gesture and speech in relationship to learners' use of technology requires greater levels of refinement; particularly in terms of the researcher's ability to transcribe how gestures such as prosthetic deictics are constructed by different learners across an array of mobile devices and settings. A more complete understanding of how pointing is achieved with the device would ideally enable more refined and generalisable transcription procedures for the future. Learning how to analyse language learners' gestural meanings in relationship to technology use may represent a developmental research skill which could also support our understanding of the kinds of visual literacies which language learners, and teachers, may increasingly have to acquire in the future.

Multimodal transcription procedures in the present study allowed for an analysis of the different modes operating together but also allowed for the researcher to present examples where learners' gesture and speech became asynchronous and incoherent in meaning. The use of screenshots in the multimodal transcripts enabled gesture types, constructed with the hands, to be shown in relationship to the gestural phrase. In terms of the transcription of prosthetic deictic gesture types, the learners' focus on particular aspects of their setting, and how this appeared to interlocutors, could usefully be revealed to the reader of the transcript via inclusion of screenshots in ways which could indicate aspects of context. However, it is important to note that in recently publishing this research in an applied

linguistics and technology journal (see Lee et al., 2019) the multimodal nature, length, and layout of the transcripts caused problems; resulting in the negotiation excerpts having to be heavily edited. This was in terms of reproduction of the screenshots and the space which these transcripts would take up in the pages of the journal. This is an area where researchers and publishers would benefit from working closely together.

The negotiation of meaning is achieved in highly complex and distorted ways across Skype VC on mobiles from beyond the classroom. Varonis and Gass' original analytic framework involves the researcher coding examples of non-understanding through determining triggers, indicators, and resolutions; based exclusively around the learners' attention to language use. In a mobile, technology-mediated interactions beyond the classroom, the implementation of the same coding scheme involves complicated decisions around how best to define seemingly blurred areas which would not have arisen had learners been completing a task within a classroom. For example, my coding decisions entailed readdressing concepts such as Varonis and Gass' 'comprehension check' category in terms of determining how this particular move operated for learners via Skype VC on mobiles from beyond the classroom. This was difficult in terms of the reliability of double coding with a lack of previous data examples to draw on. However, I argue that comprehension checks within this communicative environment involve learners attempting to establish intersubjectivity through confirmation as to whether or not an interlocutor can clearly see a specific object across Skype. It becomes clear from the research that there is much greater work required in order for researchers to develop effective and generalisable methods to both transcribe and discuss learners' use of visuals across forms of technology in relationship to language use. Furthermore, the study shows that the manner in which the coding categories within the original framework were designed to function was transformed by combinations of the learners' use of technology and the settings in which they were situated. For example, learners within the present study used forms of clarification such as repetition, based on an interlocutor's previous turn, to try to find out what certain objects looked like and to comprehend where particular objects were situated: "It's a face of a man. He goes outside of the wall. It's near to the door?" If spoken turns were examined without taking account of gesture and the wider contexts in which the interactions take place, the analysis would only be able offer a partial, and sometimes potentially misleading account, of the types of interaction which occur between language learners under these sets of conditions.

I conclude that learners' use of online gestures which occur with speech can become the subject of scrutiny by the researcher, and learners themselves, if interactions across technology are video-recorded. However, certain gestures will be lost due to constraints such as the boundaries of the mobile screen. Stimulated recall can be exploited as a qualitative methodological but also pedagogic tool which can serve as a multimodal reminder of the original event (the task). Combinations of video stimuli, questions, and encouragement of dialogue act in ways which can prompt learners to reflect on the use of gesture and emergent language use in relationship to aspects of their deployment of technology; and in terms of their interaction with the world around them. I argue that without the researcher exploiting some form of video recording, language learners will be unable to talk about their gestures which occur with speech because these are formed spontaneously, as an online exchange unfolds, and are shared between learners in fleeting and complex ways. Stimulated recall is useful in terms of the depth of insight which it can reveal for the researcher and, when used flexibly, it can allow them to collaborate with learners in ways which can indicate how exploratory forms of pedagogy might operate. It is possible to exploit stimulated recall to obtain learner-centred data which can then be triangulated with other data sources such as multimodal transcripts in order for the researcher to attempt to see the bigger research picture; and to answer questions from wider perspectives than perhaps would have been possible if reliant on a single method such as observation.

In the present study, I used stimulated recall in a very loose and flexible manner, for example, I wanted to find out how the tool could help me to understand the extent to which language learners might be supported to reflect on their multimodal communication. One difficulty in using stimulated recall as a method is that it is difficult for the researcher to ascertain whether learners are actually recalling their thoughts from the time of the task or whether they are simply reviewing the task in the present moment as they watch the video. The learners taking part in the current study used the video stimuli to discuss gestures from subjective vantage points with certain common themes discernible across the data. Use of stimulated recall appears to entail learners accessing thoughts from the past but also forming new connections, based around their interaction with the video and the researcher during the interview process itself. In hindsight, it would have been more productive to have spent twenty minutes before the interviews in order to train the learners to talk about their gestures. It became evident that learners became more proficient at recognising

features of their multimodal communication as the interviews progressed and sometimes fully grasped what was required of them too late within the proceedings.

6.4 Contribution of the study to research

The concept of multimodality has been used as an effective approach to guide research which aims to elucidate the manner in which second language teachers and learners interact across videoconferencing tools accessed on desktop technology. This previous research is extremely important in that it has illustrated that online teaching and learning involves considerably more than a focus on language use. This is because participants exploit multiple modes to interact, engage in processes considered beneficial to SLA; and build online social presence (see chapter 2, section 2.8). A contribution of the present study to research is to have instead examined multimodal forms of communication within MALL in order to illustrate the ways which language learners use gesture and speech to negotiate for meaning via Skype from beyond the classroom.

Learning beyond the classroom has seemingly ‘come of age’ and this study contributes to emergent research in MALL, and its parallel pedagogies (see chapter 2 section 2.9), in ways which suggest that L2 communicative exchanges can be supported in context-sensitive ways and from an array of everyday settings out in the world. At the time of writing, no research had been conducted into how language learners negotiate for meaning via mobile devices across Skype or from settings beyond the classroom. As a result, there was no information available for researchers or teachers interested in studying modes such as gesture and in understanding how language learners’ multimodal forms of communication might be enabled or constrained via mobile devices across Skype from an array of settings. Pachler et al. (2010) have posited that mobiles are particularly appealing from educational perspectives in that they provide affordances for meaning and for engagement with the world around us (p. 7). Their suggested affordances have included some of areas explored within the present study:

- Multimodality
- Portability
- Functionality
- Multimedia usage
- Ubiquity

- Social interactivity
- Context-sensitivity

There is also an increasing body of literature to guide research into an examination of gesture within L2 face-to-face teaching and learning contexts, based around the mode's contribution to processes considered beneficial to SLA. Conversely, there is a dearth of studies which have chosen to specifically examine gesture in relationship to language learners' speech across forms of technology-mediated communication. Sime (2008) notes that studies in the field of English language teaching in general have failed to ask learners themselves how they interpret gesture. This study placed learners at the centre of proceedings and gained a perspective on their multimodal experiences from beyond the classroom in order to help teachers to comprehend how learners might be supported to reflect on and learn from modes such as gesture in speech in exploratory forms of pedagogy. They were included as stakeholders in the research with their contribution seen as enormously valuable from a number of perspectives.

Due to the absence of comparable studies into the role of gesture via mobiles from beyond the classroom, a hybrid framework for an analysis of the video data needed to be developed. Another contribution to the research is that a reconceptualization of the negotiation of meaning framework by Varonis and Gass (1985) and integration with a gesture-speech unit of analysis and transcription scheme (McNeill, 1992; McNeill & Duncan, 2000) enables SLA to be perceived as a multimodal phenomenon. This is in terms of understanding how language learners attempt to bring together different semiotic systems in relationship to their use of technology and the settings around them. An interdisciplinary approach to multimodal research within SLA, moves away from a reliance on describing language learners' use of modes such as gesture as non-verbal, in that they are detached from language use, or are perceived as exclusively relevant to affective issues within language learning. This gap in the literature has left SLA researchers and language teachers, interested in understanding more about how gesture might contribute to aspects of language acquisition within online contexts, lacking information in terms of how they might go about analysing gesture seriously or working closely with learners to help them through their use of gesture. The approach which I have taken, to both the negotiation of meaning and the study of gesture within this, is therefore different to previous studies in SLA in face-to-face or online learning contexts. I have examined how negotiation operates multimodally from beyond the classroom and, in the

process, have challenged the conventional interpretation of this SLA framework as pertaining solely to language use. The study contributes to an understanding as to the multi-faceted roles of gesture within negotiation and appears to support the notion that language use is achieved multimodally and that gestures serve important communicative and speaker-internal functions. The socio-cognitive nature of gesture also challenges the continued narrow dichotomies in SLA which fail to draw on both social and psycholinguistic perspectives. Due to adopting an alternative research design and lens, I discovered that the negotiation of meaning does not need to operate in the decontextualised manner which the majority of previous studies have suggested.

In terms of information about language learners' use of gesture in a mobile learning context, the original transcription scheme designed to analyse gesture with the hands (see McNeill, 1992, 2000; McNeill & Duncan, 2000) was adapted to try to take account of how language learners point with an artificial device. There had been innovative research in the field of multimodality which had suggested that people form prosthetic deictic gestures through their use of technology such as cameras and video (Jaworski & Thurlow, 2011). However, at the time of writing this thesis, there was no transcription scheme or terminology which could depict how language learners point with forms of mobile technology when deployed to support language acquisition. To this end, the ability to conceptualise deictic gestures to include the manner in which learners point at objects within their use of Skype might be employed in research agendas which aim to understand how the use of mobile technologies can support aspects of acquisition in multimodal ways. Similarly, the notion that learners are seen to engage in processes which can benefit their language acquisition from spaces outside the classroom implies that tools such as Skype on mobile devices can be exploited by teachers in ways which are different to more formal classroom approaches but which, nevertheless, can be justified to institutions as being supportive of language learning.

6.5 Limitations of the study

The present study was small-scale in design and examined L2 interactions across a limited number of participants and settings. Studies involving more participants and conducted over more extended periods of time are required in order to more fully understand the use of multimodal forms of communication across mobile technologies beyond the classroom when underpinned by SLA theories. The task design was also exploratory in its inception

and application and only worked well for certain learners. I consider that task design for this type of mobile learning environment is in its infancy and would require greater levels of piloting and refinement to evaluate efficacy from both teachers' and learners' perspectives. It would be important to talk to learners themselves in order to hear their evaluation of the usefulness of certain task types implemented from beyond the classroom in relationship to others.

Language learners' use of gesture across Skype VC in a mobile learning environment is immediately differentiated from gesturing within face-to-face contexts. This is not just in terms of how the hands are mediated by mobile screens but also in terms of how pointing is achieved with a mobile device. There has been much valuable work conducted in the field of gesture studies and SLA within face-to-face scenarios, however, gestures between online learners are subjected to levels of distortion and transformation in ways which could only be initially explored within the present study. As the literature within face-to-face contexts suggests that gestures are such an important facet of L2 communication and acquisition, I argue for a parallel arena of developmental research which examines gesture and speech within online language learning contexts across a wide range of scenarios and digital technologies.

It is important to note that some of the coding decisions in the analysis need to be treated with a degree of caution. This is because the original framework (Varonis & Gass, 1985) in which they were conceived was not designed to accommodate the evolution in digital technologies. As a result, I consider that the coding categories could not fully take account of the types of interactions which the negotiation of meaning via mobiles from beyond the classroom entailed. An example of this limitation is evidenced in the levels of ambiguity around coding decisions where it was unclear to me, and my second coder, exactly how multimodal queries and forms of clarification, based on learners' use of the technology, should be conceptualised in relationship to theories of acquisition such as negotiation. Some examples of these are the visual comprehension checks and incidents of non-understanding where a problem with the sound quality appeared to be the reason why the negotiation of meaning had occurred in the first place. These are arguably new but poorly understood categories within the negotiation of meaning which would only pertain to the use of technology. An additional area for development would thus be to establish an extended series of more precise coding categories around the negotiation of meaning which is more suited to the analysis of Skype communication across a variety of devices. The

definition of codes such as triggers, indicators, and responses within qualitative data, collected from online scenarios, would benefit from levels of refinement in general through further fine-grained analysis to establish how entire communities of learners achieve these multimodally in relationship to aspects of acquisition.

It is also important to note that I was conducting the research as somewhat of an insider. For example, I had worked for many years at the school in which the research took place, having been a previous employee, and was very practised and familiar with the types of communicative language teaching which the learners would have been previously exposed to within their classroom lessons. Whilst I had not taught the learners who took part, I felt that they responded to me with levels of co-operation which might not have been so in evidence had I been a complete outsider or had not been experienced in working with learners' spoken forms of communication. It would be important to involve different types of teachers with different degrees of experience and methodological backgrounds in order to more comprehensively understand how gesture and speech might be exploited for the purposes of learning through the implementation of tasks and subsequent forms of reflection.

The use of qualitative methods in order to explore the role of gesture also entails learners exploiting a comprehensive arsenal of multimodal resources which includes their use of gaze, visuals, interaction with objects, proxemics, and the negotiation of spatial layout. Therefore, a focus on gesture and language use fails to fully address or explain learners' use of other modes within their interactions. Data from Skype also produces a mediated version of the use of modes such as gesture and generates data sets which are frequently difficult to deal with in terms of their levels of complexity. One considerable limitation to this type of research is that we cannot easily compare data to other examples of semiosis within SLA. Consequently, it might be difficult to argue that some of the findings would necessarily be generalisable beyond the small group of learners studied. This is because there are no other studies which have looked at gesture within the negotiation of meaning across mobile devices. However, in terms of gesturing with the hands, the findings confirm the role of gesture in ways which are generalisable to other learning contexts: relevance to SLA; use of iconics and deictics; a focus on lexis; multimodal forms of input and output; noticing; repair work; and benefits in learners enabled to talk about gesture and language use (see McCafferty & Stam, 2008 for an overview).

It is also important to note that the task type will have influenced and manipulated both the learners' negotiation of meaning and their use of the mobile device with which to point at the objects. Even if learners did not follow the task, most of them grasped the notion that they would be required to show and share objects across Skype VC. It is possible that a different approach to tasks, for example, asking learners to discuss a topic on Skype would not have resulted in the negotiation of meaning or the examples of prosthetic deictic gesturing within this study. However, a different approach to the task would have begged the question why learners had been asked to use Skype on mobile devices or why they had been taken out of the classroom for a task which could have been completed within the classroom setting.

6.6 Future research directions

For an overview of suggested future research directions in the field of SLA and multimodal communication within language learning with technology, see Table 11 (p. 239 this chapter). The methodology adopted for the purposes of this research illustrates one approach in which the negotiation of meaning was explored from a multimodal perspective which focussed on gesture in relationship to learners' spoken language within exchanges across Skype via mobiles deployed beyond the classroom. However, a multimodal interpretation of the Varonis and Gass framework and the use of a gesture-speech unit of analysis (McNeill, 1992; McNeill & Duncan, 2000) might be employed to underpin multimodal analysis of desktop as well as mobile forms of communication in order to answer a range of different questions about gesture in relationship to spoken language and use of technology. Future research agendas might focus on further refinement and categorisation of Varonis and Gass' coding scheme when gesture and speech, and possibly a range of other modes, are analysed across language learners' use of different technologies. I consider that the area of prosthetic deictic gesturing warrants further investigation due to the increasing interest in language learning with mobiles from beyond the classroom and the array of platforms which now enable real-time video communication across different devices. Research could involve the further development and refinement of transcription procedures designed to illustrate how pointing with devices, in relationship to sharing aspects of the world, might support forms of L2 talk from cognitive and social perspectives. Such a research agenda might take the form of qualitative studies which focus on the manner in which different settings impact on aspects of gesture and language use in order to enable a better understanding of acquisition from beyond the classroom for

learners and teachers. There are also related research areas in terms of establishing which task types are most valuable to learning when dyads are encouraged to interact from unconventional spaces and from multimodal perspectives. As this study has shown, learning beyond the classroom places learners in new and often unpredictable communicative scenarios. In order for researchers to elucidate the advantages and disadvantages of learning from beyond the classroom, it is vital to talk to participants and gain an understanding of their perspective on the experience.

Conceptualising theories such as the negotiation of meaning from wider perspectives also empowers language learners who may struggle with the continued emphasis on language skills in isolation within the language classroom. Neu (1990) previously found that in taking account of modes such as gesture, language learners are seen to represent more effective communicators; even if their language proficiency is challenged. The availability of different types of settings which language learners encounter beyond the classroom affords them an entire array of affordances to exploit in order to express themselves in ways which the conventional classroom cannot offer. It would be interesting to explore how pedagogic tasks implemented beyond the classroom could be integrated into an entire syllabus which valued opportunities to learn languages within, but also beyond, the classroom walls. There is also research scope to examine how teachers themselves exploit gestures within their online exchanges with learners. It would be productive to adopt research designs which employed methodological tools such as stimulated recall to find out more about teachers' interpretation of their own and their learners' gestures in relation to explanations and understanding of language items. This would be in order to develop structured forms of multimodal pedagogy and better levels of support for learners. Conversely, the idea of peer-led tasks or exchanges via Skype beyond the classroom might also be reflected in a different type of developmental research agenda. This would address the needs of adult language learners who have not thrived in more formal language learning scenarios or who simply cannot afford the financial cost of learning within these programmes.

Table 11: Future research directions in multimodality and SLA via technology

<p>Multimodal reconceptualisation of a range of different theories within SLA which also draw on gesture-speech theories in relationship to learners' use of mobile and desktop technologies:</p> <ul style="list-style-type: none"> • The socio-cognitive functions of gesture in relationship to SLA. • The illustrative functions of gesture in terms of learners' acquisition of specific items of vocabulary within task-based paradigms. • Multimodal perspective on self and other-initiated repair work through forms of stimulated reflection based on captured data from online tasks. • Gesture as a tool for mediation. 	<p>Areas for multimodal exploration include identification of iconic, metaphoric, deictic, prosthetic deictic, beat gestures but also modes such as gaze and proxemics in relationship to the following specific theories:</p> <ul style="list-style-type: none"> • The refinement of the categories of the negotiation of meaning (Varonis and Gass, 1985) as a multimodal phenomenon. • Long' Interaction Hypothesis (1996) as a multimodal phenomenon. • Schimdt's Noticing Hypothesis (1990) as a multimodal phenomenon. • Scaffolding and the Zone of Proximal Development (Vygotsky, 1986) as a multimodal phenomenon.
<p>Learning with mobiles from beyond the language classroom:</p> <ul style="list-style-type: none"> • Further exploration of the language learning affordances of spaces such as cafes, museums, train stations etc. These are exploited as authentic settings in order to support a variety of communicative task-based approaches. Tasks are engineered to avail of the affordances of the technology and designed to exploit the potential for learner opportunities based around contextualised language use. 	<p>A focus on the following:</p> <ul style="list-style-type: none"> • The manner in which learners exploit contextual information, derived from their surroundings. Focus on the role of prosthetic deictic gestures as a form of mutual contextualization to construct meaning and to build intersubjectivity for learning purposes. • How different gesture types contribute to task-based learning outcomes through forms of peer collaboration.

<p>The semantic use of L2 gesture in relationship to other key modes such as gaze.</p> <p>Longitudinal study across three months.</p>	<ul style="list-style-type: none"> • How learners and teachers interpret their own and one another's use of multimodal resources in task-based interactions and online conversations. • How teachers can best improve their online practice by supporting learners to comprehend the meaning of linguistic items through the use of holistic multimodal ensembles in, for example, their vocabulary explanations across ubiquitous platforms such as Skype and YouTube.
<p>Adult learners who have been excluded from accessing formal language courses in the UK for a number of reasons.</p> <p>Longitudinal study over six months.</p>	<ul style="list-style-type: none"> • How excluded adult language learners might benefit from peer-to-peer interactions which they instigate from different settings, for example, their homes or other local spaces via the connectivity of the technology and a shifting locus of control. To find out if/how they learn from their multimodal forms of communication. • In what ways are multimodal versions of SLA theories relevant to informal language learners.

Bibliography

ADAMS, R. (2007). Do language learners benefit from interacting with one another? In A. Mackey (ed.), *Conversational interaction in second language acquisition* (pp. 29-51). Oxford: Oxford University Press.

ALLEN, L.Q. (1995). The effect of emblematic gestures on the development and access of mental representations of French expressions. *Modern Language Journal*, 79 (4), 521-529.

Al-SHEHRI, S. (2011). Context in our pockets: Mobile phones and social networking as tools of contextualising language learning. Paper presented at MLEARN. 10TH World conference on mobile and contextual learning. Beijing, China. October 18-21. Retrieved from <https://scribd.com/doc/70516310/mLearn-2011-BeiJing-Conference-Proceedings> (accessed 7th April 2019)

AUSTIN, N., HAMPEL, R., & KUKULSKA-HULME, A. (2017). Videoconferencing and multimodal expression of voice: Children's conversations using Skype for second language development in a telecollaborative setting. *System*, 64, 87-103.

BACHMAN, L.S. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

BALDRY, A., & THIBAUT, P.J. (2006). *Multimodal transcription and text analysis*. London: Equinox publishing Ltd.

BALEGHIZADEH, S., & OLADROSTAM, E. (2010). The effect of mobile assisted language learning (MALL) on grammatical accuracy of EFL students. *MEXTESOL Journal*, 34 (2), 77-86.

BARALT, M. (2017). Coding qualitative data. In A. Mackey & S. Gass (eds.), *Research methods in second language acquisition* (pp. 222-244). Chichester: Wiley-Blackwell.

BEATTY, K. (2013). *Beyond the classroom: Mobile learning in the wider world*. Monterey, CA: The International Research Foundation for English Language Education. Retrieved from <http://www.tirfonline.org/english-in-the-workforce/mobile-assisted-language-learning/> (accessed 7th April 2019)

BELHIAH, H. (2013). Using the hand to choreograph instruction: On the functional role of gesture in definition talk. *The Modern Language Journal*. 97 (2), 417-434.

BENSON, P. (2011). An introduction to the field. In P. Benson, & H. Reinders (eds.), *Beyond the language classroom* (pp. 7-16). Basingstoke: Palgrave Macmillan.

BENSON, P., CHIK, A., XUESONG, G., HUANG, J., & WANG, W. (2009). Qualitative research in language teaching and learning journals, 1997-2006. *The Modern Language Journal* 93 (1), 79-90.

BERWICK, R.C., CHOMSKY, N., & PIATTELLI-PALMARINI, M. (2013). *Rich languages from poor inputs*. Oxford: Oxford University Press.

BIALYSTOK, E. (1990). *Communication strategies: A psychological analysis of second language use*. Oxford: Blackwell.

- BIRDWHISTELL, R.L. (1970). *Kinesics and context. Essays in body motion communication*. Philadelphia: University of Pennsylvania Press.
- BLAKE, R. (2000). Computer mediated communication: A window on L2 Spanish interlanguage. *Language Learning and Technology*, 4 (1), 111-125. Retrieved from https://scholarspace.manoa.hawaii.edu/bitstream/10125/25089/1/04_01_article4.pdf (accessed 7th April 2019)
- BLOCK, D. (2002). 'Mc communication:' A problem in the frame for SLA. In D., Block & D., Cameron (eds.), *Globalization and language teaching* (pp. 117-133). London: Routledge.
- BOWER, J., & KAWAGUCHI, S. (2011). The negotiation of meaning and corrective feedback in Japanese/English eTandem. *Language Learning and Technology*. 15 (1), 41-71. Retrieved from https://scholarspace.manoa.hawaii.edu/bitstream/10125/44237/1/15_01_bowerkawaguchi.pdf (accessed 7th April 2019)
- BRECKINGRIDGE CHURCH, R.B., AYMANN-NOLLEY, S., & MAHOOTIAN, S. (2004). The role of gesture in bilingual education: Does gesture enhance learning? *Bilingual Education and Bilingualism* 7 (4), 303-319.
- BROWN, E. (2010). Education in the wild: Contextual and location-based mobile learning in action. A report from STELLAR Alpine Rendez-vous workshop, edited by Elizabeth Brown. Nottingham, UK: University of Nottingham, Learning Sciences Research Institute (LSRI). Retrieved from: http://oro.open.ac.uk/29882/1/ARV_Education_in_the_wild.pdf (accessed 7th April 2019).
- BÜHLER, K. (1982). The deictic field of language and deictic words. In J. Jarvella, & W. Klein (eds.), *Speech, place and action* (pp. 9-30). Chichester: Wiley.
- BURSTON, J. (2015). Twenty years of MALL: A meta-analysis of learning outcomes. *ReCALL*, 27 (1), 4-20.
- BUTTERWORTH, B., & HADAR, U. (1989). Gesture, speech and computational stages: A reply to McNeill. *Psychological Review*. 96 (1),168-174.
- CANALE, M., & SWAIN, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1 (1), 1-47.
- CARELS, P.E. (1981). Pantomime in the foreign-language classroom. *Foreign Language Annals*, 14, 407-411.
- COHEN, L., MANION, L, & MORRISON, K. (2000). *Research methods in education*. London: Routledge.
- CHANIER, T., & LAMY, M. N. (2017). Researching technology-mediated multimodal interaction. In C. A. Chapelle, & S. Sauro, (eds.), *The handbook of technology and second language teaching and learning* (pp. 428-443). NJ: Wiley & Sons.
- CHAPELLE, C. (2009). The relationship between second language acquisition theory and computer-assisted-language-learning. *The Modern Language Journal*. 93 (S1), 741-753.

- CHOMSKY, N. (1965). *Aspects of the theory of syntax*. Cambridge, Mass: MIT press.
- CHURCHILL, E., OKADA, H., NISHINO, T., & ATKINSON, D. (2010). Symbiotic gesture and the sociocognitive visibility of grammar in second language acquisition. *Modern Language Journal*, 94 (2), 234-253.
- CLARK, H.H. (2003). Pointing and placing. In S. Kita (ed.), *In pointing: Where language, culture, and cognition meet* (pp. 243-268). Mahwah, NJ: Lawrence Erlbaum.
- COHEN, L., MANION, L, & MORRISON, K. (2000). *Research methods in education*. London: Routledge.
- COMAS-QUINN, A., MARDOMINGO, R., & VALENTINE, C. (2009). Mobile blogs in language learning: Making the most of informal and situated learning opportunities. *ReCALL*, 21 (1), 96-112.
- COPE, B., & KALANTZIZ, M. (2000). Introduction. Multiliteracies: The beginnings of an idea. In B. Cope & M. Kalantziz (eds.), *Multiliteracies: Literacy, learning and the design of social futures* (pp. 3-8). London: Routledge.
- CORBALLIS, M. C (2002). *From hand to mouth: The origins of language*. Princeton, NJ: Princeton University Press.
- CRESWELL, J.W. (2013). *Qualitative inquiry and research design* (3rd ed.). Thousand Oaks, CA: Sage.
- CUMMING, G. (1992). A perspective on learning for intelligent education systems. *Journal of Computer Assisted Learning*, 9 (2), 229-238.
- DAHL, T.I., & LUDVIGSEN, S. (2014). How I see what you're saying: The role of gestures in native and foreign language listening comprehension. *Modern Language Journal*, 98 (3), 813-833.
- DE LA FUENTE, M.J. (2014). Learners' attention to input during focus on form: The role of mobile technology in the second language classroom. *Computer Assisted Language Learning*, 27 (3), 261-276.
- DENZIN, N.K., & LINCOLN, Y.S. (2011). Introduction: The discipline and practice of qualitative research. In N.K. Denzin & Y.S. Lincoln, (eds.), *The Sage handbook of qualitative research* (4th Ed) (pp. 1-19). Thousand Oaks, CA: Sage.
- DE RUITER, J.P. (2000). The production of gesture and speech. In D. McNeill (ed). *Language and gesture* (pp. 284-311). Cambridge: Cambridge University Press.
- DICKS, B., FLEWITT, R., LANCASTER, L., & PAHL, K. (2011). Multimodality and ethnography: Working at the intersection. *Qualitative Research* 11 (3), 227-237.
- DÖRNYEI, Z. (2001). *Motivational strategies in the language classroom*. Cambridge: Cambridge University Press.
- DÖRNYEI, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.

- DOUGHTY, C.J., & LONG, M.H. (2003). The scope of inquiry and goals of SLA. In C.J. Doughty & M.H. Long (eds.), *The handbook of second language acquisition* (pp. 3-16). Malden, MA: Blackwell.
- EFRON, D. (1972). *Gesture, race, and culture* (reissue of Efron, 1941 as *Gestures and environment*. NY: King's Crown Press). The Hague: Mouton & Co.
- ECKMAN, P., & FRIESEN, W.V. (1969). The repertoire of non-verbal behaviour: Categories, origins, usage, and coding. *Semiotica*, 1 (1), 49-98.
- ELLIS, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
- ENGELKAMP, J., & KRUMNACKER, H. (1980). Image and motor-processes in the retention of verbal materials. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 27(4), 511-533.
- ERICSSON, K. & SIMON, H. (1996). *Protocol analysis: Verbal reports as data* (3rd Ed.). Cambridge: Cambridge University Press.
- ESKILDSEN, S.W., & WAGNER, J. (2015). Embodied L2 construction. *Language Learning*, 65 (2), 268-297.
- FAERCH, C., & KASPER, G. (1983). On identifying communication strategies in interlanguage production. In C. Faerch, & G. Kasper (eds.), *Strategies in interlanguage communication* (pp. 210-238). London: Longman.
- FARACO, M., & KIDA, T. (2008). Gesture and the negotiation of meaning in a second language classroom. In S.G. McCafferty, & G. Stam (eds.), *Gesture: Second language acquisition and classroom research* (pp. 280-297). London: Routledge.
- FARR, W., PRICE, S. & JEWITT, C. (2012). An introduction to embodiment and digital technology research: Interdisciplinary themes and perspectives. National Centre for Research Methods Working Paper. Retrieved from http://eprints.ncrm.ac.uk/2257/4/NCRM_workingpaper_0212.pdf / (accessed 7th April 2019).
- FELIX, S., & SIMMET, A. (1981). Natural processes in classroom L2 learning. Revised version of a paper presented at the IIIème Colloque Group de Recherche sur L'acquisition de Langues, Paris, May 15-17.
- FLEWITT, R. (2006). Using video to investigate pre-school classroom interaction: Education research assumptions and methodological practices. *Visual Communication* 5, (1), 25-50.
- FLEWITT, R., HAMPEL, R., HAUCK, M., & LANCASTER, L. (2014). What are multimodal data and transcription? In C. Jewitt, (ed.), *The Routledge handbook of multimodal analysis* (2nd Ed.) (44-59). London: Routledge.
- FOSTER, P. (1998). A classroom perspective on the negotiation of meaning. *Applied Linguistics*, 19 (1), 1-23.

- FOSTER, P., & OHTA, A. S. (2005). Negotiation for meaning and peer assistance in second language classrooms. *Applied Linguistics*, 26 (3), 402-430.
- FRICK-HORNBURY, D., & GUTTENTAG, R.E. (1998). The effects of restricting hand gesture production on lexical retrieval. *The American Journal of Psychology*, 3 (1), 43-62.
- GASS, S., & MACKEY, A. (2000). *Stimulated recall methodology in second language research* (1st Ed.) Mahwah, NJ: Lawrence Erlbaum Associates.
- GASS, S. & MACKEY, A. (2017). *Stimulated recall methodology in second language research* (2nd Ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- GEERTZ, C. (1973). Thick description: Toward an interpretative theory of culture. In C. Geertz (ed.), *The interpretation of cultures* (pp. 3-30). New York: Basic Books.
- GLASER, B.G., & STRAUSS, A.L. (2017). *The discovery of grounded theory: Strategies for qualitative research*. London: Routledge.
- GOFFMAN, E. (1959). *The presentation of self in everyday life*. Harmondsworth: Penguin.
- GOFFMAN, E. (1964). The neglected situation. *American Anthropologist*, 66 (6, part 2), 133-136.
- GOLDIN-MEADOW, S. (2014). Gesture in all its forms: following in the footsteps of Adam Kendon In M. Seyfeddinipur, & M. Gullberg (eds.), *From gesture in conversation to visible action as utterance* (pp. 289-308). Amsterdam. John Benjamins.
- GOODWIN, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32, 1489-1522.
- GOODWIN, C. (2003). *Conversation and brain damage*. Oxford: Oxford University Press.
- GOODWIN, C. (2007). Environmentally coupled gestures. In S. Duncan, J. Cassell, & E. Levy (eds.), *Gesture and the dynamic dimensions of language* (pp. 195-212). Amsterdam: John Benjamins.
- GOODWIN, C. (2014). The intelligibility of gesture within a framework of co-operative action In M. Seyfeddinipur, & M. Gullberg (eds.), *From gesture in conversation to visible action as utterance* (pp.199-216). Amsterdam: John Benjamins.
- GUICHON, N., & MC LORNAN, S. (2008). The effects of multimodality on L2 learners: Implications for CALL resource design. *System*, 36 (1), 85-93.
- GULLBERG, M. (1993). The role of co-verbal gestures in second language discourse: A case study. *Working Papers in Linguistics 40*: 41-60. Lund: Dept. of Linguistics. Lund University.
- GULLBERG, M. (1998). *Gesture as communicative strategy in second language discourse*. Lund: Lund University Press.

- GULLBERG, M. (2003). Gestures, referents, and anaphoric linkage in learner varieties. In C. Dimoth & M. Starren (eds.), *Information structure, linguistic structure and the dynamics of language acquisition* (pp. 311-328). Amsterdam: John Benjamins.
- GULLBERG, M. (2006). Some reasons for studying gesture and second language acquisition. *IRAL*, 44, 103-124.
- GULLBERG, M. (2010). Methodological reflections on gesture analysis in second language acquisition and bilingualism research. *Second Language Research*, 26 (1), 75-102.
- GULLBERG, M. (2011). Multilingual multimodality: Communicative difficulties and their solutions in second-language use. In J. Streeck, C. Goodwin, & LeBaron, C. (eds.), *Embodied interaction: Language and body in the material world* (137-151). Cambridge: Cambridge University Press.
- GULLBERG, M., DE BOT, K. & VOLTERRA, V. (2008). Gestures and some key issues within the study of language development. *Gesture*, 8 (2), 149-179.
- GUTH, S., & HELM, F. (2010). *Telecollaboration 2.0: Language, literacies and intercultural learning in the 21st Century*. Bern: Peter Lang.
- HAMMERSLEY, M. (1992). *What's wrong with ethnography? Methodological explorations*. London: Routledge.
- HAMMERSLEY, M. & ATKINSON, P. (2007). *Ethnography: Principles in practice* (3rd Ed.). London: Routledge.
- HAMPEL, R. (2003). Theoretical perspectives and new practices in audio-graphic conferencing for language learning. *ReCALL*, 15(1), 21-36.
- HAMPEL, R. (2006). Rethinking task design for the digital age: A framework for language teaching and learning in a synchronous online environment. *ReCALL*, 18(1), 105-121.
- HAMPEL, R. (2019). *Disruptive technologies and the language classroom: A complex systems theory approach*. Switzerland: Palgrave Macmillan.
- HAMPEL, R., & HAUCK, M. (2006). Computer-mediated language learning: Making meaning in multimodal virtual learning spaces, *The JALT CALL journal*, 2 (2), 3-18.
- HAMPEL, R., & STICKLER, U. (2012). The use of videoconferencing to support multimodal interaction in an online language classroom. *ReCALL*, 24(2), 116-137.
- HATCH, E. M. (1978). *Second language acquisition: A book of readings*. Rowley, Mass: Newbury House.
- HAVILAND, J. (2000). Pointing, gesture spaces, and mental maps In D. McNeill (ed.), *Language and gesture* (pp. 13-46) Cambridge: Cambridge University Press.
- HAYASHI, M. (2003). *Joint utterance construction in Japanese conversation*. Amsterdam: John Benjamins.

- HEATH, C., HINDMARSH, J., & LUFF, P. (2010). *Video in qualitative research: Analysing social interaction in everyday life*. London: Sage.
- HELLERMANN, S., THORNE, S.L., & FODOR, P. (2017). Mobile reading as social embodied practice. *Classroom Discourse*, 8 (2), 99-121.
- HENDERSON, L., HENDERSON, M., GRANT, S., & HUANG, H. (2010). What are users thinking in a virtual world lesson? Using stimulated recall reports to report student cognition and its triggers. *Virtual Worlds Research*, 3, 3-22.
- HOWATT, A.P.R. (2004). *A history of English language teaching* (2nd Ed). Oxford: Oxford University Press.
- HSIEH, H. F., & SHANNON, S.E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*. 15 (9), 1277-1288.
- HYMES, D. (1972). On communicative competence. In J.B. Pride, & J. Holmes, (eds.), *Sociolinguistics* (pp. 269-293). Harmondsworth: Penguin.
- JAWORSKI, A., & THURLOW, C. (2011). Gesture and movement in tourist spaces. In C. Jewitt (ed.), *The Routledge handbook of multimodal analysis* (1st Ed) (pp. 253-262). New York: Routledge.
- JEPSON, K. (2005). Conversations and negotiated interaction in text and voice chat rooms. *Language Learning and Technology*, 9 (3), 79-98. Retrieved from https://scholarspace.manoa.hawaii.edu/bitstream/10125/44033/1/09_03_jepson.pdf (accessed 7th April 2019)
- JEWITT, C. (2006). *Technology, literacy, learning: A multimodal approach*. London: Routledge.
- JEWITT, C. (2011). Introduction. In C. Jewitt (ed.), *The Routledge handbook of multimodal analysis* (1st Ed) (pp. 1-7). New York: Routledge.
- JEWITT, C. (2011). *The Routledge handbook of multimodal Analysis* (1st Ed.). New York: Routledge.
- JONES, R. (2015). C me SK8: discourse, technology and ‘bodies without organs.’ In C. Thurlow, & K. Mroczek (eds.), *Discourse: Language in the new media* (pp. 321-339). Oxford: Oxford University Press.
- KEATING, E., & SUNAKAWA, C. (2011). ‘A full inspiration tray:’ Multimodality across real and virtual spaces. In J. Streeck, C. Goodwin, & C. LeBaron (eds.), *Embodied interaction: Language and body in the material world* (pp. 194-204). Cambridge: Cambridge University Press.
- KELLERMAN, S. (1992). ‘I see what you mean:’ The role of kinesic behaviour in listening and implications for foreign and second language learning. *Applied Linguistics*, 13 (3), 239-258.

- KELLY, S.D., MCDEVITT, T., & ESCH, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language, *Language and Cognitive Processes*, 24 (2), 313-334.
- KELLY, S.D., ÖZYÜREK, A., & MARIS, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science*, 121 (2), 260-267.
- KELLY, S., BYRNE, K., & HOLLER, J. (2011). Raising the ante of communication evidence for enhanced gesture use in high stakes situations. *Information*, 2 (4), 579-593.
- KENDON, A. (1972). Some relationships between body motion and speech: An analysis of an example. In A. Seigman, & B. Pope (eds.), *Studies in dyadic communication* (pp. 177-210). New York: Pergamon Press.
- KENDON, A. (1980). How gestures can become like words. In F. Poyatos (ed.), *Cross-cultural perspectives in non-verbal communication* (pp. 131-141). New York: C.J. Hogrefe.
- KENDON, A. (1982). A study of gesture, some observations on its history. *Semiotic Inquiry*, 2, 45-62.
- KENDON, A. (2004). *Gesture: Visible action as utterance*. New York: Cambridge University Press.
- KENDON, A. (2011). Vocalization, speech, gesture, and the language origin debate. An essay review on recent contributions. *Gesture*, 11 (3), 349-370.
- KENNEDY, C. & LEVY, M. (2008). L'Italiano al telefono: Using SMS to support beginners' language learning. *ReCALL*, 20 (3), 315-330.
- KERN, R. (2014). Technology as pharmakon: The promise and perils of the internet for foreign language education. *The Modern Language Journal*, 98 (1), 340-357.
- KIDA, T. (2008). Does gesture aid discourse comprehension in the L2? In S.G. McCafferty, & G. Stam (eds.), *Gesture: Second language acquisition and classroom research* (pp. 131-156). New York: Routledge.
- KITA, S. (2000). How representational gestures can help speaking. In D. McNeill, (ed.), *Language and gesture* (pp. 162-185). Cambridge: Cambridge University Press.
- KNIGHT, P.T. (2002). *Small-scale research: Pragmatic inquiry in social science and the caring professions*. London: Sage.
- KRASHEN, S. (1980). The input hypothesis. In J. Atlas (ed.), *Current issues in bilingual education* (pp. 175-183). Washington DC: Georgetown University Press.
- KRASHEN, S. (1985). *The input hypothesis: Issues and implications*. New York: Longman.

- KRAUSS, R.M., CHEN, Y. & GOTTESMAN, R.F. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (ed.), *Language and gesture* (pp. 261-283). Cambridge: Cambridge University Press.
- KRESS, G. R. & VAN LEEUWEN, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. London: Edward Arnold.
- KRESS, G. (2003). *Literacy in the new media age*. London: Routledge.
- KRESS, G. R. & VAN LEEUWEN, T. (2006). *Reading images: The grammar of visual design* (2nd Ed.), London: Routledge.
- KRESS, G. R., & PACHLER, N. (2007). Talking about the ‘M’ in learning. In N. Pachler (ed.), *Mobile learning towards a research agenda*. WLE Centre. Occasional papers in work-based learning, 1, 7-32.
- KUKULSKA-HULME, A., & TRAXLER, J. (2005). Mobile teaching and learning. In A. Kukulska-Hulme & J. Traxler (eds.), *Mobile learning: A handbook for educators and trainers* (pp. 25-44). London: Routledge.
- KUKULSKA-HULME A., & SHIELD, L. (2008). An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. *ReCALL*, 20 (3): 271-289.
- KUKULSKA-HULME, A. (2012). Mobile Assisted Language Learning. In C.A. Chapelle (ed.), *The Encyclopedia of Applied Linguistics*. Retrieved from <https://onlinelibrary-wiley-com.libezproxy.open.ac.uk/doi/pdf/10.1002/9781405198431.wbeal0768> (accessed 7th April 2019).
- KUKULSKA-HULME, A. (2013). Re-skilling language learners for a mobile world. Monterey, CA: The International Research Foundation for English Language Education. Retrieved from: <http://www.tirfonline.org/english-in-the-workforce/mobile-assisted-language-learning/> (accessed 7th April 2019).
- KUKULSKA-HULME, A., LEE, H., & NORRIS, L. (2017). Mobile learning revolution: Implications for language pedagogy. In C. Chapelle, & S. Sauro (eds.), *The handbook of technology and second language teaching and language learning* (pp. 219-233). NJ: Wiley and Sons.
- KUKULSKA-HULME, A., & LEE, H. (2020) Mobile collaboration for language learning and cultural learning. In M. Dressman & R. Sandler (eds.), *The handbook of informal language learning*. (pp. 169-180) NJ: Wiley & Sons.
- KVALE, S. (1996). *Interviews*. London: Sage.
- LAI, C., & ZHENG, D. (2018). Self-directed use of mobile devices for language learning beyond the classroom. *ReCALL*, 30 (3), 299-318.
- LAMY, M.N., & FLEWITT, R. (2011). Describing online conversations: Insights from a multimodal approach. In C. Develotte, R. Kern, & M.R. Lamy (eds.), *Décrire la conversation en Ligne: Le face à face distanciel* (pp. 71-94). Lyon, France: ENS Éditions.

- LAN, Y., SUNG, Y., & CHANG, K. (2009). Let us read together: Development and evaluation of a computer-assisted reciprocal English reading system. *Computers & Education*, 53 (4), 1188-1198
- LANCASTER, L. (2012). Moving into literacy: How it all begins. In J. Larson & J. Marsh (eds.), *The Sage handbook of early childhood literacy* (pp. 313 – 328) (2nd Ed.) London: Sage.
- LANCASTER, L. (2013). Multimodal literacy. In C. A. Chapelle (ed), *The Encyclopedia of Applied Linguistics*. Retrieved from <https://onlinelibrary-wiley.com.libezproxy.open.ac.uk/doi/pdf/10.1002/9781405198431.wbeal1354> (accessed 7th April 2019)
- LANTOLF, J.P. (2000). Introducing sociocultural theory. In J.P. Lantolf (ed.), *Sociocultural theory and second language learning* (pp. 1-26). Oxford: Oxford University Press.
- LANTOLF, J.P., & THORNE, S.L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford: Oxford University Press.
- LAVE, J. and WENGER, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.
- LAZARATON, A. (2004). Gesture and speech in the vocabulary explanations of one ESL teacher: A microanalytic inquiry. *Language Learning*, 54, 79-117.
- LEA, M., and LILLIS, T. (2016). Reflections 5. Revisiting the question of transformation in academic literacies: The ethnographic imperative. In T. Lillis, K. Harrington, M., Lea & S. Mitchell (eds.), *Working with academic literacies: Case studies towards transformative practice* (pp. 383-390). Colorado: The WAC Clearinghouse.
- LEBARON, C., & STREECK, J. (2000). Gestures, knowledge, and the world. In D. McNeill (ed.), *Language and gesture* (pp. 118-138). Cambridge: Cambridge University Press.
- LECOMPTE, M., & SCHENSUL, J. (1999). *Analysing and interpreting ethnographic data*. Walnut Creek: Alta Mira Press.
- LEE, H., HAMPEL, R., & KUKULSKA-HULME, A. (2019). Gesture in speaking tasks beyond the classroom: An exploration of the multimodal negotiation of meaning via Skype videoconferencing on mobile devices, *System* 81, 26-38.
<https://doi.org/10.1016/j.system.2018.12.013>
- LEE, I. (2001). Online interaction: negotiation of meaning and strategies used among learners of Spanish. *ReCALL*, 13 (2), 232-244.
- LEE, J. (2008). Gesture and private speech in second language acquisition. *Studies in Second Language Acquisition*, 30 (2), 169-190.
- LEVY, M., & STOCKWELL, G. (2006). *CALL dimensions: Options and issues in computer-assisted-language-learning*. New York: Routledge.

- LILLIS, T. (2008). Ethnography as method, methodology, and “Deep Theorizing:” Closing the gap between text and context in Academic writing research. *Written Communication*, 25 (3), 353-388.
- LINCOLN, Y.S., & GUBA, E.G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.
- LONG, M. H. (1996). The role of the linguistic environment in second language acquisition. In W.C. Ritchie, & T.K. Bhatia (eds.), *Handbook of second language acquisition* (pp. 413-468). San Diego, CA: Academic Press.
- LOTHERINGTON, H., & JENSON, J. (2011). Teaching multimodal and digital literacy in L2 settings: New literacies, new basics, new pedagogies. *Annual Review of Applied Linguistics* 31, 226-246.
- MA, Q. (2017). Technologies for teaching and learning L2 vocabulary. In C.A. Chapelle, & S. Sauro, (eds.), *The handbook of technology and second language teaching and learning* (45-61). NJ: Wiley & Sons.
- MACEDONIA, M., & VON KRIEGSTEIN, K. (2012). Gestures enhance foreign language learning. *Biolinguistics*, 6 (3-4), 393-416.
- MACKEY, A. (2002). Beyond production: Learners’ perceptions about interactional processes. *Journal of Educational Research*, 3 (4), 379-394
- MACKEY, A. (2007). *Conversational interaction in second language acquisition*. Oxford: Oxford University Press.
- MACKEY, A., & GASS, S.M. (2012). *Research methods in second language acquisition: A practical guide*. Chichester: Wiley Blackwell.
- MAVERS, D. (2012). Transcribing video. National Centre for Research Methods Working Paper. Retrieved from http://eprints.ncrm.ac.uk/2877/4/NCRM_working_paper0512.pdf (accessed April 2019)
- MAXWELL, J.A. (1992). Understanding and validity in qualitative research. *Harvard Educational Review*, 62 (3), 279-301.
- MCCAFFERTY, S.G. (1998). Nonverbal Expression and L2 Private Speech. *Applied Linguistics*, 19 (1), 73–96.
- 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.
- MCCAFFERTY, S.G. (2008). Material foundations for second language learning: Gesture, metaphor; and internalization. In S.G., McCafferty & G. Stam (eds.), *Gesture: Second language acquisition and classroom research* (pp. 47-65). New York: Routledge.

- MCCAFFERTY, S.G. & AHMED, M.K. (2000). The appropriation of gestures of the abstract by L2 learners. In J.P. Lantolf (ed.), *Sociocultural theory and second language learning* (pp.199-218). Oxford: Oxford University Press.
- MCCAFFERTY, S.G., & STAM, G. (2008). *Gesture: Second language acquisition and classroom research*. New York: Routledge.
- MCNEILL, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: The University of Chicago Press.
- MCNEILL, D. (2000). *Language and gesture*. Cambridge: Cambridge University Press.
- MCNEILL, D. (2005). *Gesture and thought*. Chicago: The University of Chicago Press.
- MCNEILL, D. (2006). Gesture: A psycholinguistic approach (article draft for psycholinguistics section of The encyclopedia of language and linguistics). Retrieved from http://mcneilllab.uchicago.edu/pdfs/gesture.a_psycholinguistic_approach.cambridge.encyclopdf. (accessed 7th April 2019)
- MCNEILL, D., & DUNCAN, S. D. (2000). Growth points for thinking-for-speaking. In D. McNeill (ed.), *Language and gesture* (141-161). Cambridge: Cambridge University Press.
- MCNESS, E., ARTHUR, L., & CROSSLEY, M. (2015). 'Ethnographic dazzle' and the construction of the 'other:' Revisiting dimensions of insider and outsider research for international and comparative education. *Compare: A Journal of Comparative and International Education*, 45 (2), 295-316.
- MILES, M.B. & HUBERMAN, A.M. (1994). *Qualitative data analysis*. Thousand Oaks: Sage.
- MONDADA, L. (2014). Pointing, talk, and the bodies: Reference and joint attention as embodied interactional achievements. In M. Seyfeddinipur, & M. Gullberg (eds.), *From gesture in conversation to visible action as utterance* (pp. 95-124). Amsterdam: John Benjamins.
- MORI, J. & HAYASHI, M. (2006). The achievement of intersubjectivity through embodied completions: A study of interactions between first and second language speakers. *Applied Linguistics*, 27 (2), 195-219.
- MORTENSEN, K. (2016). The body as a resource for other-initiation-repair: Cupping the hand behind the ear. *Research on Language and Social Interaction*, 49 (1), 34-57.
- MORTENSEN, K. (2012). Conversation analysis and multimodality. *The Encyclopedia of Applied Linguistics*. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781405198431.wbeal0212> (accessed April 7th 2019)
- NAH, K.C., WHITE, P., & SUSSEX, R. (2008). The potential of using a mobile phone to access the internet for learning EFL listening skills within a Korean context. *ReCALL*, 20 (3), 331-347.
- NEGUERUELA, E., LANTOLF, J.P., REHN JORDAN, S., & GELABERT, J. (2004). The 'private function' of gesture in second language speaking activity: A study of motion

verbs and gesturing in English and Spanish. *International Journal of Applied Linguistics*, 14 (1), 113-147.

NEGUERUELA, E., & LANTOLF, J.P (2008). The dialectics of gesture in the construction of meaning in second language oral narratives. In S.G. McCafferty, & G. Stam (eds.), *Gesture: Second language acquisition and classroom research* (pp. 88-106). New York: Routledge.

NEU, J. (1990). Assessing the role of non-verbal communication in the acquisition of communicative competence in L2. In R. Scarcella, E.S. Andersen, & S.D. Krashen (eds.), *Developing communicative competence in a second language* (pp. 121-138). New York: Newbury House.

NORRIS, S. (2004). *Analyzing multimodal interaction: A methodological framework*. New York: Routledge.

NORRIS, S. (2011). Three hierarchical positions of deictic gesture in relation to spoken language: A multimodal interaction analysis. *Visual Communication*, 10 (2), 129-147.

NUNAN, D., & RICHARDS, J.C. (2015). *Language learning beyond the classroom*. New York: Routledge.

OGATA, H., YIN, C., PAREDES, R.G., SAITO, N.A., & YANO, Y. (2006). Supporting mobile language learning outside classrooms. Paper presented at Advanced learning technologies IEEE international conference. Kerkrade, The Netherlands. July 5 – 7. Retrieved from <https://www.computer.org/csdl/proceedings/icalt/2006/2632/00/263200928-abs.html> (accessed 7th April 2019)

O'HALLORAN, K.L., & SMITH, B.A. (2011). Multimodal studies. In K.L. O'Halloran & B.A. Smith (eds.), *Multimodal studies: Exploring issues and domains* (pp. 1-13). New York: Routledge.

OLSHER, D. (2008). Gesturally-enhanced repeats in the repair turn: Communication strategy or cognitive language learning tool? In S. McCafferty, & G. Stam (eds). *Gesture: Second language acquisition and classroom research* (pp. 109-130) New York: Routledge.

ÖZYÜREK, A. (2014). Hearing and seeing meaning in speech and gesture: Insights from brain and behaviour. *Philosophical Transactions of the Royal Society*. Retrieved from <http://rstb.royalsocietypublishing.org/content/royptb/369/1651/20130296.full.pdf> (accessed 7th April 2019)

PACHLER, N. (2009). Research methods in mobile and informal learning: Some issues. In G. Vavoula, N. Pachler, & A. Kukulska-Hulme (eds.), *Researching mobile learning: Frameworks, tools and research designs* (1-15). Oxford: Peter Lang.

PACHLER, N., BACHMAIR, B., & COOK, J. (2010). *Mobile Learning: Structures, agency, practices*. London: Springer.

PAIVIO, A. (2014). *Mind and its evolution: A dual coding theoretical approach*. New York: Psychology Press.

- PEGRUM, M. (2014). *Mobile learning: Languages, literacies and cultures*. Basingstoke: Palgrave Macmillan.
- PICA, T., KANAGY, R., & FALODUN, J. (1993). Choosing and using communication tasks for second language research and instruction. In G. Crookes, & S. Gass (eds.), *Tasks and second language learning: Integrating theory and practice* (pp. 9-34). Clevedon, Avon: Multilingual Matters.
- PICA, T. (1994) Research on negotiation: What does it reveal about second language learning conditions, processes, and outcomes? *Language Learning*, 44 (3), 493-527.
- PLATT, E., & BROOKS, F. B. (2008). Embodiment as self-regulation in L2 task performance. In S.G. McCafferty, & G. Stam (eds.), *Gesture: Second language acquisition and classroom research* (pp. 66-87). New York: Routledge.
- QUINTILIANUS, M. F. (1922). *The institutio oratoria of Quintilian*. The Loeb Classical Library. New York: G.P. Putnam and Sons (for section on gesture see volume IV, Book XI, 111, 61-149).
- RICHARDS, K. (2003). *Qualitative research in TESOL*. Basingstoke: Palgrave Macmillan.
- ROBSON, C. (2011). *Real world research* (3rd Ed.). NJ: Wiley & Sons.
- ROMMETVEIT, R. (1974). *On message structure: A framework for the study of language and communication*. London: John Wiley & Sons.
- ROS I SOLÉ, C., CALIC, J., & NEIJMANN, D.N. (2010). A social and self-reflective approach to MALL. *ReCALL*, 22, (1), 39-52.
- ROYCE, T.D. (2013). Multimodal communicative competence in second language contexts. In T.D Royce & W. Bowcher (eds.), *New directions in the analysis of multimodal discourse* (pp. 361-390). London: Routledge.
- SALDAÑO, J. (2013). *The coding manual for qualitative researchers*. London: Sage.
- SATAR, M.H. (2015). Sustaining multimodal language learner interactions online. *CALICO*, 32 (3), 480-507.
- SEYFEDDINIPUR, M., & GULLBERG, M. (2014). *From gesture in conversation to visible action as utterance*. Amsterdam: John Benjamins.
- SCHEGLOFF, E.A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge: Cambridge University Press.
- SCHIMDT, R.W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11 (2), 129-154
- SCHREIER, M. (2013). Qualitative content analysis. In U. Flick (ed.), *The Sage handbook of qualitative data analysis* (pp. 170-183). London: Sage.

- SCOLLON, R., & SCOLLON, S.W. (2003). *Discourses in place: Language in the material world*. New York: Routledge.
- SCOLLON, R., & SCOLLON, S.W. (2011). Multimodality and language: A retrospective and prospective view. In C. Jewitt (ed.), *The Routledge handbook of multimodal analysis* (1st Ed) (pp. 170-180). New York: Routledge.
- SHARPLES, M. (2009). Methods for evaluating mobile learning. In G. Vavoula, N. Pachler, & A. Kukulska-Hulme (eds.), *Researching mobile learning: Frameworks, tools and research designs* (pp.17-39). Oxford: Peter Lang.
- SILVER, C., & LEWINS, A. (2014). *Using software in qualitative research: A step-by-step guide* (2nd Ed.). London: Sage.
- SIME, D. (2008). Gesture and the negotiation of meaning in a second language classroom In S.G. McCaffery, & G. Stam (eds.), *Gesture: Second language acquisition and classroom research* (pp. 280-297). New York: Routledge.
- SMITH, B. (2003). Computer-mediated negotiation: An expanded model. *The Modern Language Journal*, 87 (1), 38-57.
- SMITH, B. (2005). The relationship between negotiated interaction, learner uptake, and lexical acquisition in task-based computer-mediated communication. *TESOL Quarterly*, 39 (1), 33-58.
- SKEHAN, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.
- SLOBIN, D.I. (1996). From 'thought and language' to 'thinking for speaking.' In J.J. Gumperz, & S.C. Levinson (eds.), *Rethinking linguistic relativity* (pp. 70-96). Cambridge: Cambridge University Press.
- SONG, Y. (2008). SMS enhanced vocabulary learning for mobile audiences. *International Journal of Mobile Learning and Organisation*. 21 (1), 81-98.
- SQUIRE, K.D. (2009). Mobile media learning: Multiplicities of place. *On the Horizon*, 17 (1), 70-80.
- STAM, G. (2008). What gestures reveal about second language acquisition. In S.G. McCafferty, & G. Stam (eds.), *Gesture: Second language acquisition and classroom research* (pp. 231-255). New York: Routledge.
- STICKLER, U., & HAMPEL, R. (2015). Transforming teaching: New skills for online language learning spaces. In R. Hampel & U. Stickler (eds.), *Developing online language teaching: Research-based pedagogies and reflective practices* (pp. 63-77). Basingstoke: Palgrave Macmillan.
- STOCKWELL, G. (2007). Vocabulary on the move: Investigating an intelligent mobile phone-based vocabulary tutor. *Computer Assisted Language Learning*, 20 (1), 365-383.

- STOCKWELL, G. (2016). Mobile language learning. In F. Farr, & L. Murray (eds.), *The Routledge handbook of language learning and technology* (296-307). New York: Routledge.
- STOCKWELL, G., & HUBBARD, P. (2013). Some emerging principles for mobile-assisted language learning. Monterey, CA: The International Research Foundation for English Language Education. Retrieved from <http://www.tirfonline.org/english-in-the-workforce/mobile-assisted-language-learning> (accessed 7th April 2019)
- STOKOE, W.C. (1980). Sign language structure. *Annual review of Anthropology*, 9, 365-390.
- STREECK, J., GOODWIN, C., & LEBARON, (2011). *Embodied interaction: Language and body in the material world*. Cambridge: Cambridge University Press.
- SUEYOSHI, A., & HARDISON, D.M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, 55, 661-699.
- SWAIN, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S.M. Gass, & C.G. Madden (eds.), *Input in second language acquisition* (pp.235-253). Rowley, Mass: Newbury House Publishers.
- SWAIN, M. (1995). Three functions of output in second language learning. In G. Cook, & B. Seidlhofer, *Principle and practice in applied linguistics: Studies in honour of H.G. Widowson* (pp. 125-144). Oxford: Oxford University Press.
- TARONE, E. (1980). Communication strategies, foreigner talk, and repair in interlanguage. *Language Learning*, 30 (2), 417-428.
- TEN HAVE, P. (2007). *Doing conversation analysis: A practical guide*. London: Sage.
- THOMAS, J. (1988). The role played by metalinguistic awareness in second and third language learning. *Journal of Multilingual and Multicultural Development*, 9 (3), 235-246.
- THOMAS, M., & REINDERS, H. (2010). *Task-based language learning and teaching with technology*. London: Continuum.
- TRAXLER, J. (2013). Mobile learning for languages: Can the past speak to the future? Monterey, CA: The International Research Foundation for English Language Education. Retrieved from <http://www.tirfonline.org/english-in-the-workforce/mobile-assisted-language-learning/> (accessed 7th April 2019)
- TUDINI, V. (2012). *Online second language acquisition: Conversation analysis of online chat*. London: Continuum.
- VAN DER ZWAARD, R. & BANNINK, A. (2014). Video call or chat? Negotiation of meaning and issues of face in telecollaboration. *System* 44, 137-148.
- VAN LIER, L. (2004). The semiotics and ecology of language learning: Perception, voice, identity and democracy, *Utbildning & Demokrati*, 13 (3), 79-103.

- VARONIS, E.M., & GASS, S.M. (1985). Non-native/non-native conversations: A model for negotiation of meaning. *Applied Linguistics*, 6, (1), 71-90.
- VAVOULA, G., PACHLER, N., & KUKULSKA-HULME, A. (2009). *Researching mobile learning: Framework, tools and research designs*. Oxford: Peter Lang.
- VYGOTSKY, L. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- VYGOTSKY, L. (1987). *The collected works of L.S. Vygotsky. Vol. 1: Problems of general psychology*. (Including the volume Thinking and Speech) New York: Plenum Press.
- WANG, Y. (2006). Negotiation of meaning in desktop videoconferencing supported distance learning. *ReCALL*, 18 (1), 122-146.
- WARE, P.D., & O'DOWD, R. (2008). Peer feedback on language form in telecollaboration, *Language Learning and Technology*, 12 (1), 43-63. Retrieved from https://scholarspace.manoa.hawaii.edu/bitstream/10125/44130/1/12_01_wareodowd.pdf (accessed 7th April 2019)
- WEBER, S. (2008). Visual images in research. In J.G. Knowles & A.L. Cole (eds.), *Handbook of the arts in qualitative research* (pp. 41-53). London: Sage.
- WERTSCH, J.V. (1991). *Voices of the mind*. Cambridge, MA: Harvard University Press.
- WIDDOWSON, H.G. (1990). *Aspects of language teaching*. Oxford: Oxford University Press.
- WONG, L.H., CHEN, W., & JAN, M. (2012). How artefacts mediate small-group co-creation activities in a mobile-assisted seamless language learning environment. *Journal of Computer Assisted Learning*, 28, 5, 411-424.
- WYLIE, L. (1985). Language learning and communication. *The French Review*, LVIII, 6: 777-785.
- WU, T., SUNG, T., HUANG, Y., YANG, C., & YANG, J.-C. (2011). Ubiquitous English learning system with dynamic personalized guidance of learning portfolio. *Educational Technology & Society*, 14, 4, 164-180.
- WUNDT, W. (1921/1973). The language of gestures (Thayer, J. S. Greenlauf, C. M., Silberman, M. D., Trans.). (From *Völkerpsychologie: eine Untersuchung der Entwicklungsgesetze von Sprache, Mythos und Sitte*. Stuttgart: Alfred Kröner Verlag, 1921. Vol.1, First part, Chap.2.). The Hague: Mouton.
- YANGUAS, I. (2010). Oral computer-mediated interaction between L2 learners: It's about time. *Language Learning and Technology*, 14 (3), 72-93. Retrieved from https://scholarspace.manoa.hawaii.edu/bitstream/10125/44227/1/14_03_yanguas.pdf (accessed 7th April 2019)

APPENDIX 1 Letter about the research and meeting

Hello,

My name is Helen Lee and I am a PhD student at The Open University, UK. I am researching the use of mobile devices in language learning tasks outside the classroom in places such as restaurants and museums, located near your school. My project is based on understanding Skype video communication from mobile devices and identifying features of students' communication when they interact in a second language from outside the classroom.

In this project, you'll have an opportunity to practise your speaking and listening skills while you have a coffee! You will be able to connect and talk with another learner to explore some places together by using a mobile device (phones, tablets, tablet/laptop) and Skype video in order to complete a speaking task. You will also be interviewed and have an opportunity to talk with me about your communication following the speaking task.

Your participation would be greatly appreciated. It will take around two hours of your time after your school day at a time which is convenient for you. I will record the speaking task on video and video record your interview as well. You can stop participating in the project at any time.

If you are interested in taking part in this project, please see the director of the school in her office and she will put your name on a list. There will be a meeting held after school on **18th January at 4.00pm in room 8** if you would like to learn more and to ask me any questions which you might have. Attending this meeting does not mean that you have to take part in the project; and you can withdraw at any point during the project.

I can be contacted by email at the Open University on email: or you can phone or text me if you have any questions:

Thank you and I look forward to meeting you

APPENDIX 2 Participant consent form

Agreement form to participate in project: Language learning via Skype video on mobiles from outside the classroom.

I

(Print name)

agree to take part in this project.

Agreement to participate in this project

- I have had the purpose of this research explained to me in detail
- I understand that I will be leaving the school building and travelling to another nearby location with Helen (the researcher)
- I understand that I can refuse to participate or to stop participating in this research at any point
- I understand that my anonymity and confidentiality will always be protected
- I understand that screenshots (images) of my face and body will be used for the purpose of educational research and publishing
- I agree that the research can be used for educational research and publishing

I

(print name)

assign the copyright for my contribution for use in educational research and for publishing

Helen can be contacted at or via mobile (redacted) If you have any problems or complaints about the research you can also contact my supervisor (redacted)

Learner A:

(Image redacted)

- **You are going to help your partner to locate this picture from her location (this is a small section of a much larger picture on her wall)**
- Help your partner to find the picture by answering her questions.
- Share and discuss the picture together. Here are some possible questions for you to ask your partner when they have located the picture.
- What is happening in the picture?
- What is the mode of transport?
- Which city does the picture depict?
- Have you ever visited this city and country?
- Would you like to visit and why?
- **Ask your partner to explain her menu and some of the dishes that she would like to order**

(Image redacted)

- Describe the dish above and find out if your partner has ordered it.
- Discuss the food on the table together. Here are some possible questions to ask your partner.
 - a) What have you ordered from the menu and why?
 - b) What is your favourite dish and why?
 - c) How do you make dish?
 - d) What does it taste like?

Now tell your partners about your own country and its traditional dishes.

Learner B

- Your partner is going to help you to find a picture which is positioned in your location. You don't know which picture it is, so you will have to ask her some questions to find out. Here are some possible questions to ask your partner:
 - a) Is it a photo or a painting?
 - b) Who is in the picture?
 - c) Where is the picture located?
- Now try to see if you can find the right picture based on your partner's description.
- When you have found the correct picture, share this and discuss it with your partner. She will also ask you some more questions about the picture which you can try to answer.
- Next, you are going to help your partner to find the painting shown below. Describe this small section of the painting and then see if your partner can find the painting. (if she needs help, it is hanging in the room on the ground floor to the right of the main entrance)
(Image redacted)
 - a) Who is the painter?
 - b) What's the name of the painting?
 - c) When was it painted?

Share and discuss the whole painting. Do you like the painting and why?

(Image redacted)

- What do you think these strange objects in the picture above are? Ask your partner to find them in her location and to give you some more information about them. When and where were they made?
(Image redacted)
- Help your partner to find and then climb these stairs in her location. The stairs contain the lyrics to a famous song. See if your partner can find the stairs and tell you what the rest of the lyrics are.

(Image redacted)

- Describe the painting above and see if your partner can find it in their location

Appendix 4: Interview guide


- 1) Can you tell me about communicating on Skype when you have a tablet/
small laptop?
- 2) Can you tell me about communicating on Skype outside the classroom?
- 3) What was happening there?
- 4) Can you explain what the problem was at the time?
- 6) Can I ask what you were gesturing here?
- 7) What did you mean by that gesture?
- 8) Can you remember what you said?
- 9) Can you remember the language you used?
- 10) Was that gesture for you or your partner?
- 11) Can you remember what your partner was doing with their hands?
- 12) Did you see your partner's gesture at the time?
- 13) Can you remember what they said?
- 14) Can you remember what language they used?
- 15) Do you remember whether or not you responded?
- 16) Could you tell me what you were doing with the mobile device?
- 17) Do you remember having any problems with the mobile device here?

APPENDIX 5: Transcription key for gesture and speech via Skype VC

Gesture with the hands via Skype VC

[]	Encloses the gesture phrase, indicating a distinguishable beginning and ending to the learner's movement (preparation and retraction phases)
Bold	Bold typeface in the speech mode indicates the point where the gestural stroke formed with the hands co-occurs with speech

Prosthetic deictic gesture with the mobile device via Skype

[]	Encloses the gestural phrase, indicating a roughly distinguishable beginning and ending to the learner's movement (preparation and retraction phases)
Bold	Bold typeface in the speech mode indicates the point the stroke formed with the device and camera synchronises with the speech mode
	Indicates an example of negotiation where the learner's deployment of mobile device and Skype VC is used to construct a 'prosthetic deictic'