

Open Research Online

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

Negotiation For Meaning In Audio And Video Synchronous Computer-mediated Communication

Thesis

How to cite:

Li, Chenxi (Cecilia) (2021). Negotiation For Meaning In Audio And Video Synchronous Computer-mediated Communication. PhD thesis The Open University.

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

© 2020 Li Chenxi



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Version of Record

Link(s) to article on publisher's website:
<http://dx.doi.org/doi:10.21954/ou.ro.000133c8>

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



NEGOTIATION FOR MEANING IN AUDIO AND VIDEO SYNCHRONOUS COMPUTER- MEDIATED COMMUNICATION

Chenxi (Cecilia) Li

**Thesis Submitted to the Open University for the
Degree of Doctor of Philosophy**

Faculty of Wellbeing, Education and Language Studies,

The Open University, UK

Supervised by Dr. Timothy Lewis, Dr. Qian Kan and Dr. Lijing Shi

July, 2020

Abstract

After the outbreak of Covid-19 across the world, video conferencing tools have been widely used for online teaching all over the world. In synchronous computer-mediated communication (SCMC), where text, audio, video, and many other semiotic resources are simultaneously available, a huge challenge for interlocutors is how they can make good use of modes for their communication. Different modes of communication afforded by different types of technology can affect the way people communicate, and thus affecting the process of second language acquisition. Therefore, this study focuses on meaning negotiation episodes and aims to explore the following two research questions: (1) How do students negotiate meaning in audio SCMC and in video SCMC? (2) What roles do multiple modes and semiotic resources play in meaning negotiation episodes in video SCMC?

Four dyads of Chinese postgraduate English language learners performed two types of lexically seeded information gap tasks in audio and video SCMC environments respectively. Meaning negotiation episodes were identified for data analysis. Video stimulated recall interviews were conducted to obtain participants' thoughts during meaning negotiation episodes. Three types of data analysis were carried out, including: (1) an interaction analysis of all audio SCMC negotiated interactions; (2) a statistical analysis of students' gaze directions during meaning negotiation episodes in video SCMC; and (3) a multimodal analysis of students' verbal interactions, gaze directions, facial expressions, and gestures.

The three types of in-depth analyses have led to the important findings. The study has proposed expanded meaning negotiation routines specifically for audio and video SCMC. The gaze analysis discovers a statistically significant positive relationship between the amount of time interlocutors spend looking at each other's video images and the success of meaning negotiation. The multimodal analysis has revealed different levels of multimodal communicative competence and identified a range of relationships between different mode(s) in video SCMC.

Key words: meaning negotiation, audio/video SCMC, multimodal analysis

Acknowledgement

First of all, I would like express greatest thanks to my supervisors Dr. Timothy Lewis, Dr. Qian Kan and Dr. Shi Lijing for their patient guidance in my growth as a researcher, rigorous feedback for every single draft and enormous support in nurturing the production of this thesis over the past five years. My thanks also goes to Dr. Uwe Baumann who was my supervisor for the first two years of my doctoral study. I am truly grateful for the generous research fund offered by the Open University and the support from my colleagues and friends at the department of WELS.

I am also very grateful for the generous help from the Institute of Online Education at Beijing Foreign Studies University, without which this thesis would not have been possible. My most important thanks go to Professor Jinlan Tang, who has kindly given the permission and her full support for me to carry out the research project and collect data in the department. I am also very thankful for Professor Ligao Wu and Lecturer Chen Li for participating in this study as online teachers. Thank you very much to all my research participants in both the pilot study and the main study who have been actively engaged in the research project.

My special thanks are given to my beloved one, Dr. Aistis Atminas, who has given me all his love, care and patience in my hardest days. Without his company, help and encouragement, I would not have been able to complete this long and hard journey all by myself. He is my light, both in my academic career and in my life.

I would like to especially thank my parents who have given me their selfless love and support. I appreciate their constant effort to offer me the best possible education opportunities at all times. I also want to especially thank my grandmother who grows me up from a toddler with her kindness and love, and who cooks the most delicious dumplings in the world.

Next, I express my gratitude to my favourite singers, Karry Wang, Roy Wang and Mayday, whose stories, songs and lyrics have given me huge strength to carry on through the hardships.

Last but not least, I want to thank myself for not having given up on the Ph.D. despite having thought about it for a hundred times. Wish myself a bright future, a happy and healthy life with all loved ones forever.

Declaration of Authorship

I declare that this thesis has been composed solely by myself and that it has not been submitted, either in whole or in part, in any previous application for a degree. Except where otherwise acknowledged, the work presented is entirely my own.

References to Relevant Works:

Li, C. C., and Lewis, T. (2018). Negotiation for Meaning Routines in Audio SMC Interactions: An Expanded Framework. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, 8(3), 50-72.

This publication presents part of my work from the thesis, particularly related to the interaction analysis in Chapter 5.

Table of Content

1 INTRODUCTION	20
1.1 Rationale for the research topic and context	20
1.1.1 The widespread of SCMC for online education worldwide and in China	20
1.1.2 Research in SCMC for language learning	21
1.2 The structure of the thesis	23
2 LITERATURE REVIEW	25
2.1 Introduction of the literature review	25
2.2 The Interaction hypothesis and negotiation for meaning	30
2.2.1 The Interaction hypothesis	30
2.2.2 Negotiation for meaning	31
2.2.3 A classroom-based model of negotiation for meaning routines	37
2.2.4 Negotiation routine stages	38
2.3 Meaning negotiation in different SCMC contexts	41
2.3.1 Meaning negotiation in text-based SCMC	41
2.3.2 Meaning negotiation in audio SCMC	48
2.3.3 Meaning negotiation in video SCMC	51
2.3.4 Conclusion of Section 2.3.....	75
2.4 Multimodal studies in video SCMC	76

2.4.1 Introduction of multimodality.....	76
2.4.2 Key branches of multimodality studies in video SCMC.....	79
2.4.3 Multimodal orchestration: Comprehensive multimodal studies.....	98
2.4.4 Conclusion of Section 2.4.....	105
2.5 Chapter conclusion and research questions.....	106
3 METHODOLOGY.....	107
3.1 Introduction.....	107
3.2 Philosophical stances.....	109
3.2.1 Positivism and post-positivism.....	110
3.2.2 Constructivism.....	111
3.2.3 Pragmatism.....	112
3.2.4 Rationale for choosing pragmatism.....	113
3.3 Research design.....	116
3.3.1 Approach to theory development.....	116
3.3.2 Methodological choices: mixed methods.....	117
3.3.3 Methodological strategies: educational intervention research.....	119
3.4 Chapter conclusion.....	121
4 DATA COLLECTION METHODS AND PROCEDURES.....	122
4.1 Research context, access, and ethics.....	122
4.1.1 The synchronous audio/video SCMC in the research context.....	122
4.1.2 Access and ethics.....	123

4.2 Research design and data collection procedures	125
4.2.1 Pilot study	125
4.2.2 Participants in the main study.....	126
4.2.3 Research design and rationale in the main study.....	129
4.2.4 Data collection procedures in the main study.....	134
4.3 Chapter conclusion	140
5 INTERACTION ANALYSIS AND FINDINGS	141
5.1 Rationale and coding scheme	141
5.1.1 Rationale for interaction analysis.....	141
5.1.2 Interaction analysis method and coding scheme	142
5.2 Interaction analysis	144
5.2.1 Conforming to previous frameworks of meaning negotiation.....	144
5.2.2 Two new stages: Confirmation of Trigger (CT) and Clarification of Indicator (CI)	147
5.2.3 Technical issues	154
5.3 Findings	154
5.3.1 Three possible routines.....	154
5.3.2 An expanded meaning negotiation routine in audio SCMC.....	156
5.4 Chapter conclusion	158
6 GAZE ANALYSIS AND FINDINGS	159
6.1 Rationale, coding scheme, and coding method	159

6.1.1 Rationale for a statistical gaze analysis	159
6.1.2 The coding scheme	160
6.1.3 How to identify gaze	165
6.2 The coding process	175
6.2.1 What episodes are included?	175
6.2.2 Coding with ELAN	176
6.3 Statistical gaze analysis and findings	177
6.3.1 The overview of data	177
6.3.2 Statistical analysis	181
6.3.3 Other possible factors	187
6.3.4 Final statistical findings	189
6.4 Chapter conclusion.....	190
7 MULTIMODAL ANALYSIS AND FINDINGS	192
7.1 Data analysis methods.....	192
7.1.1 Rationale for the multimodal analysis	192
7.1.2 What episodes are included for the multimodal analysis?	193
7.1.3 The critical incident approach	194
7.1.4 Multimodal analysis and transcripts	196
7.2 Multimodal analysis.....	198
7.2.1 Dyad 4 Task 3: 'saltshaker' episode	198
7.2.2 Dyad 1 Task 5: 'razor' episode	209
7.2.3 Dyad 3 Task 5: 'magnifying glass' episode	221

7.2.4 Dyad 2 Task 6: 'nail polish' episode.....	232
7.3 Findings	256
7.3.1 Proposing a multimodal communicative competence pyramid	256
7.3.2 The relationships among different modes/channels in meaning negotiation in video SCMC	260
7.4 Chapter conclusion.....	264
8 DISCUSSION	266
8.1 How do students negotiate meaning in audio SCMC and video SCMC?	266
8.1.1 Meaning negotiation routines in audio SCMC	266
8.1.2 Meaning negotiation routines in video SCMC-Directing the use of modes	271
8.1.3 Meaning negotiation routines in different modalities	280
8.2 What roles do multiple modes and semiotic resources play in meaning negotiation episodes in video SCMC?	283
8.2.1 The role of the video channel in meaning negotiation in video SCMC	284
8.2.2 The role of multimodal interactions in meaning negotiation in video SCMC	287
8.2.3 The differences between video SCMC and face-to-face communication	292
9 CONCLUSION	296
9.1 Summary of the research findings	296
9.2 Contribution to knowledge	298
9.3 Limitations of the study	299
9.3.1 Limited episodes in the multimodal analysis.....	299

9.3.2 Task design	299
9.3.3 Contextual factors	300
9.4 Implications.....	300
9.4.1 Implications for future research	300
9.4.2 Implications for online teaching practice	301
REFERENCES	302
APPENDICES.....	330
Appendix 1: Mock IELTS Speaking Test.....	330
Appendix 2: Opinion Gap Tasks	331
Appendix 3: Task 3	332
Task 3: Spot-the-differences Sheet A (video conferencing).....	332
Task 3: Spot-the-differences Sheet B (video conferencing).....	332
Appendix 4: Task 4	333
Task 4: Spot-the-differences task for Student A (audio conferencing)	333
Task 4: Spot-the--differences task for Student B (audio conferencing)	333
Appendix 5: Task 5	334
Task 5: Problem-solving task for Student A (video conferencing)	334
Task 5: Problem-solving task for Student B (video conferencing)	335
Appendix 6: Task 6	336
Task 6: Problem-solving task for Student A (audio conferencing)	336
Task 6: Problem-solving task for Student B (audio conferencing)	337

Appendix 7: Information sheet for participants	338
Appendix 8: Sample signed consent form	339
Appendix 9: Interview questions	340

List of abbreviations

CALL: computer-assisted language learning

DVC: desktop videoconferencing

F2F: face-to-face

L1: first language

L2: second language

MNE: meaning negotiation episodes

NfM: negotiation for meaning

NS: native speaker

NNS non-native speaker

SCMC: Synchronous computer-mediated communication

SLA: second language acquisition

VSRI: video stimulated recall interview

TS: task sheet

PVI: peer's video image

OD: other directions

UI: unidentifiable gaze

Abbreviations for meaning negotiation stages

T: trigger

I: indicator

R: reply

RR: response to reply

C: confirmation

RC: re-confirmation

CT: confirmation of trigger

CI: clarification of indication

List of Figures

Figure 1: Locating the research topic in the existing literature	26
Figure 2: Model for negotiation of meaning (Varonis and Gass, 1985)	38
Figure 3: Smith's (2003) Model of Computer-Mediated Negotiated Interaction. Adapted and expanded from Varonis and Gass (1985)	43
Figure 4: An online pedagogical interaction from different perspectives by Guichon and Wigham (2016, p.64)	83
Figure 5: Alignment of trainee hors-champ video with trainee and student webcam videos in ELAN by Guichon and Wigham (2016, p.71).....	84
Figure 6: Continuum of framing choices at minute seventeen of the interaction (2016, p.72)	85
Figure 7: A sample transcription by Lee et al. (2019, p.32)	88
Figure 8: The research onion model by Saunders, Lewis, and Thornhill (2019, p. 5)	108
Figure 9: The interface of the video conferencing platform.....	123
Figure 10: Theoretical frameworks for analysing meaning negotiation routines.....	143
Figure 11: Negotiation for meaning pathways as exemplified in three audio SCMC interactions	155
Figure 12: A negotiation for meaning routine for audio SCMC interactions: An expanded model	157
Figure 13: Linear regression between the time of gaze directed at peer's video image	185
Figure 14: The multimodal communicative competence pyramid in video SCMC	258
Figure 15: A negotiation for meaning routine for audio SCMC interactions: An expanded model	268
Figure 16: Three pathways of negotiation for meaning in audio SCMC interactions.....	269
Figure 17: Linear regression between the time of gaze directed at peer's video and number of successful MNEs	286
Figure 18: The multimodal communicative competence pyramid for meaning negotiation in video SCMC.....	289

Figure 19: The size of interlocutors' face in the video SMC interface294

List of Tables

Table 1: Subcategories of Negotiation Routine Stages (summarised by Smith, 2003)	39
Table 2: A summary of studies on meaning negotiation in video SCMC contexts.....	55
Table 3: Six types of framing of online teachers in video SCMC summarized from Codreanu and Celik (2013)	82
Table 4: Classification of screen shots by Guichon and Wigham (2016, p.70)	84
Table 5: Referential properties of 'tunnel earring' identified by Cohen and Wigham (2019, p.480)	90
Table 6: Transcription and coding of Phase 1 by Cohen and Wigham (2019, p.464)	91
Table 7: Degrees of the utilisation of webcam and screenshot examples in video SCMC summarized from Develotte et al. (2010)	94
Table 8: Types of Gaze in DVC summarized from Satar (2013).....	96
Table 9: Transcription with verbal description of multiple modes by Satar (2016, p. 314)....	101
Table 10: Transcription with screenshots by Satar (2016, p. 317)	101
Table 11: The Multimodal analytical framework used by Wigham (2017).....	104
Table 12: Summary of methodological choices for this study	109
Table 13: Participants' scores in the mock IELTS speaking test.....	128
Table 14: Research procedure for the main study	134
Table 15: Subcategories of negotiation stages by Smith (2003, p. 43)	143
Table 16: A summary of MNEs in audio SCMC	154
Table 17: Gaze information in meaning negotiation episodes in video SCMC	179
Table 18: The Y variable: the number of successful meaning negotiation episodes.....	183
Table 19: Final data for regression analysis of gaze directions and meaning negotiation success.....	184
Table 20: The regression analysis result between gaze time spent on peer's video image and the number of successful MNEs	185

Table 21: The mock IELTS speaking test scores and the number of successful MNE in video SCMC.....	188
Table 22: Gaze time on the task sheet and the number of successful MNEs.....	188
Table 23: The regression analysis result between gaze time spent on peer's video image and the number of successful MNEs	189
Table 24: Screen time for each dyad in their selected meaning negotiation episodes	195
Table 25: Participants' multimodal communicative competences	258
Table 26: A comprehensive summary of meaning negotiation routines in different communication environments	281

1 Introduction

1.1 Rationale for the research topic and context

1.1.1 The widespread of SCMC for online education worldwide and in China

The last 20 years have witnessed the rapid development of online learning and teaching across the world. Synchronous Computer-Mediated Communication (SCMC) has attracted an increasing amount of attention in recent years. Particularly with the outbreak and spread of COVID-19 across the world in early 2020, traditional face-to-face teaching has been temporarily yet primarily replaced by online teaching using SCMC technology. In their summary of 20 countries' higher education intra-period digital pedagogy responses, Crawford et al. (2020) report that SCMC technology has been used by universities in at least nine countries, including China, Germany, Italy, South Korea, Jordan, Malaysia, Singapore, Egypt, and United Arab Emirates.

In China, the Ministry of Education (MOE) has suggested flexible online teaching methods using SCMC and other forms of online teaching (Huang et al., 2020). Dingtalk (<https://www.dingtalk.com>), an audio/video conferencing package specially designed for synchronous online teaching and learning, has been used by more than 5 million students from more than 10,000 universities and primary schools (Huang et al., 2020). In the private sector, according to the survey by iiMedia Research (2020), synchronous audio/video conferencing has become the most widely used online learning tool, used by 83% of primary and secondary school students. Therefore, research on SCMC for online language learning is highly needed and has significant practical and pedagogical values for schools and universities worldwide.

1.1.2 Research in SCMC for language learning

Among different subjects using SCMC tools, language teaching and learning online have received considerable attention from researchers. As Bax (2003, 2011) predicted, computer-assisted language learning (CALL) has become 'normalized' in the 21st century. In their recent review of SCMC for *Language Teaching Research*, Cunningham and Akiyama (2018) conclude that the field is undergoing reconceptualization and expansion with the advancement of technology and diversification of participants. Following this development, the main research agenda in SCMC has progressed from examining the effectiveness of online learning to how the affordances of different types of technology can be best used for language learning online (Cunningham and Akiyama, 2018). The central argument in this field is that since communication is mediated by technology, the affordances of the technology play an essential role in how learners communicate and learn languages in the mediated environment (Hampel and Stickler, 2005; Hampel 2006; Stockwell, 2010). With the development of SCMC from asynchronous to synchronous communication, from email or text-chat messages to audio and video conferencing environments, the various modes of communication used by interlocutors have been scrutinized ever more closely (e.g. Guichon and Wigham, 2016; Satar, 2016; Wigham, 2017). These different modes form a complicated communication environment where interlocutors can communicate through one single mode or a combination of different modes. So, how the multimodal SCMC environment affects language learning online in audio and video conferencing environments has become an important question (e.g. Smith, 2003; Yanguas, 2010; Sauro, 2011; Hampel and Stickler, 2012; Wang and Tian, 2013; Satar and Wigham, 2017; Lee et al., 2019). This study attempts to contribute to answering this big question.

To answer such a big question, researchers need to focus on a specific aspect of language learning. In their review of theory in CALL research and practice, Hubbard and Levy (2016) observe that among many second language acquisition (SLA) theories, the *Interaction Approach* has been extensively cited as a theoretical base in SCMC research, especially in

studies that involve text-based, audio or video conferencing as a basis for language learning interactions. This approach emphasizes the importance of learner interaction in SLA (Long, 1996). As Ellis (2000, p.209) points out, learning arises not through interaction but in interaction. Central to the interaction hypothesis is the notion of 'negotiation for meaning', which refers to the process of solving the non-native speaker's¹ non-understanding in the target language (Long, 1988). According to Long (1996), meaning negotiation episodes can be observed in both NS-NNS and NNS-NNS interactions. Therefore, negotiation of meaning seems to be a suitable linguistic episode to focus on when studying language learning through audio/video SCMC.

As for participants, this study hopes to focus on Chinese English language learners for the following reasons. Despite many studies examining meaning negotiation episodes in audio/video SCMC (e.g. Wang, 2006; Lee, 2006; Yanguas, 2010; Jung and Jie, 2012; Wang and Tian, 2013; Guo and Möllering, 2016), very few researchers have involved Chinese participants learning English as a second language. Yet, some studies describe Chinese students as goal-oriented language learners who do not appear to value spending time discussing the language and negotiating meaning (Littlewood, 2007; Wen, 2018). 'Chinese English language learners' are a huge group of people with varying educational, social, and cultural backgrounds. China is a huge country, and each province or area has its unique history,

¹ The notion of 'nativeness' in the terms 'native speaker' and 'non-native speaker' has been critically discussed from different perspectives in social science such as critical discourse, applied linguistics, and language and culture studies. These discussions concern the speaker's identity, self-recognition, childhood experience, language proficiency, and cultural, educational and political background (e.g. Davis, 2004, 2012; Holliday, 2006; Doerr, 2009). However, in the research field of English language teaching (ELT), particularly in interaction and second language acquisition (SLA) studies that are closely related to the current thesis, the abbreviations 'NS' and 'NNS' have been widely used for decades (Long, 1983; 1996; Smith, 2003, 2004; Wang & Tian, 2010, 2013; Zwaad and Bannink, 2014, 2016). Therefore, this study chooses to follow SLA studies' tradition and uses 'NS' and 'NNS'.

culture, economic and educational development. It is questionable whether all 'Chinese students' share such stereotyped features as described by the above studies. In addition, those few studies (Zheng, 2011; Chen, 2014; Feng, Chen, Shen, 2015) where researchers investigate meaning negotiation patterns in online SCMC environments with Chinese English language learners still focus on text-chat conversations rather than oral interactions in audio and video SCMC environments. Therefore, meaning negotiation in audio and video SCMC by Chinese English language learners remains a highly under-researched area. It is not clear whether Chinese English language learners are willing to negotiate meaning in online oral interactions, and if they are, what patterns they use to negotiate meaning in audio and video SCMC. The significant number of students using online audio and video conferencing tools for language learning in China and the lack of research on Chinese learners make it necessary for this study to focus on this particular group of language learners as research participants. The findings of this study will not only contribute to the field of multimodality research but also have significant pedagogical and practical values for millions of Chinese students learning English online through audio/video SCMC tools.

1.2 The structure of the thesis

The above paragraphs briefly introduced the research topic of the current study and highlight its importance. Section 1.2 will present how this thesis is structured to address the research topic in the context of online language learning by Chinese students.

This study aims to answer the following two research questions:

- (1) How do students negotiate meaning in audio SCMC and in video SCMC?
- (2) What roles do multiple modes and semiotic resources play in meaning negotiation episodes in video SCMC?

The structure of the thesis is as follows. The current chapter (Chapter 1) will offer a broad overview of the research topic and introduces some relevant background information about SCMC in China. It will be followed by a comprehensive literature review, where the relevant literature of the three research fields (Interaction studies, SCMC, and multimodality) will be reviewed, and research gaps will be identified as providing the basis for research questions (see above). Chapter 3 will move on to explain the philosophical stance taken and the research methodology suitable for answering the research questions. More specific research design, data collection techniques, and procedures will be presented in Chapter 4. The data analysis and findings will be divided into three independent chapters because each type of data analysis will focus on a particular type of data and use a distinctive analytical method. Chapter 5 will offer an interactional analysis of meaning negotiation routines in audio SCMC. Chapters 5 and 6 will focus on meaning negotiation episodes in multimodal video SCMC. Specifically, a statistical analysis will be presented in Chapter 6 to establish the relationship between the direction of gaze and meaning negotiation outcomes in video SCMC, while Chapter 7 will focus on the multimodal analysis of how multiple modes and semiotic resources are used to negotiate meaning in video SCMC. Based on these findings, Chapter 8 will provide answers to the two research questions, discuss the contribution to the research field. Finally, Chapter 9 will present a brief conclusion to the whole thesis, reflect on the limitations of the study, and offer suggestions for future research and pedagogical/practical implications.

2 Literature Review

2.1 Introduction of the literature review

This study aims to take a multimodal approach to examining negotiation for meaning episodes in video-based synchronous computer-mediated communication. The research topic is at the intersection of three key research fields, as illustrated in Figure 1. They comprise: (a) the interaction hypothesis and negotiation for meaning, (b) SCMC studies, and (c) multimodality studies. Section 2.1 aims to briefly introduce the relevant research fields, locate the research topic, and outline the literature review structure. A more detailed review of the relevant theories and studies and the identification and justification of the specific research questions will be presented in the remaining parts of the chapter.

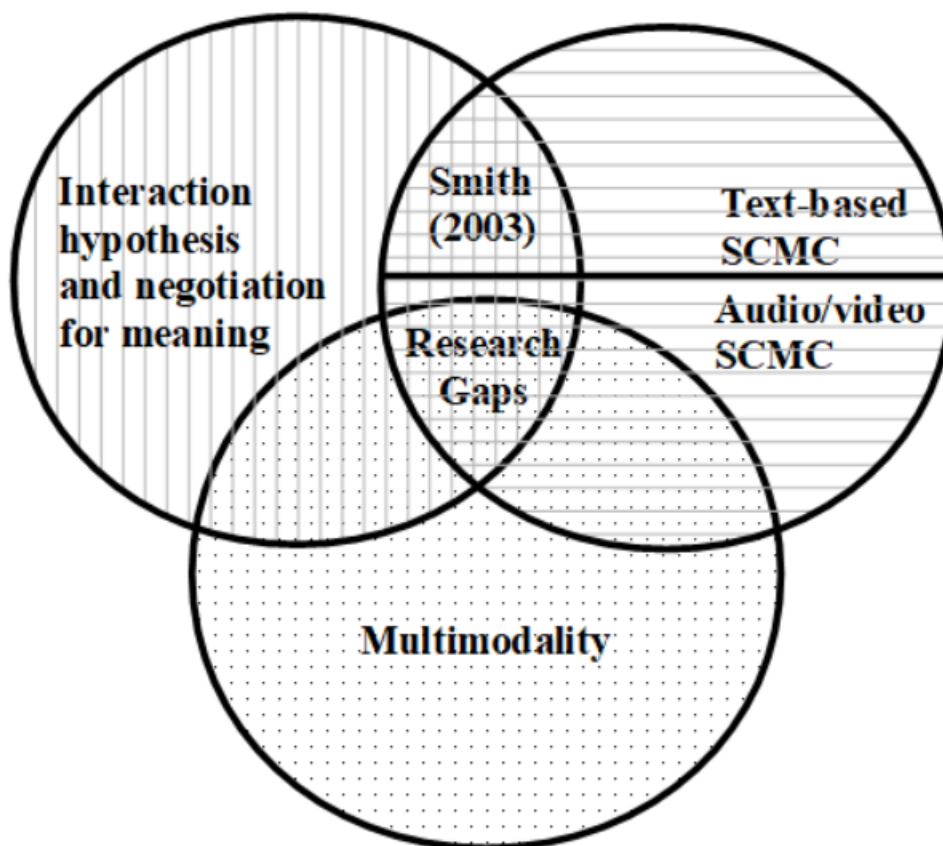


Figure 1: Locating the research topic in the existing literature

The first research area (the circle with vertical stripes) covers the Interaction Hypothesis and negotiation for meaning, which flows from it. Interaction has been identified as one of the key factors for SLA (e.g. Long, 1983, 1996; Ellis, 2003). As Ellis (2003, p.209) points out, 'learning arises not *through* interaction but *in* interaction.' The interaction hypothesis assumes that 'negotiation for meaning, and especially negotiation that triggers interactional adjustments by the native speaker or more competent interlocutor, facilitates acquisition because it connects input, internal learner capacities, particularly selective attention, and output in productive ways' (Long, 1996, pp.451-452). Negotiation for meaning (NfM) has been a central but contested feature of interaction (Varonis and Gass, 1985). Despite some questioning of the role of interaction for SLA (e.g. Mackey and Goo, 2007; Adams, 2007), the Interaction Hypothesis has

long been considered to be one of the ways in which SLA takes place (e.g. Long, 1996; Ellis, 2003; Foster and Otha, 2005). Originally observed and proposed in classroom-based settings, the negotiation for meaning framework (Varonis and Gass, 1985) has been adapted and employed to analyse different forms of online communication, ranging from asynchronous email conversations and forum discussions to synchronous audio/video communication, from text-chat messages (e.g. Smith, 2003) to voice-based and webcam-mediated interactions (e.g. Wang, 2006, 2007), from computer-mediation (e.g. Smith, 2003; Yanguas, 2010; Guichon and Cohen, 2016) to mobile device mediation (e.g. Lee et al., 2019). In this study, the term 'negotiation for meaning', 'meaning negotiation', and 'negotiated interaction' are used interchangeably to refer to the same concept.

The second field of interest (the circle with horizontal stripes) is SCMC, which belongs to the broader research area of computer-assisted language learning (CALL). As Bax (2003, 2011) foresees, CALL has become normalized as technology has been fully integrated into second language learning, teaching, and research. Among various approaches to CALL, SCMC has been one of the most commonly used and widely researched (Cunningham and Akiyama, 2018). Ever since the existence of SCMC, researchers have attempted to apply and adapt SLA theories derived from face-to-face classroom-based settings to examine interactions in SCMC environments. For example, Chapelle (1997) is one of the earliest researchers to employ the interactionist approach to investigate language learners' discourse in online SCMC environments, demonstrating how SLA research methods can be applied to CALL research contexts. Furthermore, Smith (2003) expands the negotiation for meaning routines framework by Varonis and Gass (1985) to address the differences between text-based SCMC and face-to-face communication. According to Hubbard and Levy (2016), among many SLA theories, the Interaction Hypothesis has been extensively referenced as a theoretical base in SCMC research, especially in studies that involve text-based or audio- or video-conferencing as a basis for learner interaction and exchange. In Figure 1, this strand of research is presented in the area with a grid background where the circle with vertical stripes (interaction hypothesis

and negotiation for meaning) and the circle with horizontal stripes (SCMC studies) overlap with each other.

The third relevant research field is the study of multimodality (the circle with dots), which approaches representation, communication, and interaction as something more than language (Jewitt, 2009, p. 15). It is widely believed that multimodality plays an essential role in communication (Kress and Van Leeuwen, 1996, 2001; Norris, 2004; Levine and Scollon, 2004; Jewitt, 2009). As Jewitt (2009) explains, 'the starting point for multimodality is to extend the social interpretation of language and its meaning to the whole range of representational and communicational modes or semiotic resources for making meaning employed in a culture - such as image, writing, gesture, gaze, speech, posture' (p.15). She considers multimodality more as an approach or a research perspective that can be applied to a wide range of research subjects than a research subject itself. Actually, the study of multimodal meaning making dates back to days before SCMC became broadly available for distance education, and it initially dealt with the use of gesture (kinesics) and posture (proxemics) in face-to-face communication for SLA purposes (the purple area) (e.g. Birdwhistell, 1952; McNeil, 1992; Faraco and Kida, 2008), classroom-based multimodal teaching and learning (e.g. Kress and van Leeuwen, 2001; Bourne et al., 2004), and image interpretation and media communication (e.g. Kress and Van Leeuwen, 1990, 1996, 2006). When SCMC began to be used for language learning in the late 1990s, it gave rise to a new area of research because communication is mediated by computers, and the technological mediation can both 'afford' and 'constrain' the meaning negotiation process in SCMC environments. Especially in video SCMC where the webcam makes the visual channel available, the main research question in the overlapping field of SCMC (the circle with horizontal stripes) and multimodality studies (the circle with dots) is how different modes (such as gaze, gesture, posture, and facial expression, as well as verbal communication) are used for communication in computer-mediated learning and teaching contexts (e.g. Develotte, et al., 2010; Hampel and Stickler, 2012; Codreanu and Celik, 2013; Satar, 2013; Guichon and Cohen, 2014, 2016; Guichon and Wigham, 2016, Kern, 2014; Lee

et al., 2019). This study uses 'multimodal communicative competence' (MCC) to refer to students' ability to make use of multiple modes to negotiate meaning in video SCMC.

Research to date has arguably been located in areas where any two of three research fields overlap. However, there still exists a research gap at the intersection of all three fields, which remains unexplored, except by Lee et al. (2019), who mainly study hand gestures and negotiation for meaning in mobile video SCMC contexts. However, Lee et al. (2019) primarily focuses on hand gestures and the use of mobile devices for meaning negotiation rather than offering a comprehensive multimodal analysis of other semiotic resources including gaze, facial expression, proximity, and the relationships among these modes. Therefore, this study aims to fill this research gap by taking a multimodal approach to exploring negotiation for meaning episodes in audio and video SCMC.

This review first briefly introduces the theoretical development of the interactionist approach to SLA, and a framework for analysing meaning negotiation routines in face-to-face (F2F) classrooms (Varonis and Gass, 1985). Then it moves on negotiation for meaning in SCMC contexts, starting from an expanded framework (Smith, 2003) in text-chat written SCMC, to voice-based audio SCMC, to webcam-based multimodal SCMC environments. It will be followed by comparisons of meaning negotiation among these different SCMC contexts. Afterwards, more recent research on multimodality in SCMC for SLA will be reviewed to demonstrate theories, methods, and findings from existing literature. Finally, specific research questions will be formulated based on the literature review.

2.2 The Interaction hypothesis and negotiation for meaning

2.2.1 The Interaction hypothesis

Among different approaches to SLA, the Interaction Hypothesis has been one of the most well discussed and examined theories in the past few decades in relation to both traditional face-to-face classrooms and SCMC contexts. Based on the initial proposal by Long (1981), the current version of the Interaction Hypothesis (Long, 1996) is the result of numerous strands of discussion contributed by many researchers exploring SLA from various perspectives between the 1970s and 1990s (e.g. Ferguson, 1971; Wagner-Gough and Hatch, 1975; Krashen, 1977, 1980; Long, 1980, 1981, 1983; Swain, 1985, 1995; White, 1991, 2003; Schmidt, 1990, 1993).

As Mackey et al. (2011) recall, the roots of the interactionist approach can be traced back to the 1970s when researchers (e.g. Ferguson, 1971) study 'foreigner talk' or 'care-taker talk' and concluded that modifications (e.g. repetition, syntactic simplification) which makes input more comprehensible can promote target language acquisition. Subsequently, Krashen (1977, 1980) proposes the Input Hypothesis emphasizing two crucial elements for SLA: comprehensible input that is slightly above the learner's current level of proficiency and low levels of anxiety or negative feelings for learning L2 (second language). In 1978, Hatch (1978a, 1978b) stresses the importance of interaction for SLA. Building on the above research, Long (1980, 1981) proposes the first version of the Interaction Hypothesis, claiming that interactional adjustments (also called 'conversational modifications') during negotiation for meaning between native speakers (NS) and non-native speakers (NNS) are the key to promoting comprehensible input and second language acquisition.

Although Krashen's input theory and Long's conversational modification theory were used by many researchers to describe learner interactions, they were also criticized as being descriptive and insufficient to explain the causal relationship between interaction and SLA. For example, Swain (1985, 1995), based on her study with French immersion students in Canada, identifies the crucial role of output for SLA, arguing that pressure to produce output pushes learners to produce target language, test hypotheses, focus on form, and notice gaps in their interlanguage. Moreover, Schmidt (1990, 1993) further explores the role of input, finding out that when a learner modifies his/her speech, his/her attention may be drawn to the part of the input that has not been understood, thus optimizing the opportunity for learning to occur. Therefore, in his Noticing Hypothesis, he emphasizes that 'noticing is the necessary and sufficient condition for converting input to intake'.

Drawing on the work of these researchers, Long (1996, p.451-452) reformulates the Interaction Hypothesis as follows: Negotiation for meaning, and especially negotiation that triggers interactional adjustments by the native speaker or more competent interlocutor, facilitates acquisition because it connects input, internal learner capacities, particularly selective attention, and output in productive ways. This version of the Interaction Hypothesis emphasizes that oral interaction works by 'connecting input, internal learner capacities, and output via selective attention' (Ellis, 2008, p. 257) and identifies the 'conditions under which ideal input and interactions take place' (Chapelle, 1999, p.5).

2.2.2 Negotiation for meaning

Central to Long's (1996) Interaction Hypothesis is the notion of 'negotiation for meaning'. Researchers have different definitions and understanding of the notion due to their specific research contexts and research questions. The following section aims to present mainstream arguments in literature and justify which perspectives are suitable for the current study.

2.2.2.1 What is 'negotiation for meaning'?

Pica (1994) defines negotiation as the modification and restructuring of interaction between interlocutors when they experience comprehension difficulties (p. 494). Long (1996) describes negotiation for meaning as 'the process in which learners and competent speakers provide and interpret signals of their own and their interlocutor's perceived comprehension, thus provoking adjustments to linguistic form, conversational structure, and message content or all three, until an acceptable level of understanding is achieved' (p. 418). These 'signals' or 'negotiation strategies' such as repetitions, clarification requests, confirmation checks, and recasts are examples of negative feedback (Long, 1996; Oliver, 2000). When there is a communication breakdown between interlocutors, negative feedback can be delivered explicitly through overt error correction or implicitly through negotiation strategies (Long, 1996). Ellis (2003) summarises negotiation for meaning as 'the process by which two or more interlocutors identify and then attempt to resolve a communication breakdown' (p. 346). On the one hand, although differently phrased, all three definitions share two key messages in common: (a) all highlight that the cause of negotiation is a communication breakdown or comprehension difficulties, and (b) negotiation is the process to resolve this problem through modification/adjustments. On the other hand, these definitions all have their specific emphases. For example, Pica (1994) stresses the modification and restructuring of the interaction, while Ellis (2003) adds the identification of a communication breakdown. Long's (1996) definition clarifies the relationship between two interlocutors: one is more proficient than another in the target language. He also specifies what can be modified, such as linguistic form, conversational structure, or message content.

In the literature of conversational interaction, both 'negotiation of meaning' and 'negotiation for meaning' are used, in many cases, either interchangeably or without any clear distinction between the two being made (e.g. Foster, 1998; Foster and Otha, 2005; Wang, 2006; Wang and Tian, 2013). However, 'negotiation for meaning' has imposed itself in recent years (see Mackey, 2007; Cook, 2015) as the more logical. Negotiation in this context does not involve a

choice between competing meanings. It is an attempt to move from non-understanding to a single shared understanding. The preposition 'for' seems to indicate that meaning is socially constructed through communication (Jonassen et al., 1995) as envisaged in social constructivism. Since this study aims to investigate how two learners work together to solve a non-understanding and arrive at a shared understanding, 'negotiation for meaning' is the usage that will be adopted in this study.

In this thesis, the terms 'negotiation for meaning', 'meaning negotiation' and 'negotiated interactions' are used interchangeably to denote the same concept, as explained above.

2.2.2.2 What counts as 'meaning'?

The discussion of meaning is mainly concerned with the extent to which meaning in 'negotiation for meaning' should include form. Form here is often used to refer to grammar and linguistic structures (Long and Robinson, 1998). The idea of 'focus on form' has been proposed and widely discussed (e.g. Long, 1991; Long and Crookes, 1992; Long and Robinson, 1998), especially as opposed to 'focus on meaning'. Among various definitions for 'focus on form', Long and Robinson's (1998) definition has 'offered researchers and practitioners greater direction for practical implementation' (Doughty and Williams, 1998):

Focus on form refers to how focal attentional resources are allocated. Although there are degrees of attention, and although attention to forms and attention to meaning are not always mutually exclusive, during an otherwise meaning-focused classroom lesson, focus on form often consists of an occasional shift of attention to linguistic code features – by the teacher and or one or more students – triggered by perceived problems with comprehension or production (*op. cit.*: 23; emphasis in original).

Researchers hold different opinions towards whether 'meaning' negotiation for meaning should include 'form' as well. For example, Lyster (2002) argues for a broader view of negotiation that

accounts for corrective feedback and distinguishes between form-focused negotiation and meaning-focused negotiation in teacher-student interaction. However, Pica (1994) believes that target language accuracy (or form) plays only a secondary role in negotiation, as 'negotiation, by definition, focuses on the comprehensibility of message meaning, and on the message's form only insofar as that can contribute to its comprehensibility' (pp. 517–518). According to Pica's definition, pronunciation and (sometimes) spelling should be included in 'meaning', as these factors relate to comprehensibility.

Since the current study focuses on 'negotiation for meaning' episodes in learner-learner interaction for communication purposes in spoken interactions, rather than teacher-learner interaction for pedagogical purposes, Pica's (1994) perspective on meaning is what this study will espouse.

2.2.2.3 Interlocutors in negotiation for meaning episodes

In most of Long's studies and in the interaction studies reviewed by him in 1996, the focus is on investigating meaning negotiation in conversations between a native speaker (NS) and a non-native speaker (NNS) (Ferguson, 1971; Wagner-Gough and Hatch, 1975; Krashen, 1977, 1980; Long, 1980, 1981, 1983; Swain, 1985, 1995; White, 1991, 2003; Schmidt, 1990, 1993). However, in the updated version of the interaction hypothesis, he enlarges this to include both the 'interactional adjustments by the native speaker or the more competent interlocutor' (Long, 1996, p. 451- 452). This change indicates that negotiation can occur among two types of dyads: a native speaker and a non-native speaker (NS/NNS) or two non-native speakers (NNS/NNS), where one is more competent than the other.

Existing research has shown that whether the language learner's interlocutor is an NS or an NNS has an essential influence on the number and the ways of meaning negotiation. For example, Varonis and Gass (1985) examine whether the fact that interlocutors are native

speakers or not can affect the number of meaning negotiation episodes. They compare NS/NS, NS/NNS, and NNS/NNS conversations and find out that negotiation of meaning is most prevalent among NNS/ NNS pairs. They stress the importance of NNS/NNS peer interaction. Mackey et al. (2003) compare the types of interaction feedback and modified output in negotiation for meaning episodes among NS/NNS and NNS/NNS in a face-to-face classroom context. It is found that although 'NSs provided significantly more feedback than NNS, feedback from NNS offered significantly more opportunity for modified output than that from NSs' (p. 54). Such differences in negotiation for meaning between NS/NNS and NNS/NNS also holds in SCMC contexts. For example, Sotillo (2005) explores corrective feedback in negotiation for meaning episodes in a text-based SCMC context and observed that NNSs provide more explicit feedback and more extensive error correction to ELS learners than NSs. Similarly, in text-based SCMC, Liu (2017) finds that negotiated interactions and successful negotiation episodes occur more frequently among NNS dyads than NS/NNS dyads.

The above paragraph demonstrates the necessity to acknowledge and distinguish the differences between negotiation for meaning by NS/NNS dyads or NNS/NNS dyads. The current study chooses to focus on NNS dyads' peer interactions, which refers to 'any communicative activity carried out between ... L2 learners, not native speaker peers, and not teachers' (Philip et al., 2013, p.3). The reason for this choice is that L2 learner peer interaction is under-researched, especially in second language learning and teaching contexts in China, as described in the Introduction. More importantly, peer interaction between NNSs has been found to have important pedagogical values for SLA in various ways (Philip et al., 2013). For example, having examined group work activities in second language classrooms, Long and Porter (1985) find that peer interactions can offer opportunities for learners to practice communication patterns, take on new conversational roles and engage in negotiation. Similarly, Swain (2000) argues that collaborative dialogue between L2 learners engaging in information gap tasks can mediate language learning. In NNS/NNS peer interaction, since peers are more approachable than teachers, the less competent peer might feel less afraid of making mistakes

and have more chances to experiment with their language use (Damon and Phelps, 1989; Topping and Ehly, 1998). Meanwhile, more competent learners can benefit from being pushed to articulate explanations to their partners and from being placed in a teacher's position (van Lier, 1996; Watanabe and Swain, 2007). Therefore, the current review focuses primarily on meaning negotiation between NNS/NNS dyads.

2.2.2.4 The role of negotiation for meaning in SLA

The fact that the Interaction Hypothesis remains a 'hypothesis' indicates that there are disagreements and questions over the effectiveness of interaction in promoting SLA. Such disagreement has led many researchers to seek to verify the effectiveness of interaction for SLA. While conceding that 'interaction plays a strong facilitative role in the learning of lexical and grammatical target items', Mackey and Goo (2007) suggest that the paucity of empirical studies dealing with negotiation 'render[s] any arguments for the efficacy of one kind of feedback over another premature' (p.440). Adams (2007) observes that 'while research indicates that negotiation for meaning may be quite frequent in learner-learner interactions, there remains little evidence of its effectiveness in promoting learning' (p. 33). However, Adams's own study leads her to the conclusion that 'learner-learner interactions can promote second language learning, suggesting that the benefits of interaction are not limited to the native speaker-learner context' (p.51). Moreover, there is some debate concerning (a) the pedagogical value and feasibility of meaning-focused negotiation in student-student interaction (Aston, 1986; Foster, 1998), and (b) its effectiveness in drawing learners' attention to form in teacher-student interaction (Lyster, 2001, 2002). Despite uncertainty about the developmental value of negotiation, a significant amount of effort has been expended by SLA scholars in studying the relative effectiveness of the individual feedback moves (e.g. clarification requests, prompts) associated with negotiated interactions (Mackey and Goo, 2007). Whatever its precise role in supporting language acquisition, negotiation certainly facilitates continued interaction between interlocutors. Foster and Ohta (2005) take the middle way that both

acknowledges the positive role of negotiation for SLA yet does not exclude other possible pathways to SLA by emphasizing that negotiation is 'one of a range of conversational processes that facilitate SLA as learners work to understand and express meaning in the L2' (p.402).

Among different approaches to exploring negotiation for meaning, two are commonly used by most researchers: the conversation analysis approach, which focuses on interactional feedback moves and negotiation strategies, and the interactionist approach which identifies and examines meaning negotiation routines in different contexts (e.g. Varonis and Gass, 1985; Smith, 2003). These two approaches are not mutually exclusive, and there does not seem to be a clear boundary to distinguish the two from each other. However, it is clear from the existing literature that the latter is more frequently used in SCMC research (Smith, 2003, Wang 2006; Jung and Jie, 2012). Since the current study will build on the findings of the existing SCMC studies, the interactionist approach is more suitable to analyse stances of negotiation for meaning. The following section introduces the framework by Varonis and Gass (1985) for the analysis of meaning negotiation routines.

2.2.3 A classroom-based model of negotiation for meaning routines

Varonis and Gass (1985) introduce a model for analysing the patterns of meaning negotiation between non-native speakers. This model consists of two main parts: a trigger and a resolution which involves three phases, as shown in Figure 2. The first part, a *trigger* (T), is an utterance that causes non-understanding to the hearer. Then, the hearer signals non-understanding through an *indicator* (I). A *response* (R) phase is when the speaker remedies the non-understanding. The last phase occurs when the hearer utters a *reaction to the response* (RR). Varonis and Gass (1985) explain that such negotiation episodes occur during NNS-NNS

interaction primarily because of a lack of shared background and shared incompetence in the target language.

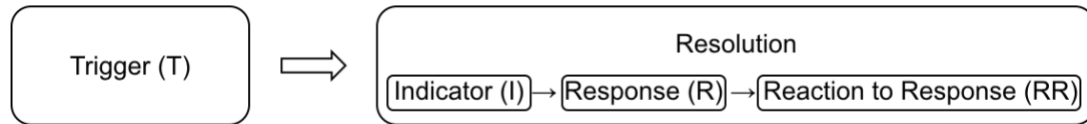


Figure 2: Model for negotiation of meaning (Varonis and Gass, 1985)

This model is the first model to systematically define and examine every single stage of a meaning negotiation routine. Although this routine was concluded from face-to-face oral communication, it has been widely used in many different research contexts, including synchronous multimodal video conferencing SCMC (e.g. Yanguas, 2010, Van der Zwaard and Bannink, 2014, 2016; Guo and Möllering, 2016). The model has also been further specified and expanded by later researchers to account for the effects of different modes on meaning negotiation routines in text-based and audio-based SCMC (e.g. Smith, 2001, 2003; Li and Lewis, 2018). The next section moves on to review how this model has been expanded in text-chat SCMC contexts.

2.2.4 Negotiation routine stages

Based on Varonis and Gass' (1985) model of meaning negotiation, many researchers have further developed subcategories within each of the four meaning negotiation routine stages (e.g. Bremmer et al., 1988; Pica, 1991; Rost and Ross, 1991; Pellettieri, 1999). Smith (2001; 2003) summarises the meaning negotiation component for SCMC interaction as shown in Table 1. The following paragraphs offer definitions of these subcategories and provide examples for demonstration.

Trigger	Indicator	Response	Reaction to Response
Lexical	Global	Minimal	Minimal
Syntactic	Local	RT + Lexical	Metalinguistic talk
Discourse	Inferential	Rephrase/	Task appropriate response
Content		Elaboration	Testing Deductions

Table 1: Subcategories of Negotiation Routine Stages (summarised by Smith, 2003)

2.2.4.1 Trigger

Research shows that most triggers of negotiation routines are lexical in nature (Smith, 2003). *Lexical triggers* are those cases where the problematic utterance can be clearly linked to a specific lexical item. *Syntactic triggers* refer to non-understandings caused by structural or grammatical construction. *Discourse triggers* are related to the general coherence of the conversation, such as non-understanding caused by an inability to reference the antecedent of a pronoun correctly. Content triggers are defined as instances where the entire content of a previous message is in some way problematic. New types of triggers have been identified in more recent studies. For example, Jung and Jie (2012) report many occurrences of non-understanding caused by pronunciation, which are categorized as *phonological triggers*.

2.2.4.2 Indicator

According to Varonis and Gass (1985). Indicators or signals are executed by the initiator of the negotiation routine and can be explicit or implicit. Rost and Ross (1991) classify indicators into three subcategories including global, local, and inferential. *Global indicators* are those cases where the respondent indicates non-understanding without identifying specific problems, as opposed to *local indicators*, where the respondent explicitly identifies the trigger (e.g. 'What's

his name again?'). Local indicators can also include clarification requests (e.g. 'What does wrench mean?') and confirmation checks (e.g. 'Do you mean *machine*?' after the interlocutor wrote *mascien*). *Inferential indicators* occur when the respondent tests out a hypothesis and in doing so signals their non-comprehension. An example of this is when the respondent asks 'OK, so that means he is tired?' after his interlocutor has attempted to describe a man who is bored (Smith, 2003).

2.2.4.3 Response

The third stage of a negotiation routine is the response. Smith (2003) summarises three main types of responses: minimal response, repeating the trigger (with lexical or syntactic modifications), and rephrasing or elaboration. A minimal response is a concise reply (usually one or two words) to the indicator with little new input (e.g. --'Do you mean machine?', --'Yes'). *Repeating the trigger* (in most cases, a lexical item) with lexical modification to the surrounding text is a learner's attempt to clarify the speaker's intended meaning. However, the respondent does not address the fundamental problem signalled in the indicator phase. *Rephrasing and elaborating* seems to be most helpful to resolve the interlocutor's non-understanding and enhance meaning negotiation. Apart from these cases, sometimes the respondent may state an inability to respond.

2.2.4.4 Reaction to Response

Reaction to response is the optional final stage in a negotiation routine when the learner signals (s)he is ready to resume the original conversation (Varonis and Gass 1985). This phase usually takes the form of a short and explicit statement of understanding (e.g. 'OK'; 'Good'; 'I understand'), which are defined as *minimal reactions to response*. In other cases, learners comment explicitly on what was the cause of the problem, which is categorized as a '*metalinguistic*' reaction to response. Smith (2003) adds two new subcategories which emerged

from his study. *Task appropriate responses* (TAR) refers to 'utterances that are contextually relevant to the preceding stretch of discourse and that implicitly show a degree of understanding of the target element' (Smith, 2003, p. 44). Another type of response is *testing deductions* (TD). It occurs when the learner reacts to the recent input provided in the response stage, makes certain inferences, and tests out their current understanding regarding the trigger or the problematic utterances. Test deduction in the RR stage is very similar to the inferential indicators in the second (I) stage of a negotiation routine.

These subcategories are crucial in helping us understand why and how a meaning negotiation routine starts, moves on and ends. Therefore, Table 1 is often used as the theoretical framework and coding scheme for researchers to analyse the features of interactions and explore the mechanism of how meaning negotiation leads to language learning in different technology-mediated contexts (e.g. Smith, 2003; Lee, 2006; Wang, 2006; Yamada and Akahori, 2007; Yanguas, 2010; Jung and Jie, 2012; Guo and Möllering, 2016).

2.3 Meaning negotiation in different SCMC contexts

Section 2.3 will review meaning negotiation studies in three different SCMC contexts, including text-based SCMC, audio SCMC and video SCMC.

2.3.1 Meaning negotiation in text-based SCMC

2.3.1.1 An expanded model for meaning negotiation in text-based SCMC

Since there are so many differences between communication in text-based SCMC and in face-to-face classrooms, it is necessary to examine to what extent the model of meaning negotiation

by Varonis and Gass (1985) is suitable for analysing computer-mediated negotiation. In Smith (2003), fourteen NNS/NNS dyads collaboratively completed four communicative language tasks (two jigsaw and two decision-making tasks). These tasks were 'seeded' with eight target lexical items so that there were opportunities for learners to negotiate meanings with their interlocutors. The study shows that the Varonis and Gass (1985) model of meaning negotiation is largely applicable to text-based SCMC environments. However, Smith (2003) identifies some new patterns in computer negotiated communication between NNSs and therefore proposes an expanded model (Figure 3).

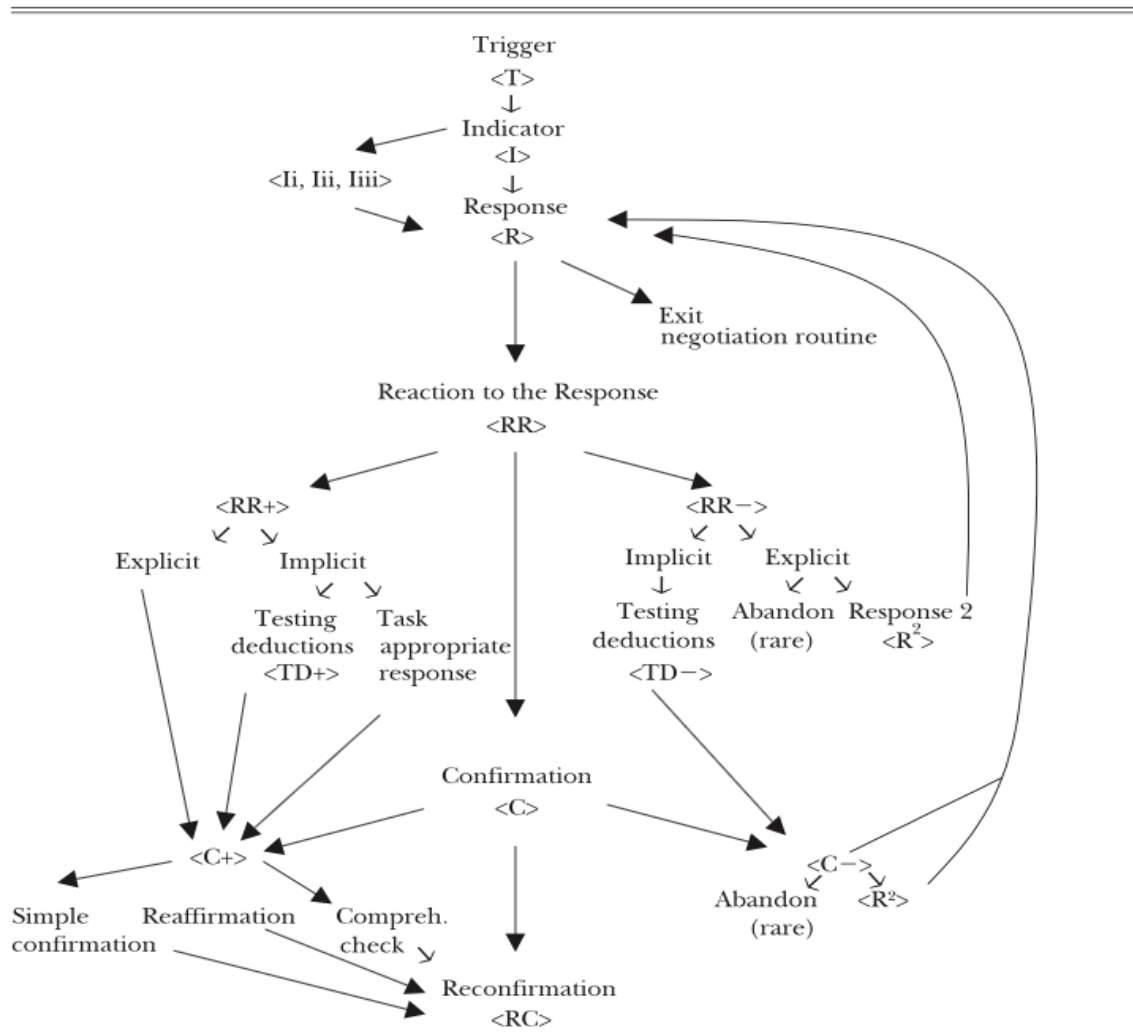


Figure 3: Smith's (2003) Model of Computer-Mediated Negotiated Interaction. Adapted and expanded from Varonis and Gass (1985)

In the expanded model of computer-mediated negotiated interaction (Figure 3), Smith (2003) adds two more possible and optional stages for a negotiation routine: confirmation (C) and reconfirmation (RC). Following the reaction to response, there can be a confirmation phase, where the respondent either confirms (C+) or disconfirms (C-) the degree of understanding by the initiator. Smith (2003) lists three types of reconfirmation, including simple confirmation, reaffirmation (with new information/input) and comprehension check (e.g. 'Got it?'). While in the case of a negative confirmation, the respondent reinitiates the response phase with further input, or in very rare cases, simply abandons the negotiation routine (Smith, 2003). The final

phase in the expanded model is reconfirmation, which usually takes the form of minimal reconfirmation (e.g. 'OK' or 'Yes'), or a simple appreciation (e.g. 'Thanks'). Smith (2003) explains that confirmation and reconfirmation stages occur very often in computer-mediated negotiation interactions due to learners' particular demands for explicit acknowledgements of understanding/non-understanding in text-based SCMC environments. In addition, Smith's (2003) framework also provides many possible routes of meaning negotiation and includes the cases of abandoned or failed meaning negotiation routines, as shown in Figure 3. This model specifically addresses the differences in meaning negotiation caused by different modes of communication and acknowledges a wider range of possible routes and results during negotiation for meaning episodes.

Smith (2003) justifies the need for an adapted and expanded model of meaning negotiation by comparing the differences between the modes of communication in both text-based SCMC and classroom-based face-to-face interactions. First, learners in text-based SCMC tend to make use of 'simplified registers, such as using shorter sentences, abbreviations, simplified syntax, the acceptance of errors, and the use of symbols and emoticons to express emotion' (Smith, 2003, p. 39). Another feature of text-based SCMC is that it contains more overlaps in turn-taking than face-to-face communication. Smith (2003) stresses that the benefits of using SCMC for language learning lie in the written nature of text-chat interactions, which allows more time for learners to attend to and reflect upon the form and content of the message while reading and typing. Moreover, since learners cannot see or hear each other in the text-based SCMC environment, they feel more obliged to express themselves more explicitly than in face-to-face communication. In this way, learners can elicit modified input from one another and are pushed to modify their own linguistic output and receive important feedback on their target language use. As a result, such negotiated interactions through text-based SCMC can be beneficial to language acquisition (Smith, 2003).

2.3.1.2 'Affordances' of text-based SCMC and its effects on meaning negotiation

As Stockwell (2010) and Hampel and Stickler (2012) stress, the 'affordances' of the technology play an important role in how learners communicate and learn languages in the mediated environment. Text-based SCMC has specific features that shape negotiated interactions and affect online language learning in particular ways. Chun (2003) argues that text-based SCMC is unique because it combines features of both speech and writing. On the one hand, as real-time interaction, it resembles oral communication in requiring students to respond to their interlocutors promptly, with neither as long a delay as in asynchronous email conversations, nor as rapidly as in face-to-face oral interactions (Smith, 2003; Shekary and Tahririan, 2006). On the other hand, characteristics resembling written communication include the reading and writing exchange between interlocutors, the use of punctuation and textual formatting in messages, the permanent record of the interaction, the push for learners to think about language use and to reflect on the message content. However, the reader and the writer in text-based SCMC reverse their roles in their mutual construction of their talk-texts (Neuage, 2004).

The hybrid nature of text-based SCMC can affect language learners' meaning negotiation in many ways. Firstly, the lack of verbal communication and visual cues in text-based SCMC (as opposed to face-to-face or video SCMC) can reduce language learners' anxiety, particularly reducing the kind of psychological pressure that is likely to lead to making mistakes (Kern, 1995; Chun, 1998), thus producing more modified input and productive output. Secondly, the anonymity of some text-based SCMC contexts also frees learners from concerns with identity issues such as race, gender, status (Warschauer, 1996). In this context, Van der Zwaard and Bannink (2014, 2016) compared (non)occurrence of meaning negotiation by NS/NNS dyads in both text-based and video SCMC. Their findings show that while learners did indicate non-understanding and negotiated for meaning in text-based SCMC, they did not indicate any non-understanding at all in video SCMC. The authors interpret such non-occurrence as NNSs' deliberate avoidance of losing face when they can be seen by the NS interlocutor. This

particular kind of pressure is minimized in text-based SCMC, so students found it easier in this environment to indicate non-understandings and to negotiate for meaning (Van der Zwaard and Bannink, 2014). Thirdly, real-time communication and reduced anxiety in text-based SCMC also contribute to building a sense of community between peers. The sense of community makes language learners contribute as often and as freely as they like, trying out language structures they would not use in oral interactions (Kelm, 1992). Other researchers have found that the need for students to read the text messages and the slow pace of the real-time interaction in text-based SCMC can promote students' noticing of language errors, and thereby enabling them to focus on form during language-related episodes ²(Lai and Zhao, 2006; Shekary and Tahririan, 2006; Zeng and Takatsuka, 2009; Worajittiphon, 2012).

Despite the benefits of text-based SCMC for meaning negotiation reported above, many researchers comparing text-based SCMC and face-to-face interactions report finding that students negotiate meaning more frequently in face-to-face interactions than in text-based SCMC (Lai and Zhao, 2006; Kaneko, 2009; Yuksel and Inan, 2014; Rouhshad et al., 2016). For example, Yuksel and Inan (2014) use a quantitative approach to compare the effect of communication mode on negotiation of meaning in face-to-face and text-based SCMC communication. Thirty-two dyads of English language learners in a Turkish university participated in both face-to-face and text-based interaction to complete two jigsaw tasks and attended a stimulated recall interview to identify the communication breakdowns during their task interactions in two different modes of communication. Statistical analysis reveals that participants tended to negotiate meaning more frequently in face-to-face interactions

² Language-Related Episodes (LREs) refers to 'any part of a dialogue where the students talk about the language they are producing, question their language use, or correct themselves or others' (Swain and Lapkin 1998, 326). LREs could include corrective feedback provided by peers which can be explicit (if it includes an overt signal that an error has occurred) or implicit (without such a signal). Both explicit and implicit feedback can lead to uptake (error correction) by learners if noticed, understood, and accepted by them. LREs differ from negotiation for meaning in that it focuses on error correction while negotiation for meaning is mainly concerned with resolving non-understandings.

(Mean=10.72) than in text-based SCMC (Mean=9.13). However, participants recalled more instances of communication breakdown in text-based interactions than in face-to-face interactions. Yuksel and Inan (2014) conclude that F2F offers a better context for the production of NfM but that SCMC leads to more instances of noticing. The number of participants (n=64) is large for an SCMC study, which seems to make their statistical findings relatively convincing. However, the study lacks some critical definitions and qualitative analysis. For example, it is not clear what counts as a meaning negotiation episode, whether incomplete or unsuccessful episodes are included, and what are the possible reasons for the statistical findings reported.

Similarly, Rouhshad, Wigglesworth and Storch (2016) also compare the nature and number of negotiations involving F2F and text-based SCMC settings. Twelve dyads of intermediate adult English learners (NNSs) were engaged in two similar decision-making tasks in the two modes of communication. It is found that the face-to-face environment generates significantly more meaning negotiation stances than text-based SCMC. This result concurs with those of Lai and Zhao (2006) and Kaneko (2009). Rouhshad et al. (2016) discuss two possible reasons for this outcome, which are related to the affordances of the two different modes of communication. One reason is that the extra processing time available to students in text-based SCMC allows students to 'scroll up' and re-read the previous messages in the event of communication breakdowns, thus removing the need to negotiate meaning (Smith, 2009; Nik, Adams, and Newton, 2012). The other reason is that some meaning negotiation episodes in face-to-face interactions are triggered by mispronunciation, which is not possible in text-based SCMC. Rouhshad et al. (2016) also report another specific finding about meaning negotiation. They found a lower number of output modifications in text-based SCMC than in F2F, which could be caused by the disrupted turn adjacency in text-based SCMC conversations (Lai, Fei, and Roots, 2008), and the low proficiency of their participants. The above comparative studies demonstrate similar findings, namely that learners negotiate meaning more often in physical face-to-face communication than in text-based SCMC. It is widely agreed among these researchers that the hybrid nature of text-based SCMC is the main cause of such differences.

Section 1.2 has discussed how the specific features of text-based SCMC affect the quantity and the ways of meaning negotiation. It demonstrates that the particular affordances of communication technology can shape and limit learners' interactions in many ways (Stockwell, 2010; Hampel and Stickler, 2012). Next, Section 2 will move on to review the features of voice-based SCMC and their impact on meaning negotiation and language learning.

2.3.2 Meaning negotiation in audio SCMC

2.3.2.1 The affordances of audio SCMC

Audio SCMC here refers to SCMC contexts where oral communication is available and is the main mode of interaction. In some audio SCMC platforms, text-chat communication or other modes such as a whiteboard, graphics, presentation slides may also be available. Terms such as 'audiographic conferencing', 'voice chat', 'voice-based SCMC' have been used to refer to different audio SCMC environments. These are all included in the review in Section 2.3.2. However, the definition of audio SCMC here excludes those showing any interlocutors' video images because video SCMC possesses different affordances from audio SCMC and will be reviewed separately.

In the existing literature on meaning negotiation in audio SCMC, two topics are predominant: the role of pronunciation in shaping meaning negotiation episodes in audio SCMC, and how text-chat is used in audio SCMC. The following review will focus on these two topics.

2.3.2.2 The effects of audio SCMC on pronunciation in meaning negotiation episodes

As emphasised above, the affordances of particular communication technology can shape ways of communication and language learning (Stockwell, 2010; Hampel and Stickler, 2012). Audio SCMC supports real-time oral communication but without any non-verbal cues. Jepson (2005) comments that 'because of the inherent absence of non-verbal communication and the focus ... on pronunciation, voice chat (audio SCMC) may be an optimal environment for pronunciation work' (p. 92). Audio SCMC has been found to enhance students' pronunciation and speaking skills because in audio SCMC students: (1) made improvements in their oral proficiency (Hampel, 2003; Satar and Özdener, 2008) and in pronunciation (Bueno, 2010); (2) become relaxed, engaged and motivated (Hampel, 2003; Rosell-Aguilar, 2005; Satar and Özdener, 2008); (3) are able to self-correct and develop a good perception of social presence when using voice-based communication tools (Yamada, 2009).

Only a few studies examine meaning negotiation routines in audio SCMC. Jepson (2005) compares text-based SCMC and audio SCMC and found more NfM episodes in audio SCMC than in text-based SCMC. The author suggests that audio SCMC offers an environment in which learners are more apt to negotiate for meaning than text-based SCMC. Similar findings have been reported by Bueno (2010). By analysing NNS peer learners' spoken interactions in voice-based SCMC, Bueno (2010) argues that audio SCMC can help develop phonological accuracy as students can notice their pronunciation issues from many communication breakdowns and produce phonetically modified output during meaning negotiation episodes. Bueno (2013) reports that approximately 40% of interactional feedback in LREs is related to pronunciation, making this medium especially beneficial for the acquisition of improved pronunciation. It is emphasised that the use of audio SCMC is beneficial for SLA as it provokes a higher quantity of interaction feedback, more trigger and indicator stances for meaning negotiation, and more phonetically modified output. Another consistent finding by Bueno (2011, 2013) is that, compared to NNS dyads with the same first language, NNS dyads with different first languages come across more phonological triggers, which leads to more instances of meaning negotiation, thereby benefitting the learners' pronunciation. This finding has also been

confirmed by Jung and Jie (2012), who examine meaning negotiation patterns by Chinese and Korean learners of English in video SCMC.

Although it is generally accepted that audio SCMC is real-time or synchronous communication, some researchers point out that voice signal transmissions in audio SCMC are usually subject to a slight yet noticeable time lag, making the production and reception less synchronous than in face-to-face interactions (Blake, 2005; Ciekanski and Chanier, 2008; Kenning, 2010). According to Kenning (2010), such time lags can confuse learners and prompt them to perform channel checks (e.g. 'Can you hear me?'), which interferes with the conversation flow. The time lags, together with the lack of non-verbal cues, also cause turn-taking issues such as overlaps, silences, interruptions, and the use of verbal backchannels. Confusion arising from not knowing when to take their turns properly may reduce learners' opportunities to indicate their non-understanding, as well as initiating meaning negotiation (Kenning, 2010). This situation demonstrates how the affordances of audio SCMC influence meaning negotiation and communication in technology-mediated communication environments.

2.3.2.3 The use of text-chat in meaning negotiation episodes in audio SCMC

Audio SCMC environments usually afford both verbal and textual communication. Researchers have explored how learners make use of the two modes in the bimodal SCMC environment to negotiate meaning (Blake, 2005; Ciekanski and Chanier, 2008; Renner, 2017). This is an initial step towards exploring meaning negotiation in multimodal video SCMC environments. In earlier audio SCMC studies on meaning negotiation, Blake (2005) observes that language learners and teachers make strategic use of the text-chat and the audio channel to communicate effectively. For example, an online tutor may prefer to use text-chat to reinforce the important points that have been expressed in the audio channel, while online learners tend to choose text-chat as a less face-threatening means of requesting linguistic assistance and indicating

non-understanding. Renner (2017) investigates language-related episodes (LREs) in Chinese-German eTandem exchanges in an audio/video SCMC. Students could choose whether to use audio or video SCMC, while text-chat is available with both. Five out of six dyads chose audio SCMC, and one chose video SCMC. In the Chinese-German task-based oral conversations, text-chat was particularly used by participants to negotiate meaning in support of their oral communication in the following circumstances: (1) when there were linguistic comprehension problems, (2) when there were acoustic comprehension problems due to low audio quality, (3) to specify lexical meaning and expressions at the sentence level and (4) when other strategies in the oral mode including repetition, translations and explanations were not successful. For example, when a German student had a linguistic comprehension problem, she asked her Chinese interlocutor to 'write it down!'. The finding indicates that when there is more than one mode available in SCMC (in this case, the audio and text-chat), students are able to choose the mode of communication which facilitates meaning negotiation.

Section 2.3 demonstrates how online learners and teachers combine the two modes (text-chat and audio mode) strategically to negotiate meaning in this bimodal SCMC environment. The next section reviews research into meaning negotiation in multimodal video SCMC environments.

2.3.3 Meaning negotiation in video SCMC

2.3.3.1 The definition and affordances of video SCMC

Video SCMC, also called webcam-mediated SCMC, refers to computer-mediated communication, where, on top of the features of audio SCMC described in Section 2, the video images of the interlocutors in the interaction are transmitted through the webcam. In other words, in video SCMC the interlocutors can see and talk to each other. In most video SCMC contexts, text-chat is also available for communication. Some video SCMC interfaces also

include graphics or presentation slides. All these different modes are available for interlocutors to choose from. One can choose to use a single mode (normally audio) or employ a flexible combination of different modes that are available within the video SCMC environment. This is the multimodal nature of the video SCMC context.

In video SCMC, due to different software interface designs, the size and location of the video image can vary. For example, in Skype video chat, the size of the peer's video image can be adjusted to full-screen size or to any other size or location. But in other SCMC software, the interface may be fixed, as are the size and the location of the peer's video image. This issue has important implications for participants' use of visual resources because the level of detailed visual information one can get from a small corner of the screen is very different from the full screen. However, despite the different sizes and locations of the video image, as long as the video images of interlocutors are on display, this type of SCMC is included in the definition of video SCMC in the current study.

Another important factor concerns the control of the webcam. Some video SCMC software allows students to have full control of their own webcam. If students choose to turn off their webcam, their communication immediately becomes audio-only. In Section 3, the review focuses on pure video SCMC where both interlocutors have their webcam on throughout the whole interaction so that there can be no confusion about the mode of communication and their affordances. However, it is also important to review the studies where students have access to video SCMC use but choose not to employ it. This is because students' choices can reflect their attitudes towards video conferencing SCMC and how students prioritize different modes of communication (Guo and Möllering, 2016).

2.3.3.2 Studies of the effects of video SCMC on negotiation for meaning

This section focuses on the effects of video SCMC on meaning negotiation. Studies included in the review are those which satisfy the following conditions: (a) they focus on patterns of negotiation for meaning; (b) they involve webcam-mediated interaction between interlocutors; and (c) they follow a task-based approach.

Although the current study has a prime focus on meaning negotiation between NNS peers, studies of NS/NNS interactions are also included in the review because the research methods employed and findings relating to the use of visual cues in video SCMC, both contribute to establishing a comprehensive view of meaning negotiation studies in video SCMC. Some of these studies focus on peer interactions between NNSs (Lee, 2006; Yanguas, 2010; Jung and Jie, 2012; Guo and Möllering, 2016), some explore teacher-learner NS/NNS interactions (Wang, 2006) and NS/NNS tandem exchanges (Wang and Tian, 2013; Van der Zwaard and Bannink, 2014, 2016). The review of NS/NNS studies will focus on the task design, research methods, and findings in respect of meaning negotiation patterns in video SCMC.

The review also includes studies that compare meaning negotiation patterns in different modes of SCMC (text-based SCMC, audio-only SCMC, video SCMC), since these studies directly address the different affordances in different modes of communication and their effects on meaning negotiation patterns (Yanguas, 2010; Van der Zwaard and Bannink, 2014, 2016).

Table 2 briefly summarizes eight studies that fall in the scope of the review, including their research questions, participants, tasks design, data collection procedures, framework and data analysis methods, and main findings related to meaning negotiation in video SCMC.

Author	Title	Research Aim	Mode	Participants /Interlocutors	Tasks	Data collection procedure	Framework and data analysis	Relevant findings
Wang (2006)	Negotiation of meaning in desktop videoconferencing-supported distance language learning	To explore the occasions of interactional modifications during video SCMC teacher-learner interactions	Video SCMC with whiteboard	Teacher-learner NS/NNS interaction in Chinese (as the target language for learners); 5 NNS learners, 1 NS teacher	Role-play conversations	Nineteen one-to-one video SCMC sessions.	Varonis and Gass (1985). A qualitative approach to analysing the types and features of meaning negotiation episodes	Video SCMC allows the participants to modify their interaction when there is a breakdown in task completion, thus facilitating L2 acquisition (p. 140). Video conferencing-supported negotiation of meaning has its own distinct features (e.g. supporting visual indicators confirm understanding or non-understanding) in comparison to f2f interaction.
Van der Zwaard and Bannink (2014)	Video call or chat? Negotiation of meaning and issues of face in telecollaboration	To compare how the nature of video and text SCMC affect meaning negotiation.	Comparison between text-based SCMC and video SCMC	8 NS/NNS dyads. Native speakers of English are Australian. NNS are Dutch	A task of cultural humour that would warrant NNS non-understanding	The first half of the task was through text-based SCMC and the second half through video SCMC	Varonis and Gass (1985). A qualitative approach to analysing meaning negotiation episodes in text and video SCMC	NNSs indicate non-understanding during text SCMC but tend to pretend understanding during video SCMC due to 'issues of loss of face' (p. 146). In meaning negotiation episodes, NNSs tend to offer a task-appropriate response in text-based SCMC and face-appropriate response in video SCMC.
Guo and Möllering (2016)	The implementation of task-based teaching in an online Chinese class through web conferencing	To explore how the teacher and learners use multiple modes (video, audio, text-chat, voting, raised-hand function, emoticons, and whiteboard. To examine whether learners engage in negotiation of meaning in the completion of tasks in video SCMC	Video SCMC (including audio, video, whiteboard)	4 NNS Chinese language learner dyads, NNS/NNS peer interaction	An information gap task and a jigsaw task (role-play conversation)	During task instructions, students and teachers use video SCMC. During peer interactions, students can make their own choices. Two online sessions were video recorded, and peer interactions were transcribed.	Varonis and Gass (1985). Mixed method. Quantitatively, the teacher and the learners' multimodal interactions in the main room (in pre- and post-task stages) were quantitatively analysed in terms of speaking dominance, video dominance, use of text-chat and use of emoticons, raised hand and polling functions. Qualitatively, Varonis and Gass' (1985) model was used to analyse different stages of meaning negotiation routines by learners in peer interaction.	Students only use the webcam when the teacher is present and do not use video during peer interactions; instead, they pay more attention to the tasks and finishing their task on the whiteboard. Audio seems to be the most useful and functional feature, while other multimodal functions (video, text) may be perceived by students as 'back-up' while the main audio channel was not available or taken. Students negotiate for meaning in peer audio interaction. Low listening and speaking proficiency is the main trigger for non-understanding. 'Indicator as correction' is identified as a new type of indicator for NfM.
Lee (2006)	Meaning negotiation in task-based video conferencing	To investigate interaction in video SCMC in order to determine its effectiveness as a tool for second language learning.	Video SCMC (with text-chat board)	Eight intermediate level Korean NNS speakers of English. 4 Dyads engaged in 5 online sessions.	Three table-filling Jigsaw tasks about poetry in English	Task interaction video recordings, post-task survey and interviews were collected.	Mixed methods. Quantitatively, the ratio of negotiated turns, the number of different types of meaning negotiation routines, and the number of different types of triggers, indicators and responses were calculated. Smith's model (2003) was used for analysing negotiation routines qualitatively.	Meaning negotiation routines in video SCMC are similar to the model proposed by Smith (2003), as compared to Varonis and Gass (1985). Meaning negotiation in video SCMC occurred mostly due to lexical or content triggers, and sometimes phonological triggers. Learners use more minimal responses than rephrasing or elaborating responses in video SCMC. Meaning negotiation in video SCMC is very similar to face-to-face interactions but less stressful and can provide positive conditions for spoken language acquisition. However, comprehension difficulties and frequent communication breakdowns in video SCMC resulted in excessive use of communication strategies (e.g. inference, summary, repetition, comprehension checks, asking for slower speech, emphasizing keywords, spelling).

Author	Title	Research Aim	Mode	Participants /Interlocutors	Tasks	Data collection procedure	Framework and data analysis	Relevant findings
Yanguas (2010)	Oral computer-mediated interaction between L2 learners: It's about time!	To examine the similarities and differences of meaning negotiation patterns in audio, video SCMC and f2f interactions.	audio SCMC, video SCMC, F2F.	Fifteen dyads of Spanish L2 learners were randomly assigned to three groups: audio SCMC, video SCMC and F2F group.	Jigsaw tasks with potential lexical triggers	Each dyad was allowed 20 minutes to complete the tasks. All task interactions were audio recorded and transcribed for analysis.	Mixed-methods. Quantitatively, the ratio of negotiated turns, negotiation outcome, and numbers of different types of trigger, indicator, response and reaction to response were calculated and compared among three groups. Qualitatively, Varonis and Gass (1985) model was used to analyse meaning negotiation routines in different modes of interaction.	The study stresses that different modes of communication allow for different ways of communicating. Varonis and Gass (1985) still holds in task-based SCMC interactions but does not quite account for many of the interchanges. Negotiations are highly sensitive to tasks. Audio SCMC forces learners to make use of linguistic resources, which can be superseded by visual cues in Video SCMC and FTF groups, but it does not appear to lead to success in the negotiation outcome. No differences are found between video and FTF groups. Furthermore, oral SCMC turn-taking patterns are shown to be very similar to FTF patterns but opposite to those found in written synchronous SCMC. Oral SCMC negotiation patterns are shown to be more versatile than text-based SCMC.
Jung and Jie (2012)	Analysis of meaning negotiation patterns between NNSs in video conferencing	To find out the patterns of meaning negotiation among NNS dyads from the same L1 (first language) and different L1 backgrounds.	Video SCMC	2 Korean and 2 Chinese learners of English.	Information gap tasks on given topics (e.g. unforgettable experience, weddings, advertisement s)	Twelve sessions, six sessions with the same L1 partner, six sessions with different L1 partner. Video SCMC task interactions and post-task interview were recorded and transcribed for analysis.	Mix-method. Quantitatively, the ratio of negotiated turns, different meaning negotiation routines in two group settings, types of triggers, indicators, responses, were calculated. Qualitatively, Smith's (2003) model was used to analyse meaning negotiation patterns.	In the same L1 groups, lexical errors, and content trigger most meaning negotiations while the content and phonological errors trigger negotiation in the different L1 group settings. There is a clear tendency to indicate non-understanding of a lexical trigger through a local indicator in the same L1 group while global non-understanding indicators caused by phonological and content triggers occur more frequently in the different L1 groups. At the response stage, in the same L1 groups, rephrasing and elaboration are the most commonly used strategies to minimise non-understanding whereas in different L1 groups, about half the responses fall into the minimal category. Video SCMC promotes negotiation of meaning and facilitates second language learning in that it helps learners recognise their pronunciation problems and grammatical errors and motivates them to speak English confidently and self-consciously.
Wang and Tian (2013)	Negotiation of meaning in multimodal tandem learning via desktop videoconferencing.	To investigate the quality of meaning negotiation by eTandem partners in video SCMC and the facilitating effects of video SCMC on SLA.	Video SCMC.	Fifteen dyads of Chinese and Australian students. 15 Australian students were divided into three proficiency levels: low, intermediate, and high.	Pair discussion on given topics (e.g. sports, movies in China and Australia).	Each week, for nine weeks, the pairs conducted a one-hour video SCMC exchange, half an hour in English and half an hour in Chinese. One dyad from each of the three proficiency levels was randomly selected for further analysis.	Mixed methods. Quantitatively, different subcategories of trigger, indicator, response, and reaction to response are counted and compared, including the use of visual cues during these stages. Qualitatively, specific meaning negotiation episodes were analysed using Varonis and Gass (1985).	Visual cues, such as nodding, laughing, and a puzzled look are found at indicator, response and reply to response stages, constituting an integral part of the negotiation process. However, students make use of video to varying degrees, with some using a variety of gestures deliberately to generate comprehensible input and output, and others ignoring the video most of the time. Text SCMC is used in indicator, response and reply to response during the video SCMC. Text-chat is mainly used for learning lexical items and confirming understanding.
Van der Zwaard and Bannink (2016)	Non-occurrence of negotiation of meaning in task-based SCMC	To investigate and compare the (non)occurrences of meaning negotiation between NS/NNS dyads in video and text-based SCMC.	video SCMC and text SCMC	Sixteen dyads of NSs and NNSs of English. All NNSs were advanced level English learners (B2-C1)	Four culturally specific jokes for all dyads and an information gap (decision making) task for six dyads.	Of the cultural joke tasks, two were completed through video SCMC, 2 through text-based SCMC. The analysis only focuses on NNS-initiated negotiation of meaning stances.	Mixed methods. Quantitatively, the number of (non)occurrences of NfM stances in two tasks in both video and text-based SCMC were counted and compared. Qualitatively, students' multimodal performances during video SCMC, and post-task stimulated recall interviews were used to analyse (non)occurrence of NfM.	More than half of the non-understandings in video SCMC and 20% of the non-understandings in text-based SCMC are not resolved because NNSs do not initiate meaning negotiation. Four possible reasons proposed by the authors include informal learning context, task design, mode of SCMC, and interlocutors' identity.

Table 2: A summary of studies on meaning negotiation in video SCMC contexts

2.3.3.3 Research methods used in meaning negotiation studies in video SCMC

Despite the disparities between different research contexts, participants, and specific research questions, the studies listed in Table 2 share many similarities in their research methods including (a) the data collection methods, (b) the type of tasks used for online interactions, (c) the analysis methods and (d) the theoretical meaning negotiation framework used for data analysis. Reviewing the similarities and differences between research methods used in the existing literature in this field will be informative for the current study's research design and for critiques of the current field.

2.3.3.3.1 Data collection methods

The most common method of data collection for video SCMC research is through the screen video recordings of the video SCMC interactions (e.g. Wang, 2006; Wang and Tian, 2013; Guo and Möllering, 2016; Van der Zwaard and Bannink, 2014, 2016). Other researchers use less precise terminology, for example, 'the conversation/interactions were recorded' in order to describe the data collection procedure (Lee, 2006; Yanguas, 2010; Jung and Jie, 2012). These studies do not clearly specify whether the data were audio- or video-recorded and they do not include any visual data in their analysis and findings. Drawing on this lesson, the current study will offer a detailed description of the specific data collection procedures.

On top of recordings, some researchers choose to conduct post-task surveys and interviews (Lee, 2006; Jung and Jie, 2012; Van der Zwaard and Bannink, 2016) about participants' (non-)understandings, as well as their feelings and thoughts, especially during meaning negotiation episodes in their video SCMC interactions. Moreover, Van der Zwaard and Bannink (2016) use a post-task video stimulated recall interview with NNS participants to enable them to elicit further clarification of (non-)understanding at particular points of their video SCMC interaction. The video stimulated recall interview will be further discussed in the Methodology

chapter. In addition, Yanguas (2010) also asks participants to complete a post-task test to examine whether they had learned the correct meaning of the target lexical items.

2.3.3.3.2 Task design

Task design plays an important role in both the (non)occurrence of meaning negotiation episodes and specific meaning negotiation patterns or routines (Smith, 2001, 2003). Researchers have reached a consensus that information gap tasks are particularly useful in eliciting meaning negotiation stances (Doughty and Pica, 1986; Pica et al., 1993; Pellettieri, 2000; Smith, 2003; Kim, 2006; Wang, 2007; Yanguas, 2010; Yilmaz, 2011; Van der Zwaard and Bannink, 2016, Kim, 2017). Although the tasks used in the studies under review are different, they all contain some form of information gap between interlocutors. A specific review of the types of tasks used in these studies and their effects on meaning negotiation is presented in Section 3.4.4.

2.3.3.3.3 Data analysis methods

In video SCMC studies, two main approaches have been adopted for examining negotiated interactions. Some researchers employed a purely qualitative approach to describe and analyse meaning negotiation patterns using the framework by Varonis and Gass (1985) (Wang, 2006; Van der Zwaard and Bannink, 2014). All the other researchers adopt a mixed-methods approach. This means that they not only analyse meaning negotiation routines qualitatively, but also attempt to offer some statistical findings, such as the ratio of negotiated turns to total turns, the number of different meaning negotiation strategies, or the number of different subcategories of each meaning negotiation stages (Lee, 2006; Yanguas, 2010; Jung and Jie, 2012; Wang and Tian, 2013; Van der Zwaard and Bannink, 2016; Guo and Möllering, 2016).

(a) Data analysis: the qualitative approach

Both the qualitative and the quantitative approaches to analysing meaning negotiation have their strengths and limitations. The qualitative approach can provide a detailed description and analysis of individual meaning negotiation episodes, with a strong focus on how different modes of SCMC affects participants' meaning negotiation routines. For example, Wang (2006) analyses all 19 negotiated interactions, with a special focus on how interlocutors make use of visual cues, such as facial expressions and hand gestures, in order to indicate or confirm their (non-)understandings. Also following the interactionist approach (Varonis and Gass, 1985), Van der Zwaard and Bannink (2014) provide a very detailed turn-by-turn analysis of NS/NNS meaning negotiation episodes in both text-based and video SCMC. They demonstrate how NNSs feigned understanding to avoid loss of face, particularly in video SCMC.

The interactionist approach that was developed and originally used by Varonis and Gass (1985) seems to be a natural and suitable way of analysing meaning negotiation patterns, yet this method still has its limitations. On the one hand, the findings of such qualitative studies tend to lack generalisability due to the small number of cases for analysis. The selection of particular cases may potentially lead to partial or biased findings and over-interpretation of particular factors. For example, Van der Zwaard and Bannink (2014) whose analysis is based on their analysis of four specific meaning negotiation episodes, conclude that the non-occurrence of meaning negotiation in video SCMC is mainly caused by face issues, meanwhile ignoring other key factors such as task design and research participants' different linguistic and cultural backgrounds. On the other hand, while both studies (Wang, 2006; Van der Zwaard and Bannink, 2014) attempt to demonstrate interlocutors' use of visual cues and stress the effects of visual mode on meaning negotiation, their analysis remains highly descriptive and unstructured. Neither of the two studies employs an established multimodal analytical framework to analyse students' use of non-verbal cues in meaning negotiation episodes in video SCMC.

(b) Data analysis: The quantitative approach in mixed-method studies

The quantitative approach offers a full picture of meaning negotiation patterns and enables comparisons across different modes of SCMC. Lee (2006), Yanguas (2010), and Jung and Jie

(2012) all calculated the rate of negotiated turns to total turns. Yanguas (2010) even reports this ratio in three modes of communication: video SCMC (47%), audio SCMC (57%) and face-to-face (50%). Lee's (2006) ratio is 47.31% in video SCMC, almost exactly the same as that reported by Yanguas (2010). Jung and Jie (2012) report 8.34% for NNS dyads with a shared L1 and 19.81% from NNS dyads with different first languages, both of which substantially smaller than those found by Lee (2006) and Yanguas (2010). However, as will be explained, the validity of such negotiation ratios is questionable, as is their comparability across different studies. First, none of these studies offers any specific definition of what counts as a 'turn'. For example, if one says something, pauses for a while and speaks again, is that one turn or two turns? How did they deal with overlaps and disruptions, or instances of disturbances caused by technical issues? Is a slight sigh a turn or not? Turns are clear in text-based SCMC, but in audio and video SCMC and F2F interactions, the counting of turns can be a highly subjective process, especially if not clearly defined. As Foster and Ohta (2005) argue, identifying meaning negotiation stages involves subjective interpretation by researchers. Furthermore, what counts as the overall number of turns is not clearly defined. For example, are the greetings and farewells at the beginning and end of the task included? Are the turns unrelated to the task discussion (e.g. dealing with technical issues) counted in the overall number of turns? The ratio of negotiated turns is valid only if these details are clearly articulated; otherwise, researchers need to be critical when interpreting such ratios. Moreover, differences in research designs, tasks, interlocutors and in the affordances of video SCMC systems make reliable comparisons of the prevalence of negotiated turns between studies all but impossible. Another issue relating to the quantitative approach is that the limited number of participants makes statistical findings less trustworthy because exceptional cases can have a disproportionate influence on the overall result.

Other reported figures in these studies include the number of different subcategories of each meaning negotiation stage, and the number of (non)occurrences of meaning negotiation in different modes of SCMC and so on. Although these statistical findings can offer a general view of the meaning negotiation patterns identified in any particular study, researchers need to

interpret the findings critically, taking into account a wide range of factors that may affect the (non)occurrence of routines of meaning negotiation.

(c) A methodological gap

The above review of the main research methods used in existing studies seems to indicate methodological gaps in the research into meaning negotiation patterns in video SCMC. Purely qualitative research is unable to provide a comprehensive account of participants' use of the visual mode, while quantitative approaches need clearer definitions for related notions such as the ratio of negotiated turns. These gaps are valuable and informative for the current study in terms of research design and data analysis methods. It is reasonable to conclude that the findings of these studies offer a substantial contribution to the field of meaning negotiation patterns in video SCMC environments. But the research methods for studying this topic needs further improvements.

2.3.3.4 Findings of meaning negotiation studies in video SCMC

The previous section (3.3) reviewed different data analysis methods used in meaning negotiation studies addressing video SCMC. This section (3.4) moves on to review the findings reported in these studies, focusing on four main aspects: (1) the quantitative findings in respect of meaning negotiation routines in video SCMC, (2) the qualitative findings in relation to specific meaning negotiation routines identified in video SCMC studies, (3) the use of visual resources during meaning negotiation; and (4) the effects of task design on meaning negotiation. These four aspects are reviewed because they are the key findings from the existing literature listed in Table 2.

2.3.3.4.1 Quantitative findings: Meaning negotiation routines in video SCMC

One important finding of those studies involving quantitative analysis is the frequency of different subcategories of each meaning negotiation stage as outlined by Smith (2003), Lee

(2006), Yanguas (2010) and Jung and Jie (2012). Wang and Tian (2013), Van der Zwaard and Bannink (2014, 2016) and Guo and Möllering (2016) are excluded here because NS/NNS meaning negotiation patterns have been shown to differ from those in NNS/NNS interactions. In terms of triggers, despite different research designs and tasks, they are most frequently identified as lexical in all four studies. Lee (2006) and Jung and Jie (2012) also report around 40% of content-related triggers in their task interactions. Local and global indicators are found to be frequently used in different studies. In the response stage, elaboration and minimal responses are mostly reported. Jung and Jie (2012) in their study comparing meaning negotiation between NNS dyads with identical and different L1s find that shared L1 dyads use more elaboration while different L1 dyads often end their meaning negotiation with minimal responses. Yanguas (2010) finds the use of visual cues in more than half of the meaning negotiation stances in video SCMC, demonstrating how learners use visual rather than linguistic resources to resolve communication breakdowns. Among different types of reaction to response, minimal reactions to responses occur most frequently, while deduction testing and other task-related responses are also commonly found at the last stage of meaning negotiation.

Apart from quantitative studies, the findings from mixed-method research on video SCMC will be reviewed in the following section. These may be categorized in terms of three key topics related to meaning negotiation: (1) meaning negotiation routines in video SCMC, as opposed to those in other modes of SCMC; (2) the use of multimodal resources in video SCMC negotiated interactions and their effects on meaning negotiation; (3) the types of tasks and their effects on meaning negotiation routines in video SCMC.

2.3.3.4.2 Qualitative findings: Negotiation for meaning routines

When analysing meaning negotiation routines in video SCMC interactions, most researchers (Wang, 2006; Yanguas, 2010; Wang and Tian, 2013; Van der Zwaard and Bannink, 2014, 2016; Guo and Möllering, 2016) adopt the framework by Varonis and Gas (1985) while others use the extended framework developed by Smith (2003) (Lee, 2006; Jung and Jie, 2012). In general

terms, these studies all report that the meaning negotiation routines examined in them were in line with the frameworks used for analysis, but also report new findings and variations. For example, Yanguas (2010) not only identifies all the stages of meaning negotiation proposed by Varonis and Gass (1985), but also finds that the negotiated interactions tended to contain more speech turns than the four stages in the model. Other researchers have also added new and different stages of meaning negotiation to the Varonis and Gass (1985) framework. The following section will review the findings reported by these studies in relation to each meaning negotiation stage.

In meaning negotiation, the trigger is the source of non-understanding. According to Smith's (2003) summary, triggers can be lexical, syntactic, discursal, or content related. Among the studies reviewed in Section 3, lexical triggers are the most frequently reported type of trigger used in negotiated interactions (Lee, 2006; Yanguas, 2010; Jung and Jie, 2012; Wang and Tian, 2013). As Yanguas (2010) comments, the main reason for the frequency of lexical triggers lies in the task design. If the tasks are lexically seeded, then it is highly likely that students will negotiate the meaning of these lexical items. However, both Wang and Tian (2013) and Jung and Jie (2012) employ open discussion or opinion gap tasks without any embedded lexical items and also identified many lexical triggers. Another source of non-understanding is the limited listening and speaking skills of the interlocutors (Wang, 2006; Wang and Tian, 2013; Guo and Möllering, 2016). Wang (2006) further confirms this finding through the post-task interview. Similarly, Lee (2006) and Jung and Jie (2012) report phonological triggers in negotiated interactions in video SCMC. They propose to add this new category of trigger related to pronunciation or listening skills for future video SCMC studies. This is because the subcategories of each meaning negotiation stage summarized by Smith (2003) relate to written SCMC rather than any form of oral communication (Jung and Jie, 2012; Wang and Tian, 2013). Moreover, the content trigger has also been reported by Lee (2006) and Jung and Jie (2012). This finding strongly suggests that video SCMC interactions are meaning-focused (Brock et al., 1986, as cited in Lee, 2006). Last but not least, Wang (2006) emphasizes that the online learning environment can also trigger non-understanding due to sound and video quality

problems (see also Wang, 2004) and the participants' lack of familiarity with the features of the video SCMC software.

According to Varonis and Gass (1985), indicators are used by the listener to signal non-understanding and initiate a meaning negotiation routine. Yanguas (2010) in his quantitative analysis of meaning negotiation routines finds that global indicators appear most frequently in all three modes of interactions including video SCMC (41%), audio SCMC (39%), and face-to-face interactions (44%). As Yanguas (2010) claims, this finding demonstrates that the turn-taking patterns of meaning negotiation episodes in audio and video SCMC are similar to those in face-to-face negotiated interactions, rather than those found in written SCMC, which contains a large number of local indicators (Smith, 2003). He also notes that indicators are omitted in 31%, 35% and 29% of the overall meaning negotiation cases in video SCMC, audio SCMC and face-to-face interactions respectively. In contrast, Jung and Jie (2012) identify 61.36% of local indications between NNS interlocutors from the same ethnic group. Only in interactions by NNSs from different ethnic groups are global indicators used frequently (51.86%), and these are predominantly in response to phonological triggers. Furthermore, based on the video SCMC data, Wang (2006) and Wang and Tian (2013) describe cases of visual indicators in video SCMC, stressing that visual cues such as hand gestures and facial expressions play an important role in the negotiation of meaning in video SCMC interactions. Therefore, it is recommended that visual indicators should be added to the subcategory of indicators in any model of meaning negotiation. Furthermore, Guo and Möllering (2016) propose another type of indicator, called 'indicator as correction' based on their data in which 'students provide corrections of their peers' mistakes' (p.11).

The response stage in meaning negotiation occurs when the interlocutor resolves the non-understanding. Minimal response, rephrasing and elaboration are the most frequently reported types of response in the studies under review. For example, Lee (2006) in his video SCMC meaning negotiation study, reports 55.62% stances of minimal response and 42.60% of rephrasing and elaboration. According to Jung and Jie (2012), NNSs from the same ethnic

group use mostly rephrasing and elaboration during the response stage while almost half of the responses by NNSs interlocutors from different ethnic groups are minimal. Similarly, in the response stage Wang and Tian (2013) reveal students' lack of conversational strategies such as paraphrasing, expansion, reduction, which is likely to be caused by students' low level of proficiency. For the same reason, Guo and Möllering (2016) only identify one type of response: repetition, which is also a simple and minimal form of response. Among all the studies, Yanguas (2010) observes the largest ratio of elaboration in audio SCMC (95%), as opposed to 42% in video SCMC and 47% in face-to-face negotiated interactions. While in video SCMC and face-to-face communication, the 'use of signs' (visual cues) occurs frequently (55% in video and 45% in face-to-face) during negotiation of meaning. Arguably, therefore, in audio SCMC, the lack of visual cues to assist elaboration during the response stage forces interlocutors to make more use of linguistic resources to resolve non-understanding.

The reaction to response stage is an optional stage in the Varonis and Gass model (1985) and usually takes the form of a minimal confirmation of understanding. Jung and Jie (2012) find 25.92% and 41.58% respectively of the meaning negotiation episodes by NNSs dyads from the same ethnic groups and NNS dyads from different ethnic groups end with a minimal response. They also report another 33.33% and 28.71% of task-related responses instead of reactions to response regarding the meaning of the target lexical item. Wang and Tian (2013) report a 47.3% average of minimal responses. In Yanguas (2010) comparative study, the percentages of minimal reactions to response are 48%, 52% and 50% in video SCMC, audio SCMC and face-to-face communication respectively, showing no significant differences between different modes. Similarly, Lee (2006) reports 55% of minimal reactions to response. However, Wang (2006) emphasizes the importance of this stage and calls for more attention to be paid to it because a large amount of modified output observed in her study demonstrates that the reaction to response can serve as a 'valuable prime when we determine the level of L2 acquisition resulting from an occasion of interactional modification' (p.138). Wang and Tian (2013) further report examples of modified output in arguing this point. Similarly, Lee (2006) observes that in more than half of cases, the reaction to response stage involves an explicit

and positive indication of understanding, thereby demonstrating the successful completion of a meaning negotiation episode. Echoing Wang (2006), Guo and Möllering (2016) also identify examples of output modification during the reaction to response stage. However, a very different finding is claimed by Van der Zwaard and Bannink (2014, 2016) who find meaning negotiation episodes in video SCMC where NNS interlocutors feign understanding by uttering a positive reaction to response for reasons of politeness, as well as to save their own face. One reasonable explanation of this behaviour is that their task design is based on a complicated cultural joke that is hard to explain and understand without the appropriate cultural background knowledge.

One of the key reasons for studying meaning negotiation is to examine whether it can facilitate L2 acquisition, in other words, to what extent Long's Interaction Hypothesis (1996) accurately represents what occurs in different contexts and modes of communication. Therefore, the success or failure of a meaning negotiation episode is important for researchers. As Wang (2006) advocates, researchers should pay more attention to the reaction to response stage as it indicates the success or failure of a meaning negotiation episode. Yanguas (2010) also examines the success of meaning negotiation across different modes and finds that although audio SCMC forces learners to make more use of linguistic resources, the success rate of meaning negotiation shows no substantial difference to those in video SCMC and face-to-face interactions. Therefore, he calls for further research to 'ascertain what scenario (mode of communication) is more beneficial for L2 learning' (p.86).

Another perspective on examining meaning negotiation routines entails verifying whether they are completely in line with Varonis and Gass (1985) framework. For example, Van der Zwaard and Bannink (2014, 2016) compare meaning negotiation routines in both text-based and video SCMC and identify many cases of non-occurrence of meaning negotiation where the NNSs came across a non-understanding but did not indicate it. Even in cases where the NNSs indicated their non-understanding several times, 'the negotiation of meaning sequence was aborted after an average of two indicators of non-understanding of the same trigger' (Van der

Zwaard and Bannink, 2014, p. 145). In contrast, Lee (2006) claims that no indicator was ignored or delayed for acknowledgement and that there were no split meaning negotiation routines such as the ones found in text-based SCMC by Smith (2003).

The two paragraphs above emphasize the importance of examining not only meaning negotiation routines, but also non-occurrence, lack of completion and other outcomes, when researching meaning negotiation in different contexts. Such information can offer important (counter)evidence for the Interaction Hypothesis (Long, 1985) and the relationship between meaning negotiation and L2 acquisition.

2.3.3.4.3 The use of visual/multimodal resources

A shared objective of the studies under review is to explore what role video SCMC plays in meaning negotiation, and more broadly, in video SCMC task interactions. The majority of studies agree that video SCMC is beneficial for the negotiation of meaning and can serve as a suitable online environment for second language acquisition, particularly because of the multimodal nature of the communication media (Wang, 2006; Lee, 2006; Yanguas, 2010; Jung and Jie, 2012; Wang and Tian, 2013; Guo and Möllering, 2016). However, some concerns have been raised regarding the negative effects of video SCMC on meaning negotiation. For example, many researchers (Lee, 2006; Wang and Tian, 2013, Guo and Möllering, 2016) have noted that students' limited use of the webcam and visual cues; also, Van der Zwaard and Bannink (2014, 2016) have consistently identified cases of the absence of meaning negotiation in the video SCMC medium. One of the reasons for negative effects may be students' unfamiliarity with the video software; therefore, training in the use of video SCMC technology could be helpful in solving this problem (Lee, 2006; Wang and Tian, 2013; Guo and Möllering, 2016). The following paragraphs will review the benefits and limitations of video SCMC for negotiating meaning reported in the existing literature and summarize suggestions offered for developing better online learning via video SCMC in the future.

One of the biggest benefits of video SCMC is that as a medium for learning interaction, it offers opportunities for interlocutors to negotiate meaning, offer peer feedback, and produce modified output, all of which can facilitate second language acquisition (Lee, 2006; Wang, 2006; Yanguas, 2010; Jung and Jie, 2012; Wang and Tian, 2013; Guo and Möllering, 2016). These studies have demonstrated that video SCMC resembles face-to-face conversations in many respects including (a) the number of meaning negotiation episodes, (b) the turn-taking patterns, (c) the communication strategies used by the interlocutors, and (d) the presence of phonological triggers.

Moreover, the use of visual cues, the multimodal environment, and the availability of text-chat are all distinctive features of video SCMC that make a difference to meaning negotiation and L2 acquisition. Wang (2006) is one of the first researchers to point out the importance of visual cues in meaning negotiation in video SCMC. Specifically, facial expressions including nodding, laughing, a puzzled look, an expression of confusion, and signs of comprehension and enjoyment are examples of indicating or confirming (non-)understandings, and they 'were often received accurately by the other party' (Wang, 2006, p.137). This finding suggests that paralinguistic cues can reduce misunderstanding and ambiguity in video SCMC and can play a crucial role in indicating (non-)understanding when sound quality issues occur (Bruce, 1996, as cited in Wang, 2006). Similarly, Yanguas (2010) and Wang and Tian (2013) also identify examples of visual indicators and minimal reactions to responses using visual cues, further confirming the critical role of visual cues in meaning negotiation. Wang and Tian (2013) add that visual cues constituted an integral part of the negotiation process' because interlocutors are found to make 'deliberate and effective use of a variety of body gestures to generate comprehensible input and output' through video SCMC (p.52).

Although video SCMC is by definition a multimodal communication environment, multimodality has not gained most SLA researchers' attention except for Wang (2006), Wang and Tian (2013) and Guo and Möllering (2016). Wang (2006) first proposes that the combination of video, text-chat, document sharing, file transfer features of video SCMC helps to 'create an effective

language learning environment conducive to task completion' (p.123). However, the challenge of multimodality has also been identified as knowing when and how to use different combinations of multiple modes to 'overcome misunderstanding and promote quality interaction' (Wang and Tian, 2013, p.53). This is the central question in the field of multimodality studies. To answer this question, Guo and Möllering (2016) present a statistical analysis of speaking dominance, video dominance, use of text-chat and use of emoticons, raised hand and polling functions in video SCMC interactions to demonstrate how participants managed to use multiple modes to interact with each other in collaborative tasks. Examples show that multiple modes can simultaneously reinforce one another. For instance, Guo and Möllering (2016) present a case where students used the 'raised hand' button to indicate the willingness to speak while other students are using the audio channel. Similarly, Wang and Tian (2013) observe that text-chat is mostly used by interlocutors to confirm understanding of lexical terms, which is also found in Hampel and Stickler (2012). Clearly, multimodality is still highly under-researched in meaning negotiation studies in video SCMC, one which deserves a great deal more research attention.

Despite the availability of visual cues and the multimodal feature of video SCMC, interlocutors still demonstrate ambiguity as they strive to express themselves more explicitly and clearly (Lee, 2006). For example, Lee (2006) finds that learners tend to use explicit local indicators rather than global indicators to achieve clarity in their negotiated interactions through video SCMC. Similarly, Wang (2006) and Wang and Tian (2013) also observe explicit statements of non-understanding during the indicator stage, as well as explicit confirmation of understanding in the reaction to response stage. This finding is in line with Smith (2003), who identifies confirmation and reconfirmation stages in meaning negotiation routines as learners attempt to achieve more explicitness during text-based SCMC interactions.

Another limitation of video SCMC for meaning negotiation is that interlocutors seem unable to make full use of the functionalities afforded by the software to facilitate their negotiated interactions (Lee, 2006; Wang and Tian, 2013; Guo and Möllering, 2016). As Hampel (2006,

p.112) claims, 'we cannot expect learners to be aware of the affordances of a new media and to know how to use them constructively'. For example, learners in Lee's (2006) study chose to sacrifice the peer's visual information so that they could type and take notes, read materials on the screen, or look up a word in an online dictionary. The lack of use of visual information adds to learners' listening difficulty and anxiety during video SCMC interactions. This shows students' unfamiliarity with video SCMC software and their difficulties in dealing with multiple modes of communication at the same time. However, Wang and Tian (2013) observe that some students actively and deliberately make use of the webcam to facilitate communication and understanding, although other students seem to almost ignore it. Guo and Möllering (2016) find that the webcam was only used when the teacher was present in the online classroom and was turned off during peer interactions. Their data suggest that audio seemed to be the default mode of communication, while text-chat and video 'may have been perceived by students as back-up when the main audio channel was not available or taken' (Guo and Möllering, 2016, p. 11).

While most researchers have a generally positive attitude towards the role of video SCMC in meaning negotiation, Van der Zwaard and Bannink (2014, 2016) hold a strongly negative opinion on this issue. In their studies Van der Zwaard and Bannink, they consistently observe frequent cases of non-occurrence of meaning negotiation in video SCMC. In contrast, the task interactions in text-based SCMC show many more episodes of meaning negotiation. The comparison seems to suggest that video SCMC 'seemed to pose a threatening and daunting communication environment where issues such as politeness and potential loss of face thwarted successful task completion' as well as meaning negotiation (Van der Zwaard and Bannink, 2014, p. 146). In their post-task interviews (Van der Zwaard and Bannink, 2016), the NNS interlocutors confessed that the video interactions were 'scarier' than the text-based ones. This finding has led the researchers to question the validity of the Interaction Hypothesis by Long (1996). In Van der Zwaard and Bannink (2018, 2019), they further reinforce the absence of meaning negotiation in video SCMC and propose a new model of negotiated interactions that categorises response stages into task appropriate responses (TAR) and face appropriate

responses (FAR). Despite Van der Zwaard and Bannink's consistent and strong critical arguments against the role of video SCMC in meaning negotiation, other researchers all confirm the positive effects of video SCMC for meaning negotiation and L2 acquisition, a conclusion supported by a good number of meaning negotiation episodes (Wang, 2006; Lee, 2006; Yanguas 2010; Jung and Jie, 2012; Wang and Tian, 2013; Guo and Möllering, 2016). To further examine this problem, more research focused on this topic is required.

Like many other educational technologies, video SCMC can be dogged by problems, including sound quality, internet connection and frozen video images. However, these technical issues, as well as the above-mentioned limitations, such as the interlocutor's feeling of ambiguity, the inability to make full use of video SCMC affordances and face issues can be overcome by training participants to use the affordances available in video SCMC tools (Wang, 2006; Lee, 2006; Wang and Tian, 2013; Guo and Möllering, 2016). Guo and Möllering (2016) argue that the issues of multiple tools and multimodality have both positive and negative effects on students' online learning. On the one hand, the visual mode can facilitate understanding during negotiated interactions; on the other hand, the array of tools and modes of communication can lead to cognitive overload, thereby preventing students from concentrating on their linguistic output. It is predicted that an improved level of familiarity with the video SCMC environment may encourage language learners' participation and have positive pedagogical implications.

2.3.3.4.4 The effects of tasks on meaning negotiation

Although this research does not focus primarily on the effects of tasks on meaning negotiation routines in video SCMC, reviewing the relevant literature will greatly contribute to informing the task design for the present study. For example, the review of this topic enables us to (a) consider the different types of tasks used in different SCMC studies, (b) summarize their effects on meaning negotiation routines, and (c) demonstrate how to distinguish between the impact of task types and of modality on meaning negotiation routines in multimodal SCMC interactions. All of these aspects are highly relevant to the research design of the current study.

It is widely agreed that task type plays an important role in negotiation of meaning in any mode of communication (Doughty and Pica, 1986; Pica et al., 1993; Pellettieri, 2000; Smith, 2003; Kim, 2006; Wang, 2007; Yanguas, 2010; Yilmaz, 2011; Van der Zwaard and Bannink, 2016, Kim, 2017). Specifically, many researchers have acknowledged the importance of information gap tasks for meaning negotiation and second language acquisition. In face-to-face communication, Varonis and Gass (1985) originally explore NNS/NNS meaning negotiation routines using a simple information gap task which asks students to 'introduce yourselves and find out about each other' (p. 72). Doughty and Pica (1986) emphasize that 'a task with a requirement for information exchange is crucial to the generation of conversational modification of classroom interaction' (p. 305).

Following this view, many researchers have used a variety of information gap tasks to elicit negotiated interactions in different online SCMC settings. For example, in text-based SCMC, Smith (2001, 2003) manages to use a jigsaw story-telling task and a decision-making task to explore meaning negotiation routines. Smith (2003) himself compares the effects of task type on the extent to which learners engaged in negotiation in text-based SCMC. He concludes that jigsaw tasks may elicit more incidental negotiation, as predicted by Pica et al. (1993), but decision-making tasks yield more negotiation sequences than jigsaw tasks when target lexical items are 'seeded' into them. Also focusing on text-based SCMC, Kim (2006) concludes that learners prefer tasks involving cognitive challenges, such as decision-making tasks. However, when comparing negotiated interactions in face-to-face and text-based SCMC, Kim (2017) finds that 'more closed-type tasks with convergent goals and only one possible outcome', such as spot-the-difference tasks, seem to generate a 'larger degree of negotiation' than tasks with open outcomes, such as decision-making tasks (p. 231), which seems to contradict Smith's (2003) findings about tasks. Moving on to audio SCMC, Hampel (2006) demonstrates a framework for transferring face-to-face classroom task design to online audio SCMC environments. She emphasizes that the task design in online environments should be 'genuinely interaction and student-centred' (p.119) and that the resources, affordances and

constraints of the online environment need to be taken into consideration when designing online tasks.

In video SCMC, researchers mainly use five different types of tasks, ranging from closed gap-filling tasks to more open-ended discussions, to more involved cultural jokes. This section summarizes the type of tasks used in the existing literature and reviews the effects of these tasks on meaning negotiation in video SCMC.

First, a relatively easy type of information gap task is a classic gap-fill, in which each interlocutor has a piece of reading material with different blanks and they have to read it to each other, summarize the information together and fill in the blanks. For example, Lee (2006) uses such jigsaw tasks about poetry in English. Guo and Möllering (2016) also use a similar gap-filling task on the topic of a family tree. Many participants in Lee's study complained that the lexical items in their poetry jigsaw tasks were difficult, which leads to the suggestion that they 'caused active meaning negotiation in this study' (p. 150). However, frequent communication breakdowns in video SCMC interactions led to students' excessive use of communication strategies (such as too much repetition) which slowed down the flow of the conversation and demotivated students from continuing to communicate in the target language. Lee reflects on the task design by confirming Ellis' (2003) argument that task difficulty can affect the amount of meaning negotiation. Moreover, it has been suggested that researchers should use more than one type of task to be able to distinguish whether it is the task type or the video SCMC medium that affects the number of meaning negotiation episodes in video SCMC interactions.

Another frequently used type of relatively closed information gap task is the decision-making task, which is similar to the one used by Smith (2003), where both students have different items, and they need to describe them to their peers and then make choices together to achieve a required goal. Lexical items are easily seeded in this type of task to elicit meaning negotiation by interlocutors. Yanguas (2010) uses this type of task for students to choose items for an outdoor racing competition. He identifies a clear focus on the meaning of the target lexical items,

which indicates that 'the nature of the task appears to be responsible for the type and focus of the negotiations' (p. 86). Echoing Yanguas, Guo and Möllering (2016) also find that students provided corrections of their peer' mistakes during their task completion, which might be helpful to draw their interlocutors' attention to language form (Long and Robinson, 1998).

An even more open type of task is role play, where each student is allocated a particular role and given certain information/instructions. Both interlocutors need to perform a role-play together according to their task instructions. Wang (2006) adopts the role-play task and requires the student participants to act as a job applicant and ask for particular information about the job during the interview. Similarly, in Guo and Möllering's (2016) study, their second task is a role play where three students need to play the part of a sick student going to the doctor and asking for sick leave from the teacher. They conclude that the information gap tasks involving one or two modes of information exchange are effective in eliciting meaning negotiation in audio/video SCMC environments.

The most open-ended type of task is opinion gap tasks. Since students usually have different knowledge, experience and opinions on a common topic, researchers can make use of this natural gap to design such tasks as a relatively open-ended peer discussion. Jung and Jie (2012) and Wang and Tian (2013) both offer students real-life topics, such as movies, sports, unforgettable experience as subjects for an open discussion. In Wang and Tian's (2013) study, Australian students have relatively low proficiency in Chinese while their Chinese interlocutors have relatively high proficiency in English. Opinion gap tasks such as open discussion can offer students flexible spaces to express themselves and produce as much 'pushed output' (O'Rourke, 2005, p.442) as they can, rather than using the same standards for students with different L2 proficiencies, which might demotivate highly proficient students or discourage those with poor L2 proficiency.

Also, in video SCMC interactions, Van der Zwaard and Bannink (2014, 2016, 2018, 2019) use another type of information gap task, culturally specific jokes, as a trigger to examine NS/NNS

meaning negotiation patterns in text-based and video SCMC and to identify the absence of meaning negotiation due to face issues. Based on this task design, they argue against the Varonis and Gass (1985) model 'which presupposes that hearers consistently and explicitly indicate non-understanding after a communication breakdown' (Van der Zwaard and Bannink, 2019, p. 119). A new model is proposed involving communication trajectories during digital task-based interactions, in order to illustrate how interlocutors, move between task appropriate responses (TAR) and face appropriate responses (FAR) during meaning negotiation episodes in video SCMC interactions. Their research identifies the importance of 'face' in video SCMC and finds out that students do not initiate meaning negotiation or do not complete the full routine of meaning negotiation proposed by Varonis and Gass (1985). It is true that interlocutors tend to perform in socially appropriate ways, either in face-to-face communication or in text-based or video SCMC interactions. The contribution of their research primarily lies in pointing out the differences of interlocutors' social pressure and relational management in meaning negotiation episodes in two different technology-mediated contexts: text-based and video SCMC. Informed by their findings, this study will pay attention to the issue of 'face' in video and audio SCMC and discuss the relative findings in the discussion chapter.

However, their research methods and the choice of tasks may be misleading. Telling and understanding cultural jokes is much more challenging for NNS than other types of information gap tasks such as spot-the-difference or decision-making. Moreover, the fact that the NNS knew that the NS was telling them a culturally specific joke puts the former into a very embarrassing situation: people tend to feel obliged to laugh when they know their interlocutor is telling a joke although they may or may not fully or even remotely understand it. Therefore, social pressure to provide face-appropriate responses rather than task-appropriate responses is partly due to the nature of the task. However, this is not a common scenario for meaning negotiation by the majority of language learners. So, their new model can only hold in particular cases where students are under high pressure to perform in a socially appropriate way due to the task design. Another possible reason for the non-occurrence of meaning negotiation in their research (Varonis and Gass, 1985) may be the fact that their study pairs were of native-

speakers with non-native-speaker interlocutors. But since the present study primarily focuses on meaning negotiation between NNS dyads, the effects of interlocutors will not be discussed in further detail.

In summary, despite the different types of information gap tasks, the above review of task design confirms that tasks play an important role in meaning negotiation patterns in video SCMC and that information gap tasks are effective in eliciting negotiated interactions in both face-to-face communication and different modes of online SCMC environments. Properly designed information gap tasks can elicit successful meaning negotiation stances (e.g. Yanguas, 2010; Wang and Tian, 2013; Guo and Möllering, 2016), while tasks that are too difficult and those which expose learners to social pressures can lead to overuse of communication strategies or the non-occurrence of meaning negotiation (e.g. Lee, 2006; Van der Zwaard and Bannink, 2014, 2016). The above review of the effects of task types in video SCMC informs task design for the current study, which will be explained in detail in the methodology chapter.

2.3.4 Conclusion of Section 2.3

Section 2.3 reviewed studies of meaning negotiation in three different SCMC contexts from text-based SCMC, to audio only SCMC, to multimodal video SCMC. The following conclusions can be drawn from the review in Section 2.3:

(1) It is widely agreed that meaning negotiation patterns are different in different modes of communication due to the affordances of the SCMC environment. However, researchers have been using a framework derived from face-to-face interactions (Varonis and Gass, 1985) and text-based SCMC interactions (Smith, 2003) to analyse meaning negotiation routines in audio SCMC and video SCMC, which have completely different affordances. Therefore, a clear research gap has been identified since no meaning negotiation routine framework has been specifically developed for negotiated interactions in audio and video SCMC.

(2) Many researchers studying meaning negotiation in video SCMC have acknowledged the important impact of the visual mode on negotiation routines. However, these studies either rely on quantitative statistical analysis, or interaction analysis, but lack any in-depth analysis of how interlocutors make use of the webcam and the visual mode to facilitate meaning negotiation in a multimodal computer-mediated communication environment. Therefore, methodologically, there is a need for a multimodal approach to analysing interlocutors' meaning negotiation episodes in video SCMC interactions.

Section 2.4 will focus on studies employing multimodal approaches to analysing video SCMC interactions.

2.4 Multimodal studies in video SCMC

2.4.1 Introduction of multimodality

Section 2.3 has reviewed studies of meaning negotiation episodes in text-chat, audio, and video SCMC and has identified a lack of research on the role of the visual mode and on multimodality in meaning negotiation episodes in video SCMC. Methodologically, none of the studies reviewed has used multimodal methods to analyse the negotiated interactions. Therefore, Section 2.4 will review studies that look into the role of the visual mode and multimodality in video SCMC, with a special focus on multimodal analytical frameworks. The concept of multimodality can be viewed as a research topic itself and/or as an approach used to explore other research topics in video SCMC interactions. Although the research topic of the studies reviewed in Section 2.4 might not be closely related to meaning negotiation, they have all employed multimodal analytical methods. The aim of this review is to summarize and synthesize the research methods used in multimodal studies, which can inform the design and the data analysis of the current study. Section 2.4 will first justify the importance of multimodality research, discuss the definitions of key concepts, and then review studies focusing on key

aspects of multimodality (including proxemics, kinesics, gaze, mode switching), followed by a summary of studies using multimodal methods to analyse the above-mentioned aspects of multimodality. Section 2.4 will conclude with a consideration of possible research gaps identified in the course of the review.

2.4.1.1 The importance of multimodality in SCMC

Both in face-to-face communication and video SCMC settings, meaning is not solely communicated verbally, but also through a wide range of resources, such as body posture, gestures, gaze, and facial expressions (McNeil, 1992; Kress and Van Leeuwen, 2001; Norris, 2004; Jewitt, 2015). It has been widely agreed that all these modes play an important role in the meaning-making process (Kress and Van Leeuwen, 1996, 2001; Norris, 2004; Levine and Scollon, 2004; Nelson, 2006; Royce, 2006; Jewitt, 2009; Chapelle, 2009; Dooly and Hauck, 2012; Cohen and Guichon, 2016; Chanier and Lamy, 2017; Lee et al., 2019). However, computer-mediated communication differs from face-to-face communication in that communication is mediated through computers and synchronous web-conferencing technology. Meanings are constructed 'through learners' physical relationship to tools' (e.g. the use of webcam), as well as 'through participants' body language on screen, through learners' engagement with still and moving images, with sounds, and with each other's language outputs' (Chanier and Lamy, 2017, p.431). Crucially, the development of internet and multimedia technology has provided opportunities for people to 'integrate imagery, voice, sound, written text, and other semiotic modes' (Nelson 2006, p.57). Such 'changes in communication modes and conventions' (Royce, 2006:36) create learning opportunities and have 'incited CALL researchers to envisage multimodality in a new light' (Guichon and Cohen, 2016, p.509). The central aim of multimodal studies is to explore this full range of communicative forms/modes and the relationships that are created between them' (Jewitt, 2015, p.69). With the increasing use of technology for online language teaching/learning and telecollaboration, it is crucial for researchers to understand how different multimodal resources are 'orchestrated' in technology-

mediated interactions (Stockwell, 2010; Cohen and Guichon, 2016). This is why Chapelle (2009) has identified multimodality as a defining characteristic of CALL.

2.4.1.2 Definition of modes, modality, and multimodality

Many researchers have sought to define mode, modality, and multimodality (Kress and Van Leeuwen, 2001; Guichon and McLornan, 2008; Guichon and Cohen, 2016; Chanier and Lamy, 2017). Kress and Van Leeuwen's (2010) definition of multimodality has been most widely used. According to their definition, multimodality is 'the use of several semiotic modes in the design of a semiotic product or event, together with the particular way in which these modes are combined – they may, for instance, reinforce each other [...], fulfil complementary roles [...] or be hierarchically ordered' (p. 20).

According to Guichon and Cohen (2016), mode is 'the type of semiotic representation (textual, aural and visual) used to present information'. Modality refers to 'the semiotic realisation of one mode, for instance, the visual modality of videoconferencing is realised through the webcam image' (p.510). Guichon and McLornan (2008) state that multimodality makes sensory information accessible in diverse semiotic modes and offers the opportunity to produce, comprehend and exchange information simultaneously through different channels.

Chanier and Lamy (2017) offer their definition of mode from a pedagogical perspective, defining mode as 'the resources to express meaning' (p.429). They believe 'computer-mediated interactive language learning is carried out through the use of modes, which are accessed and manipulated with tools to carry out certain learning objectives', and 'the integration of these three aspects [modes, tools and learning objectives] of communication makes up modality' (p.429). From this standpoint, 'multimodality is the complex relationship that develops between multiple tools and modes when they are co-deployed in different combinations, in learning situations to work toward particular objectives.' (p. 431).

All three definitions of multimodality emphasise two key elements: (1) the availability of different modes to present a piece of information, and (2) how different modes are used in combination to make meaning. In Chanier and Lamy (2017) and Guichon and Cohen (2016), the use of tools has been identified as a key element in their definition of modality. The main difference between the two is that Chanier and Lamy (2017) stress the element of learning objectives from a pedagogical perspective, while Guichon and Cohen (2016) highlight information exchange from a semiotic perspective. This study adopts Guichon and Cohen's (2016) definition because the aim of the study is to identify the role of the visual mode in multimodal meaning negotiation episodes from a semiotic perspective and the research design is similar to work by Guichon and his colleagues (see reviews in Section 2.4).

2.4.2 Key branches of multimodality studies in video SCMC

Sindoni (2013, 2014) has proposed a multimodal theoretical framework specifically for the study of video SCMC interactions. In this framework, four key elements are identified as being essential in multimodal transcription and analysis: proxemics, kinesics, gaze and eye contact, and mode switching. This framework has been used by researchers in the field as a tool for multimodal analysis (e.g. Satar, 2015, 2016; Cohen and Guichon, 2016). Section 2 uses this framework to categorise existing studies of multimodality in video SCMC, and then presents the findings in these key strands with a particular focus on multimodal research methods. Obviously, all these multimodal aspects are integrated into multimodal interactions and are closely interconnected. The studies below are chosen for review in a particular strand mainly because the primary focus of each section is on a particular mode.

2.4.2.1 Proxemics in video SCMC





In social interactions, especially face-to-face communication, individuals tend to negotiate their relationship with space and with one another in meaningful and culture-bound patterns (Sindoni,

2014). Such specific use of social distance is defined as 'proxemics' (Hall, 1966). In video SCMC, however, due to the mediation of the webcam, social distance is not directly established between interlocutors, but between an interlocutor and his/her webcam. Moreover, interlocutors are able to observe their own video image during a conversation in video SCMC, which is completely different from face-to-face interactions. This affects how interlocutors behave verbally and non-verbally such as in their speech, facial expressions, gestures and so on. Video SCMC interlocutors can change the webcam positioning to adjust the social space and the way they project themselves in front of the webcam. Therefore, the study of proxemics in video SCMC is primarily concerned with the positioning and framing of the image (Codreanu and Celik, 2013; Guichon and Wigham, 2016).

2.4.2.1.1 Categories of framing in video SCMC

Codreanu and Celik (2013) explore how multimodal interactions using webcams affect interactive learning in video SCMC language learning/teaching contexts. Specifically, their descriptive study draws on a variety of theories to investigate aspects of visual communication including framing, the degree of use of the webcam (Develotte, Guichon and Vincent, 2010), spatial context and tutors' and learners' gestures (Cosnier, 2008; Develotte, Guichon and Kern, 2008) in a video SCMC environment. Eight trainee tutors, three experienced online tutors and twenty-two undergraduate students from UC Berkeley participated in video SCMC sessions through VISU, a video conferencing software package for online teaching which affords text-chat, audio, video interactions, and personal notes etc. during online interactions. Different types of framing by online teachers and learners in video SCMC were identified including (a) centred, (b) lateral, (c) shifting, (d) lateral medium, (e) extreme close-up (where only the forehead or half of the face is visible) and (f) satellite close-up during double-tutoring (Table 3). Furthermore, the study identified and established relationships between the use of webcam framing and the online teacher's pedagogical objectives. For example, an experienced teacher intentionally moved his body towards the screen to demonstrate his attentiveness to the student. This example shows how proxemics are integrated into online teachers' behaviours to realise

certain communicative (in this case pedagogical) purposes. Experienced teachers' effective integration of physical actions such as body orientation and exaggerated gestures into pedagogical instructions can create an interactive environment and promote learners' comprehension and knowledge construction. On the contrary, bad framing (e.g. extreme close-up) by the online teacher has a negative mirror effect on the learners' framing. The authors argue that the more communication is synchronous and visual, the higher the quality of online tutoring.

Type of framing	Explanation	Sample screenshot
centred close-up	The interlocutor appears in the middle of the frame	 <p data-bbox="1066 1032 1291 1093">Centered close-up, external webcam</p>
lateral close-up	The interlocutor appears towards one side of the frame.	 <p data-bbox="1066 1323 1291 1384">Lateral close-up focus on learner (body orientation)</p>
shifting close-up	The interlocutor moves his/her body/head restlessly.	 <p data-bbox="1066 1621 1291 1682">Shifting close-up</p>
lateral medium close-up	No explanation provided.	 <p data-bbox="1066 1917 1291 1977">Lateral medium close-up</p>



<p>extreme close-up</p>	<p>The interlocutor does not show full face in the frame.</p>	
<p>Satellite close-up</p>	<p>When two interlocutors show in the same frame, either the right or the left side of the frame is bigger/smaller or higher/lower than the other, creating an imbalanced effect.</p>	

Table 3: Six types of framing of online teachers in video SCMC summarized from Codreanu and Celik (2013)

The study demonstrates different types of framing used in online teaching through video SCMC and how framing is used by online teachers to perform certain pedagogical moves. It would be more helpful if clear definitions and distinctions of different types of framing were presented on top of some screenshot illustrations. The research sheds light on how to enhance the design of interactions mediated by the webcam in video SCMC and reinforces the need to train online teachers and learners to become aware of their video images and to be able to maximise its potential to communicate more effectively in video SCMC.

2.4.2.1.2 The effects of framing in video SCMC

Guichon and Wigham (2016) investigate the meaning-making potential of the webcam in a video SCMC language teaching environment from a semiotic perspective. Specifically, they observe how online language teachers use the affordances of the webcam to perform non-verbal cues that can facilitate communication. The study draws on a theoretical framework from film analysis that identifies three different framings of video SCMC including *champ*, *contre-champ*, and *hors-champ* (Deleuze, 1983). As is presented in Figure 4, *champ* refers to what is visible within the frame of the screen, *contre-champ* is one's own video image, and the *hors-*

champ corresponds to the wider environment of the interaction, including all the elements that remain out of the frame but are still part of the pedagogical interaction (p. 65). Such multiple perspectives allow researchers to study the physical elements of the context beyond 'the screen's edge' (Jones, 2004, p. 24) and examine how the different framing affects the multimodal meaning-making potential (e.g. of gestures) through the webcam.

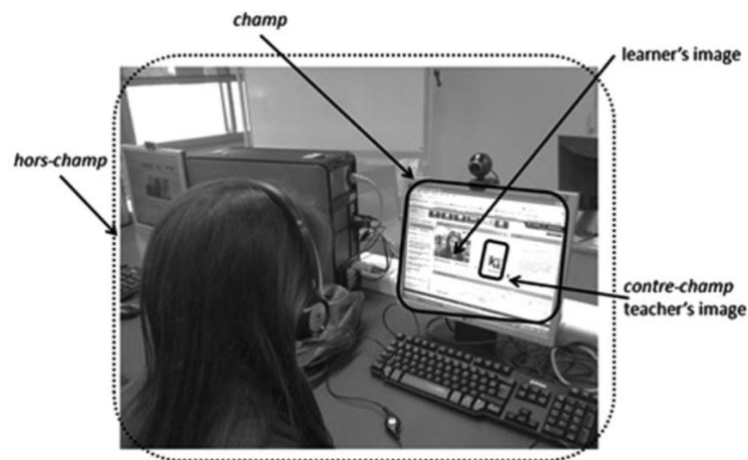


Figure 4: An online pedagogical interaction from different perspectives by Guichon and Wigham (2016, p.64)

Seven trainee language teachers and twelve students were randomly allocated to '1 to 1' or '1 to 2' groups for their online French language classes through VISU (a video conferencing tool). The video recordings from all the three perspectives (*champ*, *centre-champ*, and *hors-champ*) were collected for multimodal analysis. Furthermore, screenshot images of online teachers at a particular moment of each session were taken to analyse their framing choices. Trainee teachers and language learners also attended reflective feedback sessions and interviews in which they were asked about how they made use of the webcam and the video image of the other interlocutor during the online interactions. The screen video recordings from different perspectives were synchronised and transcribed multimodally using the multimodal annotation software ELAN (Figure 5). Six categories including audio act, silence, text-chat act, communicative gestures, movements, and extra-communicative gestures were used (see the methodology chapter for a more specific coding scheme).

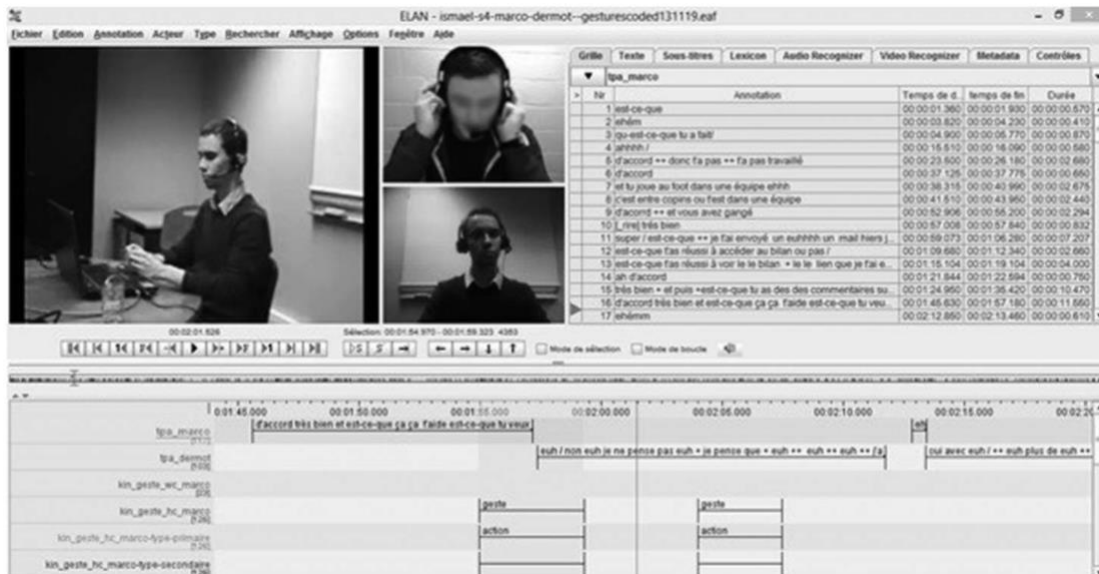


Figure 5: Alignment of trainee hors-champ video with trainee and student webcam videos in ELAN by Guichon and Wigham (2016, p.71)

The screenshots of online teachers' framing were summarised under the following categories: extreme close-up shot, close-up shot, head-and-shoulders shot, and head-and-torso shot, as is defined in Table 4, and demonstrated in Figure 6.

Framing category	Explanation
extreme close-up shot	The trainee's face saturates the frame but is only partially visible. At least one crucial element is missing (eyes, mouth or eyebrows are not visible).
close-up shot	The webcam is focused on the participant's face.
head-and-shoulders shot	The trainee's face and shoulders can be seen.
head-and-torso shot	The trainee's face, shoulders and torso can be seen.

Table 4: Classification of screen shots by Guichon and Wigham (2016, p.70)



Figure 6: Continuum of framing choices at minute seventeen of the interaction (2016, p.72)

The findings of these two studies build on one another. First, different framing choices seem to have their own benefits and limitations. While the close-up shot allows learners to clearly see the teacher's lip movements, which enhances their pronunciation and comprehension, the head-and-shoulder shot, and head-and-torso shot offer better visibility of bodily gestures. On the one hand, the framing choices directly influence the 'shared gestural space', which can make learners feel more at ease, clarify potential misunderstandings, and increase learner concentration. It is also important that the communicative gesture is held long enough to be perceived by learners. However, the comparison between *champ* and *hors-champ* reveals a huge loss of communicative effectiveness due to the reduction of the frame. On the other hand, *contre-champ* seems to promote teachers' critical semiotic awareness by providing them with a clear video image of themselves so that they can adjust their framing accordingly. Guichon and Wigham (2016) conclude that online teachers need to make flexible framing decisions according to 'perceived learning needs, pedagogical intentions, task types, familiarity with the learner and intercultural considerations' (p. 79).

Guichon and Wigham (2016) contribute to multimodal SCMC research by proposing a clear coding scheme for framing, using the innovative framework of *champ*, *contre-champ*, and *hors-champ* to examine the online teaching context beyond the screen, identifying the role of

webcam for online language teaching, and proposing the notion of critical semiotic awareness. Compared to Codreanu and Celik (2013), Guichon and Wigham (2016) further demonstrate how proxemics are used to achieve pedagogical objectives in video SCMC language teaching environments. Although based on a limited number of participants, this study is informative for future researchers and builds the foundation on which more multimodal analysis can follow, for example, focusing on learners' use of the webcam, and on particular episodes that directly relate to language learning, such as meaning negotiation.

2.4.2.2 Kinesics

According to Kendon (2004) and Martinec (2004), in face-to-face communication, kinesics focuses on hand and head movements and gestures. Other researchers also include body movements and other non-verbal behaviour (Boomer and Dittman, 1964; Ekman and Friesen, 1969). Norris (2004) points out that multimodal interaction does not deal with what people think, but with what people communicate through gestures and body movements. In video SCMC context, Sindoni (2014) describes the video image through the webcam as a 'frozen yet living image', which means the webcam is fixed and can only include a partial representation of body (as has been demonstrated by Guichon and Wigham, 2016), meanwhile presenting 'a living representation of the participant's behaviour including speech, kinetic action and postural shifts' (p. 337). Multimodal analysts need to be cautious when analysing the intention of certain behaviours as 'any interaction is culturally, socially, and individually determined' (p.338).

2.4.2.2.1 Hand gestures in meaning negotiation in video SCMC

Lee, Hampel and Kukulska-Hulme (2019) study the gestures in multimodal negotiation of meaning instances via video SCMC through mobile devices. This paper seems to be the only study that closely examines the effect of gestures on meaning negotiation episodes in video SCMC. Ten adult language learners from different countries performed speaking tasks through their mobile devices from a variety of real-world settings, such as their own home, a cafe, art

galleries, or historical buildings. The task has a natural information gap as learners were asked to take their mobile devices, walk around, locate, and share an array of real-world objects through the video SCMC interaction. Video stimulated recall interviews were carried out to gain an in-depth understanding of how learners use gestures to negotiate meaning in the mobile video SCMC context. The data analysis employs a combination of the interactionist approach and the multimodal approach. The verbal element of meaning negotiation episodes was analysed using the interactionist approach based on the model proposed by Varonis and Gass (1985). Meanwhile, learners' gestures during negotiated interactions were coded - using a multimodal analytical framework proposed by McNeill (1992) - into three categories: iconic (representations of an action or object); metaphoric (illustrating an abstract concept); and deictic (pointing gestures at concrete or abstract spaces). Figure 7 presents an example of how the authors combine both frameworks to transcribe and analyse the multimodal interaction. Data from stimulated recall interviews were presented to further demonstrate learners' rationales for their gestures during negotiated interactions.


		FAY	ANDREA	
1. A:	There is a beautiful head . 			DEICTIC (Andrea with mobile device)
2. F	I don't know if you see.			Stroke = (00:06:40)
3. A	I can see it. It's a.... You see it?			ICONIC Fay: (00:06:50)
4. F	[Um] [How do you call it?]			
5. A	I don't know the name... I don't know how what's the name.			
6. F	[Horn] [horn]			ICONIC Fay: (00:6:57)
7. A	[Like the wood]			Andrea: (00:07:01)
8. F	[Horn]			
9. A	[Yeah]			
10. F	[Like eh] I don't know. And how do you say in Eng [ah deer] deer? deer.			ICONIC Fay: (00:07:03)
11. A	Deer?			
12. F	[Head][Head of] [deer]			ICONIC Fay: (00:07:09)
13. A	Ah yeah.			

Figure 7: A sample transcription by Lee et al. (2019, p.32)

It is found that gestures support forms of negotiation by 'affording participants a range of visual and embodied clues, which operate in close conjunction with their language use' (p. 26). However, when the meaning of the gesture does not accurately reflect the meaning conveyed in the verbal conversation, interlocutors may be confused. It is stressed that the negotiated interactions during multimodal video SCMC were significantly impacted by the real-world public settings around the participants. Despite the limitation of having a small number of participants, this study illuminates the multimodal nature of language learners' communication in video SCMC and identifies the role of gestures (especially iconic and deictic) in the establishment of joint attention and negotiation of meaning. The authors encourage further research into the role of gestures in multimodal SCMC environments and suggest that teachers need to train distance language learners to develop multimodal competence to understand and make use of combinations of modes (see also Royce, 2002; Sime, 2008).

2.4.2.2.2 A combination of the multimodal approach and semantic feature analysis for lexical search episodes

Also aiming to examine the role of gestures and the visual mode in L2 learner interactions, Cohen and Wigham (2019) adopt mixed methods to compare word search episodes in both audio and video SCMC contexts. A word search, or a lexical search episode, in this study refers to the cases where a speaker wishes to label a concept but does not have or cannot recall or retrieve the necessary resources to do so (Kasper and Kellerman, 1997, as cited in Cohen and Wigham, 2019, p. 450). L2 learners were asked to explain the expression 'tunnel earring' to a language teacher in both audio and video SCMC settings. The online teacher was instructed to provide minimal feedback to ensure maximum comparability between the audio and video interactions. The interactions were screen video recorded and multimodally transcribed with the ELAN software.

Methodologically, a quantitative approach to semantic feature analysis (Beattie and Shovelton, 1999; Gerwing and Allison, 2011) is used to compare the number of semantic features or referential properties associated with the target lexical item that are mentioned by learners in audio and video interactions. Specific referential properties of the words 'tunnel earring' mentioned by learners are summarised by the researchers and presented in Table 5. Then, a further fine-grained multimodal analysis is carried out to qualitatively demonstrate how gestures are used by L2 learners in video SCMC to convey the meaning of the target lexical item. A sample line-by-line multimodal analysis combined with semantic feature coding is illustrated in Table 6. The quantitative analysis shows that learners used more semantic information in video SCMC with both visual and audio modes than in audio SCMC, although only certain referential properties such as location, position, size, process, and result have been enriched through the video SCMC. The qualitative multimodal analysis of the learners' gestures and the linguistic outputs regarding different referential properties during video SCMC were presented in three phases. In phase one, the learner embodied the salient physical semantic features of the target lexical item with rich and precise iconic and deictic gestures while the linguistic description was

small in quantity and not precise in its meaning. In the second phase, the online teacher reacted with a thinking face and a slight smile, which encouraged the learners to keep on describing the target item. In phase three, the learner managed to produce more turns and a more accurate account of the tunnel earring, while the gestures were less precise than in phase one. It is hypothesised by the authors that the gestures the learner performed in phase one helped him to plan his linguistic output in phase three, thus facilitating L2 production. This answers the question how the webcam or the visual mode can promote learners second language acquisition: through the use of gestures, learners develop and plan their language so that they can produce more and better linguistic outputs, which facilitates SLA.

Referential property type	Referential property	Description
Physical features	Location	The tunnel earring is located on the person's ear.
	Position	The tunnel earring is positioned in the earlobe rather than, for example, in the cartilage at the top of the ear.
	Shape	The tunnel earring is ring-shaped; a cylindrical, hollow piece of jewellery.
	Size	The tunnel earring is unusually big compared to a standard ear piercing.
	Material	The learner suggests that the material from which the tunnel earring is made is wood.
Relating to body modification	Process	A tunnel earring involves stretching the earlobe over time and by gradually upgrading the size of the jewellery, coaxing the hole to fit bigger and bigger tube earrings.
	Result	The large hole left after a tunnel earring is removed is often a permanent body modification.
Other learner strategies	Synonym	The learner uses words having the same or nearly the same meaning. The coders accepted, for example, ear jewel, piercing, earring.
	Comparison	The learners expressed that the item was similar to or different from another item.
	Judgement	The learner expresses a personal reaction to the target item, for example in the verbal mode through phrases such as 'it's pretty impressive' or, in the visual mode, through facial expressions.

Table 5: Referential properties of 'tunnel earring' identified by Cohen and Wigham (2019, p.480)



Line	Webcam images	Verbal	Verbal referential properties	Visual referential properties
1		a earring	Location Synonym	
2		you know I donno if you see what I mean		
3		a earring	Location Synonym	Location
4		that's		Location Shape Size Process

Table 6: Transcription and coding of Phase 1 by Cohen and Wigham (2019, p.464)

Cohen and Wigham (2019) contribute to the field of multimodal research in video SCMC in various ways. First, it employs an innovative research method combining both quantitative and qualitative analysis, meanwhile incorporating the semantic feature approach into the multimodal analysis. Secondly, the results of the qualitative and quantitative analysis reinforce each other, making the argument convincing. Moreover, the mixed-methods analysis demonstrates specifically how the visual mode can promote L2 output in video SCMC interactions. However, the findings might be of limited generalisability since the choice of the target lexical item has particular features that can be easily presented with interpretable gestures in front of the webcam. More multimodal research in video SCMC interactions needs to be done to examine whether the webcam plays similar or different roles for other types of lexical items and other linguistic knowledge.



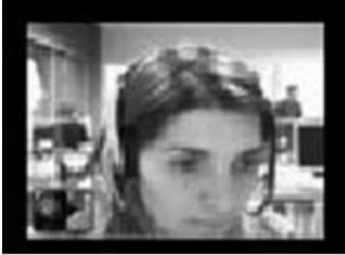
2.4.2.3 Gaze

The role of gaze has been widely studied in many research areas including language learning, psychology, communication studies, etc. Argyle et al. (1973) have summarized the following key functions of gaze in face-to-face interactions: (1) seeking for information and feedback, (2) signalling attitude, (3) controlling the synchronization of speech, and (4) managing/avoiding intimacy. In video SCMC interactions, due to the lack of a shared physical communication environment and the loss of partial body visibility, gaze has become one of the most effective resources for interpreting and making sense of the interlocutor's attitude, stance, and behaviour (Sindoni, 2014). Gaze can be determined by many factors including context, culture, media, interlocutors, tasks and so on. Moreover, in video SCMC, either through a built-in webcam or an external webcam, mutual eye contact is impossible. Therefore, Sindoni (2014) comments that the incidence and the role of gaze in facilitating SCMC interactions 'are not easily gauged by analysts'. The following review presents two studies which identify different types of gaze in video SCMC interactions (Develotte et al., 2010; Satar, 2013).

2.4.2.3.1 Five degrees of use of webcam by online language teachers

Develotte, Guichon and Vincent (2010) explore the use of webcam for teaching a foreign language in a desktop videoconferencing (DVC) environment. The study aims to identify the importance of gaze in webcamming for pedagogical purposes and to analyse the non-verbal dimensions of pedagogical communication via DVC. Five trainee teachers' performances during DVC sessions via Skype were screen video recorded and analysed. Semi-structured interviews were held with online teachers individually to obtain their perceptions of the way they teach online. Due to the technological affordances of the video conferencing interface, online teachers had complete control over the platform and can freely choose to use or not use different tools (text-chat, webcamming) and the different textual and iconic documents available to them. After analysing trainee teachers' use of the webcam, the authors identify five degrees of utilisation of the webcam medium as shown in Table 7 adapted from Develotte et al. (2010).

While mainly concerning online teachers' gaze directions, these degrees also include the use of gestures and facial expressions (see degree 3) as they are an integral part of the multimodal meaning-making process in video SCMC. The findings suggest that the webcam is used more in terms of its availability as a possible resource in case of need than as a favoured type of communication. It seems that webcam images play a complementary role in contributing to the information contained in an oral message and could potentially be distracting. However, when webcam is used, facial expressions (e.g. smile, frown) and gestures (e.g. nod) take on various empathic and interactional functions. The interview data suggest that webcams create presence at a distance, install an obvious connection between the participants, and develop the quality of the pedagogical relationships in a positive way. The authors propose a definition for 'semio-pedagogical' skills as 'the capacity to mediate a pedagogical interaction by combining or dissociating modalities (written, oral, and/or video) that are adapted to the objectives and the cognitive requisites of the task' (p. 293). Online teachers are encouraged to develop their semio-pedagogical competence by 'adjusting the tool to their objectives and the relation they wish to establish with their distant students' (p.310).

Degree 0	The teacher trainee does not appear in the video window, she is standing outside the camera focus or it is not possible to use video.	
Degree 1	The teacher trainee does not look at the computer screen.	
Degree 2	The teacher trainee looks at the open video window on the computer screen.	



Degree 3	The teacher trainee looks at the open video window on the computer screen and uses facial expressions and/or gestures to back up her message.	
Degree 4	The teacher trainee looks straight into the webcam, giving her interlocutor the impression that she is looking directly at her.	

Table 7: Degrees of the utilisation of webcam and screenshot examples in video SCMC summarized from Develotte et al. (2010)

This study contributes to the field of multimodal SCMC research by revealing how multimodal resources, especially the webcam, are used to varying degrees by online teachers for communication and pedagogical purposes in synchronous computer-mediated communication environments. By doing this, the study establishes the close relationship among three key concepts in multimodal online exchanges: proxemics, gaze, and online language teaching pedagogy. The authors' detailed description of the affordances and limitations of the DVC interface as well as the physical setting are crucial in helping readers to understand the effects of webcamming on online language teaching. As is acknowledged by the authors, it is valuable to investigate learners' use of webcam and their attitudes and preferences for future research. Online teachers are suggested to 'develop their semio-pedagogical competence and maximise the advantage of webcamming for online language teaching' (p.310).

2.4.2.3.2 Different types of gaze by online language learners

Satar (2013) reports parts of the findings of her doctoral study which investigates online language learners' social presence in video SCMC from different perspectives (see also Satar, 2015, 2016). Focusing on language learners' gaze in video SCMC, Satar (2013) aims to explore participants' strategies for using the webcam and to identify some patterns or features of gaze in online learner interactions. Five dyads of Turkish undergraduate learners of English participated in 18 DVC sessions to talk about open-ended tasks (see the review of Satar 2015 for more information). Questionnaires, interviews, and stimulated reflections were used to gain a deeper understanding of students' perspectives during the video SCMC learner interactions. Atlas.ti (a qualitative analysis software) was used to transcribe and analyse learners' gaze in video SCMC interaction, and screenshots (sometimes together with a verbatim transcription) were presented to demonstrate the features or patterns of online learners' gaze.

Five different ways in which learners used the webcam were identified including manipulating gaze constantly, strategically, avoiding gaze totally, directing gaze and free gaze (p. 138), as is defined in Table 8. However, the frequency of each type of gaze was not calculated because the study is qualitative in nature and the quality of video and the clarity of gaze were not constant across participants at all times. The thematic analysis of interviews and questionnaires reveals the following findings: (1) mutual gaze is unnatural and perhaps impossible (see also Lamy and Flewitt, 2011); (2) direct attempts at eye contact can be intimidating; (3) gaze on screen might mean an attempt at eye contact; (4) lack of mutual gaze can lead to a decreased sense of trust online; and (5) DVC requires manipulation of one's own image and gaze. The hardware features, for example, whether the webcam is inbuilt or external, can have an important impact on how participants can manipulate gaze. Satar (2013) concludes that the DVC environment, at least at that time, was unable to afford the generation of immediacy due to 'disembodied and limited representation, delays and distortions in audio and video, and the lack of eye contact (p. 139).

Type of Gaze	Definition	Screenshot example
Fixed Gaze	The speaker tries to establish direct eye contact by looking at the webcam at all times.	
Free Gaze	The speaker directs his/her gaze freely and naturally without paying any particular attention to the webcam.	
Strategic Gaze	The speaker tries to establish direct eye contact via the webcam when needed for meaning negotiation.	 
Averted Gaze	The speaker predominantly avoids direct attempts at eye contact via the webcam.	 
Directed Gaze	The speaker directs the gaze of the listener to a particular item by controlling the visual field transmitted via the webcam.	

Table 8: Types of Gaze in DVC summarized from Satar (2013)

The study contributes to the multimodal SCMC research by proposing a framework for analysing gaze in video SCMC environments and illustrating how technological limitations affect learner interactions through the webcam. Yet the frequency of each type of gaze has not been calculated. So, the role of the visual mode has not been quantitatively measured. More research is needed to investigate how paralinguistic cues, such as gaze, are used for learner interactions in video SCMC. Furthermore, since many researchers found a mismatch between the intended and perceived attempt at eye contact via direct gaze into the webcam (e.g. Grayson and Monk, 2003; Norris, 2004; Satar, 2013), it is important for online language learners and teachers to understand the difference between gaze in face-to-face communication and in video SCMC, and develop skills for interpreting mediated eye-contact in video SCMC.

Also focusing on identifying types of gaze in video SCMC, Lamy and Flewitt (2011) analyse video SCMC interactions via eTandem using MSN messenger. Using Scollon and Scollon's (2003) variant of semiotics, they identify four types of gaze: looking at the peer, own image, camera and chat window (as cited in Satar, 2013). Since the original study is written in French, the researcher of the current study is not able to offer further information of this study beyond Satar's (2013) review.

It is clear that these three studies use different methods to classify different types of gaze during video SCMC interactions. Develotte et al. (2010) rank five degrees of using the webcam and their gaze, indicating a hierarchy of competence in multimodal video SCMC. Lamy and Flewitt (2011) categorize gaze according to the part of the video SCMC interface interlocutors focus on, and Satar's (2013) framework of gaze is identified according to the learner's intention.

These three studies have demonstrated that there is no universal framework for analysing multimodal interactions in video SCMC. Researchers need to develop their own approach to multimodal analysis based on their particular research objectives, participants, devices, contexts, and the interface of the video SCMC software.

2.4.3 Multimodal orchestration: Comprehensive multimodal studies

Apart from some specific branches of multimodality, another key question in multimodality research is how people make use of different modes in combination to make meaning, or what are the relationships among different modes in a multimodal communication environment. Kress and Van Leeuwen (2001) believe there can be redundancy or complementarity among different modes. Hampel and Stickler (2012) report from their research in a multimodal online teaching environment that 'different modes can be used in a complementary, compensating and competing manner' (p. 135). Kress and Van Leeuwen (1996) emphasise that visual and verbal media 'are not simply alternative means of representing the same thing' (p.76). As Jewitt (2009) argues, 'the different aspects of meaning are carried in different ways by each of the modes in the ensemble. Any one mode in that ensemble is carrying a part of the message only: each mode is therefore partial in relation to the whole of the meaning' (p.25). Similarly, Guichon and Cohen (2016) stress that it is precisely the enriched interactional opportunities offered by the multimodal nature of technology-mediated environments that are thought to provide enhanced opportunities for second language learning.

Therefore, to investigate language learning and teaching, especially in technology-mediated environments, researchers are encouraged to examine a wide range of modes that make up a pedagogical situation, including all the available semiotic resources and the ways in which they are orchestrated (Kress, 2009; Jewitt, 2009, 2011; Stockwell, 2010, Cohen and Guichon, 2016). This idea requires a multimodal analysis to investigate how online teachers and learners 'make choices among various semiotic options in discursive practices' (Pinnow, 2011, p.384) and assess their meaning-making potential (Jewitt, 2011, as cited in Guichon and Wigham, 2016). For example, Sindoni (2013, 2014) focuses on mode switching between oral and written modes in multiparty video SCMC interactions, demonstrating a competitive relationship between the two modes. Other researchers in different SCMC settings demonstrate how different modes are used complementarily to convey meaning in video SCMC (Satar, 2016, Satar and Wigham, 98

2017; Wigham, 2017). While Satar (2016) presents a highly flexible way of using different combinations of frameworks and produces different types of transcriptions for different analytical approaches, Satar and Wigham (2017) and Wigham (2017) employ a comprehensive multimodal transcription and analysis of a variety of semiotic features including proxemics, gestures, gaze, facial expressions, and the relationships between these modes of communication. These key studies will be closely reviewed in Sections 3.1 to 3.3.

2.4.3.1 Mode switching between spoken and written interaction

Sindoni (2011, 2013, 2014) reports many examples of mode switching between the oral and the text-chat mode during multiparty video SCMC interactions. The words 'mode switching' indicates participants changing from one mode to another, suggesting a competing relationship between different modes. It seems that learners attach more 'authoritative value' to the spoken mode than the written mode (Sindoni, 2014). Sindoni (2014) emphasises the importance of studying mode switching in video SCMC as it is 'one of the most significant facilitators for turn-taking management, as a way of repairing trouble and as a strategy to maintain the flow of conversation across both spoken and written conversational threads' (p. 332).

As has been repeatedly argued, the different ways of communicating through SCMC depend on the affordances and limitations of the mediating technology. In all Sindoni's studies, the multiparty video SCMC software only allows one person to take the audio channel at a time, forcing others to use the alternative of written mode. Moreover, she only focuses on audio and text-chat modes, and appears to ignore the visual mode, which is important in multimodal communication. In addition, her studies examine multiparty communication, which can be very different from one-to-one interactions in video SCMC. Her claims concerning mode switching should not be over generalised to other SCMC settings with different numbers of participants and a different array of modes. But interlocutors' choices among different modes and their changes on the use of modes should definitely receive more research attention, as Sindoni suggests (2013, 2014).

2.4.3.2 A flexible approach to multimodal transcription and analysis

Satar (2016) reports a multimodal analysis exploring meaning-making in online language learner interactions via desktop videoconferencing (DVC) or video SCMC. Here, meaning making refers to a much wider and more general communicative process than negotiation for meaning in SLA. The theoretical framework for the analysis draws on theories from various fields: (1) studies of signs and meaning-making through semiotic systems other than language (Kress and van Leeuwen, 2001; van Lier, 2004; Sindoni, 2013); (2) interactional sociolinguistics, which explores the influence of culture, background assumptions and contextualization cues on the interpretation and negotiation of meaning (Gumperz, 1982, 2003; Andersen, 1998, 2008); (3) multimodal interaction analysis on gestures and body movements in the creation of social identities, relationships and practices (Norris, 2004); (4) conversation analysis including overlaps, backchannels, and silences in turn-taking (Jefferson, 1984; Sacks, 1992; Schegloff, 2000; Tannen, 2005, 2012); and (5) different types of overlaps including transitional, recognitional and progressional overlaps identified by Jefferson (1984) and Schegloff (2000). Five dyads of Turkish undergraduate English learners participated in three or four video SCMC sessions, producing a total of 18 video SCMC sessions. The oral tasks were designed to be open-ended discussions of a particular topic (e.g. family life, music tastes, personalities, ideal room, free time activities, etc.) in each session to stimulate interpersonal interaction. All linguistic data were transcribed, and multimodal elements were directly annotated and coded with ELAN.

The distinctive research method in Satar (2016) is that the author employs different theoretical framework(s) to produce different types of transcription for the corresponding analytical methods. The choice of theoretical framework depends on the nature of the interaction. For example, in Table 9, the long pause in line 28 combined with the facial expressions of both interlocutors indicate a moment of understanding by the interlocutor N, which is why the author chooses to use both the conversation analysis approach to transcription and a verbal description of the interlocutor's facial expressions. While in Table 10, the focus of the interaction

is the student's hand gestures, so screenshots were presented in this particular transcription to demonstrate how learners combine words, hand gestures, and head movements (accompanied by reduced proximity) to make meaning in this multimodal SCMC environment.

21	F: she is my sister /and the other is (.) me (.)\	F points to the other person in the photo. The person is wearing a headscarf.
22	N: /<hi>\	N nods; F removes the photo
23	F: /a:nd [err]\	F shows another photo
24	N: [she is] <older than you?> is [she]	N raising intonation, gaze: right
25	F: [no she's]	
26	N: <older than you?>	N gaze: right; looks at camera when finished
27	F: /<I am older than> her\	N gaze: camera, looks right when finished; F smiling voice
28	N: <ha:>/(1.0)\	N gaze: camera; laughs; intonation surprised; F smiles

Table 9: Transcription with verbal description of multiple modes by Satar (2016, p. 314)



	Verbal	Screenshot
1	H: It's near to the bed	
2	H: from above	
3	D: Near to the bed?	
4	H: Yes, bed.	

Table 10: Transcription with screenshots by Satar (2016, p. 317)

Based on interactional sociolinguistics (Gumperz, 1982, 2003) and Andersen's (1998, 2008) five codes of the body, Satar (2016) provides evidence of how physical appearance, contextualisation cues and shared cultural background influence meaning-making in DVC interactions. The multimodal interaction analysis demonstrates that learners made use of facial expressions and gestures to facilitate their task completion, which resonates with Wang (2007). As for overlaps, conversation analysis (Jefferson, 1984; Sacks, 1992; Schegloff, 2000; Tannen, 2005, 2012) proves to be 'partially applicable and useful' in explaining the overlaps in video SCMC, although 'delays in audio/video transmission seemed to be a major reason for overlaps in desktop videoconferencing' (Satar, 2016, p. 321). Despite certain limitations, the approaches and frameworks from other fields used in this study seem to be suitable methods for investigating meaning making in online multimodal interactions between language learners.

This study contributes to the field of SCMC studies at both the content level and the methodological level. Content-wise, it demonstrates how language learners make meaning in multimodal SCMC interactions; and methodologically, it exemplifies how to incorporate theories, frameworks, and approaches from other research fields flexibly into multimodal SCMC studies and examines their validity or effectiveness in technology-mediated communication environments. Another methodological contribution lies in the flexible and appropriate use of different types of transcriptions, which present various ways of multimodal 'orchestration'. Satar (2016) argues that decisions on the role of transcription in multimodal analysis and the tools used for multimodal transcription 'are closely related to the methodological choices for analysis and thus they should be well informed and carefully considered to suit the aims of the analysis' (p. 321). Future research suggestions include more exploration of the meaning-making process in multicultural SCMC settings and more investigation of the role of a wider range of semiotic resources in making meaning in multimodal video SCMC contexts. Satar (2016) emphasises that CALL researchers need to learn from other research fields and test the applicability and efficiency of their tools and methods in CALL research and develop new approaches for understanding multimodal interactions in video SCMC.

2.4.3.3 A comprehensive approach to multimodal transcription and analysis

Wigham (2017) conducts a multimodal analysis to examine how different semiotic resources are used during lexical explanation sequences between trainee teachers and learners of French in a video SCMC environment. According to the author, lexical explanation episodes, refer to the instances where the trainee teacher asks the students the meaning of a lexical item, or where students indicate a non-understanding of a lexical item. This process is similar to negotiated interactions between a language teacher and a learner. Six video SCMC sessions taught by three trainee teachers and collected for the ISMAEL corpus (Guichon et al., 2014) were selected for multimodal analysis. Lexical explanation episodes were identified and annotated on ELAN with a special focus on five semiotic features including gestures (McNeill, 1992; Kendon, 1982), actions, proxemics, head movements (Altorfer et al., 2000) and changes in gaze, as summarised in Table 11. The findings include that: (1) learners use the visual mode to project active listening strategies; (2) text-chat was used as a complementary tool to secure common ground concerning the target item; (3) verbal explanations were accompanied firstly by deictic and iconic gestures to explain meaning and then by metaphoric gestures to help foreground different properties of the target lexical item; (4) changes in gaze and proximity seem to play a role in managing interaction and signalling which verbal modality was foregrounded. Such findings reveal the complementary relationship between different modes of communication in video SCMC interactions, as was discussed by Kress and Van Leeuwen (2001) and Hampel and Stickler (2012).

Semiotic resources analysed	coding scheme/examples	established framework used
gestures	(1) iconic, (2) metaphoric, (3) deictic, (4) beats (movements to accompany the rhythm	McNeill (1992) Kendon (1982)

	of the discourse), (5) category of emblems (culturally specific gestures).	
actions	(1) writing, (2) typing, etc.	N/A
proxemics	(1) move closer to the webcam, (2) move further away from the webcam.	N/A
head movements	(1) rotational (shaking the head), (2) lateral (tilting the head), (3) sagittal (nodding).	Altorfer et al. (2000)
changes in gaze	(1) towards the screen, (2) towards the keyboard, (3) out of the webcam frame and pensive (out of the webcam frame with gaze directed upwards), (4) eyes closed.	N/A

Table 11: The Multimodal analytical framework used by Wigham (2017)

Wigham (2017) not only demonstrates how online teachers make use of the different semiotic resources available to teach vocabulary in video SCMC, but also exemplifies how to use a multimodal analytical framework with coding schemes for different semiotic features to conduct multimodal analysis for video SCMC interactions. As acknowledged by the author, a limitation of the study is the lack of stimulated recall data. And the employment of eye-tracking data would have added further information as to how participants 'read' multimodal interactions and whether modes are read simultaneously or different channels are foregrounded in reception (O'Rourke et al., 2015; Stickler and Shi, 2015). Wigham (2017) suggests that online language teachers need training to become more aware of the range of semiotic resources available to support different pedagogical practices in video SCMC teaching environments. Further research on the multimodal analysis of interaction in video SCMC is advocated by the author.

Satar and Wigham (2017) use yet another combination of different multimodal theories and analytical frameworks to reveal how online teachers use gaze, webcam proximity, gestures,

word stress and text-chat in different stages of giving online instructions (Kendon, 1982; McNeill, 1992; Norris, 2004; Lamy and Flewitt, 2011; Seedhouse, 2008; Jones, 2012; Satar, 2013; Sindoni, 2013; Jewitt et al., 2016; Guichon and Wigham, 2016). Since the multimodal analysis is similar to Wigham (2017), a specific review is not necessary.

2.4.4 Conclusion of Section 2.4

Section 2.4 focused on reviewing multimodality studies in video SCMC. The review critiqued a variety of multimodal analytical frameworks and offered specific examples of how these frameworks can be used or combined to analyse multimodal interactions in video SCMC. The following conclusions can be drawn from the review. First, multiple modes and semiotic resources are important in meaning making in video SCMC, but the study on this topic is still in its early stages and requires more attention. Second, methodology-wise, there is no universal multimodal analytical framework or well-established research methods for analysing multimodal data, which means that researchers need to make their own framework choices based on their specific research questions, context, the affordances of the technology, participants, task design, etc. (e.g. Satar, 2016; Guichon and Wigham, 2016; Satar and Wigham, 2017; Wigham, 2017; Cohen and Wigham, 2019; Lee, et al., 2019). It is also possible to take an inductive approach and propose new frameworks or theories from the data (e.g. Develotte et al., 2010; Satar, 2013; Codreanu and Celik, 2013). Last but not least, topic-wise, the existing multimodal studies have focused on topics such as multimodal meaning making, enhancing teacher competence, lexical searches, lexical explanations, and instruction giving. Apart from Lee et al. (2019), no other studies have examined how language learners make use of multimodality in meaning negotiation episodes in video SCMC interaction between peers. But even Lee et al. (2019) has a strong focus on gesture rather than offering a comprehensive multimodal analysis of other semiotic resources including gaze, facial expression, proximity, and the relationships among these modes. Therefore, it can be concluded that more research needs to be done to

reveal how learners make use of multiple modes during meaning negotiation episodes in video SCMC interactions.

2.5 Chapter conclusion and research questions

This literature review attempted to review studies at the intersections of three important conceptual fields: meaning negotiation, computer-mediated communication, and multimodality, as is presented in Figure 1 at the beginning of the chapter. The review started by locating the research topic in the existing SLA literature (Section 2.1), then moved on to review the key concepts and theoretical developments of the Interaction Hypothesis and meaning negotiation (Section 2.2). Section 2.3 focused on the intersection between meaning negotiation and different types of SCMC. While Section 2.4 reviewed the intersection between video SCMC and multimodality. Section 2.3 identified the lack of a meaning negotiation routine framework specifically developed for negotiated interactions in audio and video SCMC contexts. Moreover, a methodological gap was demonstrated in the shape of the lack of an in-depth multimodal approach to analysing how learners make use of different semiotic resources to negotiate meaning in video SCMC. While Section 2.4 provided a review of multimodal studies in video SCMC, it also indicated a lack of multimodal studies particularly dealing with meaning negotiation episodes. Therefore, based on the above review, the current study aims to explore the following questions:

- (1) How do students negotiate meaning in audio SCMC and in video SCMC?
- (2) What roles do multiple modes and semiotic resources play in meaning negotiation episodes in video SCMC?

3 Methodology

3.1 Introduction

When it comes to research methodology, researchers have used a range of complicated terminologies to refer to similar or different concepts or even different layers of complex ideas, such as philosophical stance, paradigms, research design, approaches, methodology and methods. Since there is no universal definition of these terms, it is hard for people to clearly understand the relationships among them. Saunders, Lewis, and Thornhill (2019) propose a research onion model (as shown in Figure 8) that clearly demonstrates the relationships among the most important concepts related to research methodology for social science studies. This chapter will generally follow this model because it offers a clear structure for the researcher to present different methodological concepts and justify the choices for the current study. However, this model was originally developed for business studies. Therefore, some adaptations will be implemented to be suitable for this particular study. Specific adaptations will be presented where appropriate. This chapter will be divided into three main parts. After the introduction, Section 3.2 will focus on the philosophical stances, including the ontological, epistemological, and axiological choices. Then the chapter will move on to introduce and justify the research design for the current study, which includes the approaches to theory development and methodological strategies in Section 3.3. And Section 3.4 will conclude the chapter.

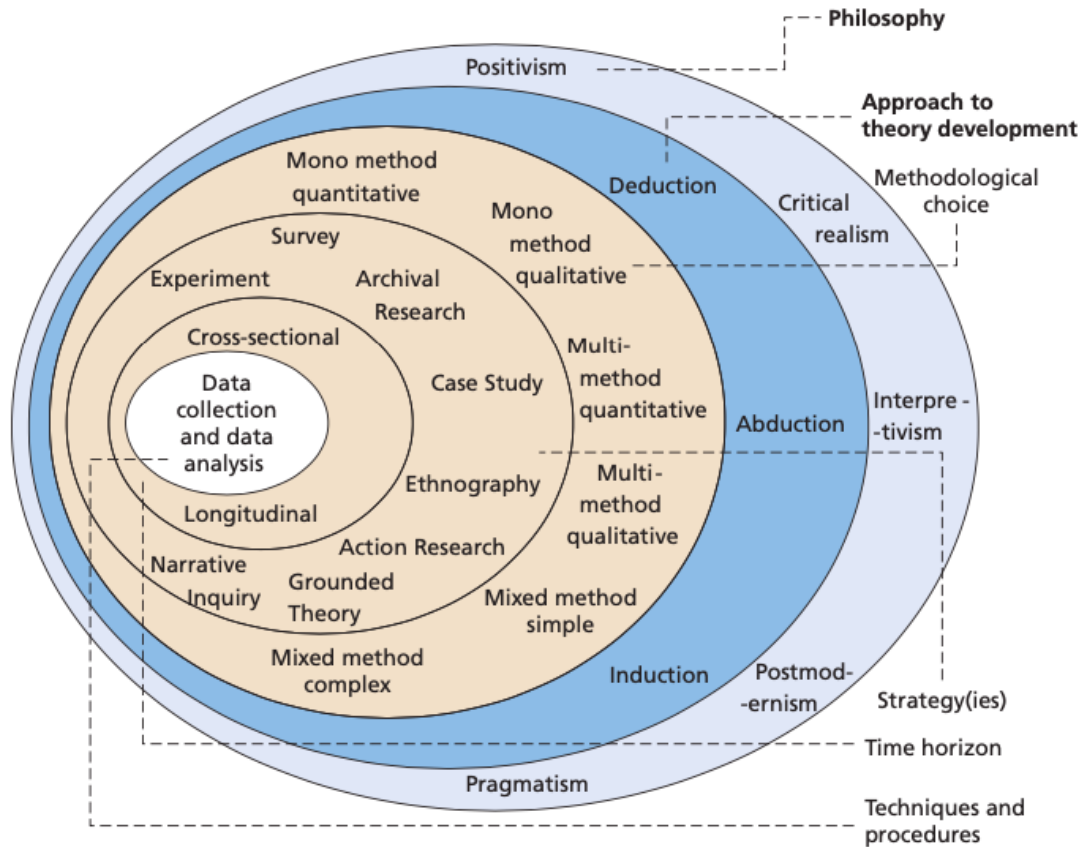


Figure 8: The research onion model by Saunders, Lewis, and Thornhill (2019, p. 5)

All choices concerning methodology at different layers in the research onion should be based on the research questions identified from the literature review, including:

- (1) How do students negotiate meaning in audio SCMC and in video SCMC?
- (2) What roles do multiple modes and semiotic resources play in meaning negotiation episodes in video SCMC?

To clearly present the methodological decisions at different layers, this section first lists all the methodological stances for the current study in Table 12 and then offers a specific explanation and justification. There are a few differences between the structure of this chapter and the research onion model. For example, in the 'philosophy' layer, this study will focus on the philosophical stances that are most closely related to the current study, including positivism and post-positivism, constructivism, and pragmatism. Similarly, in the 'strategy' layer, educational intervention research will be introduced as the main methodological strategy,

although it is not listed in the research onion model. The 'time horizon' topic will be discussed in Chapter 4 as it is highly practical, which is closely related to specific data collection methods and procedures.

Philosophical stance		Pragmatism
Research Design	Approach to theory development	A combination of inductive and deductive reasoning
	Methodological choices	Mixed-methods: with both qualitative and quantitative methods.
	Methodological strategies	Overall: Educational intervention research Multimodal analysis: Case study
Specific Research methods	Data collection methods	1. Screen video recordings of students' audio/video SCMC task interactions 2. Video stimulated recall interviews
	Data analysis methods	1. Interaction analysis of meaning negotiation routines in audio SCMC 2. Gaze analysis of meaning negotiation episodes in video SCMC 3. Multimodal analysis of critical incidents (4 meaning negotiation episodes) in video SCMC

Table 12: Summary of methodological choices for this study

3.2 Philosophical stances

The philosophical stances, which underpin 'research paradigms' are 'a basic set of beliefs that guide action' (Guba, 1990, p. 17). Different terminology has been used by researchers to refer to similar concepts, for example, 'worldview' by Creswell (2014), or 'epistemologies' (about the nature of knowledge and knowing) and 'ontologies' (about the nature of reality) by Crotty (1998). Despite the use of different terminology, it is widely agreed that philosophical stance is important for researchers because it can inform why they choose qualitative, quantitative, or

mixed-method approaches for their research (Creswell, 2014) and 'help both the researcher and the readers of the research to appraise the systematicity and coherence of the research and the plausibility of the conclusions made' (Riazi, 2016, p. 278). Philosophical stances can be shaped by a variety of factors such as different discipline orientations, researchers' inclinations, and the research questions. This study engages with three well established and widely used philosophical stances, including (post)positivism, constructivism, and pragmatism. These philosophical stances entail different ontological and epistemological positions, which can inform different research methodologies and designs.

3.2.1 Positivism and post-positivism

Ontologically, positivists believe in an objective reality that exists 'out there' in the world. In other words, there is one universal reality or truth (Bryman, 2004). This truth does not change in different contexts, so once it is found, it can be generalized. Therefore, the epistemological stance associated with positivism is to use objective measurement to explore universal rules that govern behaviour (Richard, 2003). This philosophical stance supports quantitative research methods, particularly in science subjects such as physics or chemistry. However, when researching social science, the positivist notion of absolute truth is challenged (Phillips and Burbules, 2000). So, postpositivism is proposed, which rejects the concept of absolute truth and focuses on identifying and assessing the causes that influence outcomes. The corresponding epistemology is to develop numeric measures of observations to study the behaviour of individuals with quantitative methods, although being objective is still essential for postpositivists (Creswell, 2014). The axiological stance (the relationship between the research and the researcher) usually taken by positivists and postpositivists tends to be that the researchers' beliefs, assumptions, social, cultural, and educational backgrounds should be eliminated so that the research is as value-free as possible (Maxwell, 2005). Methodologically, postpositivist research usually starts from a theory or a hypothesis and analyses the data to confirm or reject it.

In the current study, students' multimodal behaviours are observed, counted, and calculated in relation to meaning negotiation outcomes (see Chapter 6 for a statistical approach to gaze analysis). Using quantitative analysis to explore the relationship between two factors is in line with the ontological, epistemological, and axiological stances of a postpositivist approach.

3.2.2 Constructivism

Constructivism, or interpretivism³ as has been referred to by Saunders, Lewis and Thornhill (2019) in their research onion model, holds an opposing view to positivism, believing that 'individuals seek understanding of the world in which they live and work' and that 'individuals develop subjective meanings of their experiences— meanings directed toward certain objects or things' (Creswell, 2014, p. 37; see also Berger and Luckmann, 1967; Lincoln and Guba, 1985; Crotty 1998; Mertens, 2010). Therefore, their ontological stance is that multiple realities exist. As explained by Creswell (2014), 'often these subjective meanings are negotiated socially and historically, [...] formed through individuals' interaction with others and through historical and cultural norms that operate in individuals' lives' (p.37).

Epistemologically, constructivists aim to become the insiders of certain cultures. They study, obtain and analyse the views of participants. Meanwhile, constructivists recognize that their interpretation of the data can be shaped by their personal, cultural, and historical experiences. With regard to methodology, constructivists usually take qualitative approaches to their research design and generate a theory from the data, rather than starting with a theory or hypothesis as positivists do (Creswell, 2014). Therefore, in terms of their axiological stance, constructivists believe that researchers are part of what is being studied, and researchers' value-laden interpretation can be a key contribution to the research. Researchers are

³ This thesis uses the term 'constructivism' instead of 'interpretivism' because the former is widely used in the research field of SLA and CALL.

encouraged to be highly reflexive and articulate how the values of the researcher have influenced their research.

Following the constructivist approach, the current study uses interviews to glean participants' attitudes towards meaning negotiation and audio and video SCMC. Interlocutors' answers to these topics may vary depending on their different personal, educational, cultural, and work experiences. All of this information is helpful for gaining a better understanding of their multimodal performances during meaning negotiation in video SCMC (see Chapter 7 for the multimodal analysis).

3.2.3 Pragmatism

While positivism and constructivism are two ends of a spectrum of different philosophical stances regarding researchers' ontological and epistemological stances, another philosophical stance, pragmatism, is not on this spectrum at all because 'pragmatism is not committed to any one system of philosophy and reality' (Creswell, 2014, p.39). Biesta (2010) defines pragmatism as follows:

We should not expect that pragmatism [...] can provide the philosophical framework for mixed methods research. The main reason for this is that, perhaps unlike many other philosophies, pragmatism should not be understood as a philosophical position among others, but rather as a set of philosophical tools that can be used to address problems—not in the least problems created by other philosophical approaches and positions (p. 3).

Central to pragmatism is the idea of focusing on the research questions and taking the combination of pluralistic approaches that are useful in solving the research questions (Patton, 1990; Morgan, 2007; Tashakkori and Teddlie, 2010; Biesta, 2010; Creswell, 2014). Ontologically, pragmatists believe that truth is what works at the time, which could be an

external reality independent of the mind and/or multiple realities within people's minds, 'but they believe that we need to stop asking questions about reality and the laws of nature' (Cherryholmes, 1992, as cited in Creswell, 2014, p.40). However, pragmatists acknowledge that social, historical, political, and other contextual factors can play an important role in research. This philosophical stance offers a philosophical basis for mixed-methods research that combines qualitative and quantitative methods. It allows researchers the freedom to choose the research designs and procedures, data collection, and data analysis techniques. A common criticism of pragmatism is that the methods used are not compatible with each other with regard to their ontological and epistemological assumptions, therefore, the combination does not necessarily make sense. To solve this problem, pragmatists are encouraged to offer a strong rationale for the mixed methods used and to acknowledge the strengths and weaknesses of this philosophical stance (Creswell, 2014).

3.2.4 Rationale for choosing pragmatism

This study takes pragmatism as the philosophical stance which guides the research design, procedures, data collection and analysis because it allows the researcher to make innovative choices of the research methods to answer the research questions.

As has been justified in the literature review, the aim of the current study is to examine meaning negotiation routines in audio and video SCMC interactions and to explore the role of multimodality in video SCMC. The topic of this study lies at the intersection of three research fields, including interaction studies on meaning negotiation, SCMC and multimodality. A wide range of research techniques has been used in these three fields.

First, most interaction studies take a constructivist philosophical stance because interactionists believe that meaning is constructed through human interactions rather than that there is an objective reality independent of human behaviours and contexts (e.g. Varonis and Gass, 1985;

Long, 1996). The task-based language pedagogy frequently used by interaction studies also shares the constructivist assumption that meaning is socially and culturally constructed (e.g. Smith, 2003; Wang, 2006; Jung and Jie, 2012; Wang and Tian, 2013). This is why some researchers prefer to use the term 'negotiation for meaning' rather than negotiation of meaning (Jonassen et al., 1995; Mackey, 2007; Cook, 2015). However, existing literature has already established initial theories such as the meaning negotiation routines by Varonis and Gass (1985) and the interaction hypothesis by Long (1996). What other researchers in this field are mainly engaged in is to verify or falsify the routine or the hypothesis in different contexts. From a methodological point of view, the examination of theories seems to fall towards the positivist end of the spectrum.

Second, for SCMC studies, the central argument is that the ways in which people communicate with each other and make meaning are shaped by the technological affordances and limitations of the communication environment. The stress on contextual factors conforms to constructivism which believes in socially and culturally constructed meaning (e.g. Stockwell, 2010; Hampel and Stickler, 2012). However, many researchers in this field aim to test the effectiveness of online learning through SCMC environments with mixed methods, such as combining interaction analysis with the comparison between pre-test and post-test results (Smith, 2004; see also Stickler and Hampel, 2015). In mixed-methods SCMC studies, the quantitative approach is taken, following an experimental design to make measurements and demonstrate the overall effectiveness of certain SCMC environments, while the in-depth qualitative data can provide specific insights into participants' online learning experiences. Such a mixed-methods approach is supported by a pragmatic philosophical stance.

Third, multimodality studies mainly argue that multimodal and semiotic resources such as gaze, facial expressions, gestures and even the physical environment can all play an important role in the meaning-making process (e.g. Kress and Van Leeuwen, 1996, 2001; Norris, 2004; Royce, 2006; Jewitt, 2009). From this point of view, multimodality has a constructivist philosophical stance as it emphasizes that meaning is socially constructed through multimodal interactions.

As a newly emerging research field, there are no well-established research methods in multimodal studies. Therefore, the richness of multimodal data drives researchers to use innovative combinations of qualitative methods (such as multimodal analysis, gaze analysis, conversation analysis) and quantitative methods (such as eye-tracking, regression analysis) to explore the roles of multimodality in communication. This is the pragmatist strand in multimodal research. Moreover, some multimodalists, such as Perniss (2018), urge that 'it is time to reconceptualize our object of study and to usher in a new paradigm of language theory, a paradigm that focuses on multimodal language, that aligns with the real-world use of language and focuses on doing language' (p.2, see also Andresen, 2014; Kendon, 2014). Similarly, Wildfeuer (2014), from the field of film discourse interpretation study, also proposes the new multimodal paradigm for film analysis.

In summary, all the three key research fields to which the current study belongs share a constructionist philosophical stance in its origin, but many studies in these fields have employed a variety of both qualitative and quantitative methods for their research design, data collection and analysis, that are underpinned by pragmatism. Therefore, to allow more choices of research methods for answering the research questions, this study follows a pragmatic philosophical stance and adopts a mixed-methods approach.

Ontologically, the quantitative analysis follows a postpositivist approach and focuses on the causal relationship of students' multimodal performances and their meaning negotiation outcomes (see Chapter 6). The qualitative analysis of participants' oral interactions (Chapter 5) and multimodal performances (Chapter 7) takes a constructivist approach that believes in the socially constructed nature of reality.

Epistemologically, this study chooses a relatively middle point of the spectrum, neither being a complete outsider, remote from the research nor trying to become an insider and interacting too much with the participants. By choosing a relatively neutral standpoint, the researcher aims to explore participants' natural and undisturbed performances during meaning negotiation

episodes in audio and video SCMC rather than shaping their thoughts and behaviours as a teacher.

As a result, axiologically, the researcher attempts to be as objective and distant from the participants as possible. For example, the researcher invited two other teachers to give task instructions and feedback for all sessions. But some interactions with the participants were necessary. For example, during the task sessions, the researcher was present as an 'assistant' to make screen video recordings and helped with some technical issues. The researcher also conducted one-to-one interviews with participants after all their task sessions. On all these occasions, the researcher tried to be neutral and have as little influence as possible on their performances. For instance, the researcher always sought to ask neutral questions during the interview and avoided unnecessary social activities with all participants before all the data were collected. However, it is still necessary to be reflexive. The researcher acknowledges that the shared linguistic and cultural background (Chinese) of the researcher with participants, her educational background (bachelor's degree in China, master's degree and Ph.D. in the UK), and social experiences (working as an English language teacher in China) can have an influence on participants' performances and the researcher's data interpretation. Such influences are almost unavoidable and can also offer 'a source of insights' (Maxwell, 2005, p.38).

3.3 Research design

3.3.1 Approach to theory development

The approach to theory development, also known as the 'logic of enquiry', is related to the type of reasoning in an argument. The two most basic and commonly used approaches to theory development include the inductive approach and the deductive approach. The inductive approach is a 'bottom-up' process, in which researchers start from data and aim to look for

patterns and rules to generate theory through research (Braun and Clarke, 2006, p.83). This exploratory process is suitable for qualitative research. On the contrary, the deductive approach is a 'top-down' process 'driven by the researcher's theoretical or analytic interest in the area' (Braun and Clarke, 2006, p.86). Therefore, researchers following this approach start from a theory or hypothesis and use data to confirm or reject the theory, typically in quantitative research.

The current study takes a combination of both approaches to theory development. As has been proposed in the literature review, the first research question ('How do students negotiate meaning in audio SCMC and in video SCMC?') aims both to test the existing meaning negotiation routine by Smith (2003) and explore if further patterns occur in synchronous audio and video SCMC contexts. To answer this question, this study uses the deductive approach to test the previous model and generate new patterns from the data for possible expansion or modification of the previous theory. The second research question ('What roles do multiple modes and semiotic resources play in meaning negotiation episodes in video SCMC?') mainly requires an inductive approach since the answer needs to be derived from the analysis of participants' multimodal performances. However, existing studies have some shared agreement on the high importance of gaze in video SCMC (e.g. Develotte, Guichon and Vincent, 2010; Satar, 2010, 2013; Lamy and Flewitt, 2011; Wang and Tian, 2013). This agreement can be tested deductively using the multimodal data obtained in the current study. In summary, a combination of the inductive and deductive approaches will be used in answering both research questions. This is also in line with the philosophical choice of pragmatism, for which mixed methods can be chosen in combination to answer the research questions.

3.3.2 Methodological choices: mixed methods

Following the philosophical choice of pragmatism and the combination of inductive and deductive approaches, this study chooses to use a combination of both qualitative and quantitative research methods for the research design, data collection and analysis.

Although some quantitative elements can be used for triangulation and generalisability purposes, this study still is still highly qualitative in nature. Denzin and Lincoln's definition (2008, p4) is as follows:

Qualitative research is a situated activity that locates the observer in the world. It consists of a set of interpretive, material practices that make the world visible. These practices transform the world. They turn the world into a series of representations, including field notes, interviews, conversations, photographs, recordings, and memos to the self. At this level, qualitative research involves an interpretive, naturalistic approach to the world. This means that qualitative researchers study things in their natural setting, attempting to make sense of, or interpret phenomena in terms of the meanings people bring to them.

The current study conforms to the definition of qualitative research. The key aim of the research is to look into a social phenomenon (meaning negotiation between interlocutors) in its customary setting (audio and video SCMC environments) and use recordings as data for analysis.

However, as Stickler and Hampel (2015, p. 386) point out, online language learning is such a 'multi-faceted and fast-changing activity' that it 'cannot easily be captured with one research approach'. For example, based on the research question about the roles of multiple modes and semiotic resources in meaning negotiation episodes in video SCMC, the data needed to answer such questions are screen video/audio recordings of students' SCMC interactions and most of the data are multimodal. Multimodal data can be highly complex in their nature and can be analysed qualitatively and quantitatively. Multimodal performances in video SCMC can be analysed from a range of different perspectives. For instance, the direction of students' gaze has been found to play an important role in video SCMC interactions (e.g. Develotte, Guichon,

and Vincent, 2010; Lamy and Flewitt, 2011; Wang and Tian, 2013; Satar, 2013, 2016; Wigham, 2017; Lee et al., 2019), but no statistical evidence has been provided to confirm this descriptive claim. A quantitative analysis will be helpful to establish a statistical relationship between participants' gaze directions and their meaning negotiation outcomes. In addition, different modes and semiotic resources, including facial expressions, gaze and gestures can all contribute to the meaning-making process. A qualitative multimodal analysis will be helpful to demonstrate empirically how participants make use of a wide range of modes and semiotic resources to negotiate meaning in video SCMC.

Therefore, this study takes a mixed-methods approach, with qualitative methods as the main approach for data analysis but with some quantitative elements. The choice of this mixed-methods approach can offer deep insights into both the role of multimodality and participants' thoughts during synchronous audio and video SCMC, as well as provide further triangulation and generalisability with the statistical results.

3.3.3 Methodological strategies: educational intervention research

On the whole, this study is a piece of educational intervention research. Participants engaged in an online speaking course as an intervention, which would otherwise not be available to them. Pressley, Graham and Harris (2006) summarise the defining features of this type of research strategy. Educational intervention research is diverse in its theoretical orientations, the complexity of the interventions, research approaches, types of measurements, amount of evidence. The interventions always need to be ethical and the effects of interventions are usually summarised in meta-analyses and other integrative documents.

Methodologically, educational intervention research usually employs an experimental or quasi-experimental method to investigate the effects of the intervention by comparing the controlled

group and the experiment group. This type of research tends to be quantitative in nature. However, Pressley, Graham and Harris (2006) strongly emphasize the importance of qualitative and mixed-methods educational intervention research, as it can 'provide detailed portraits of how an intervention can be delivered, the challenges in doing so and the rich array of outcomes and relationships that might be influenced by the intervention' (p.6). They believe that 'the most informative educational intervention research programmes of the future are going to use multiple research approaches, with analyses complementing one another to provide information about the various aspects and impacts of the intervention' (p. 8). This is also in line with a pragmatic philosophical stance and the mixed-methods approach chosen for this study. The value of such mixed-methods educational intervention research often lies in its contribution to new theories about when and how the intervention works. In other words, this type of educational intervention study does not necessarily evaluate the intervention outcome but observes and explores what occurs during the intervention. Encouraged by this call, the current study takes a mixed-methods approach to exploring how meaning negotiation occurs in audio and video SCMC and how multiple modes are used in their multimodal interactions. This research strategy is suitable for the exploratory nature of the current study.

The diversity of data sources and data analysis methods are extremely important in developing the validity of mixed-methods educational intervention research (Pressley, Graham and Harris, 2006). On the one hand, educational intervention researchers are advised to obtain 'a great deal of available data [...] that are as revealing as possible about an educational intervention and its impacts' (p.13). In the current study, to explore meaning negotiation routines in audio and video SCMC and the role of multimodality in video SCMC negotiated interactions, it would be helpful if data about what students are thinking at certain points of interactions can be collected. On the other hand, multiple types of data analysis methods, such as combining qualitative and quantitative approaches, can 'explicate both the outcomes produced by the interventions as well as the nature of the interventions themselves' (p.13). In this study, qualitative data analysis includes the analysis of students' meaning negotiation routines (see Chapter 5) and the multimodal analysis of their use of multiple resources in video SCMC (see

Chapter 7); while quantitative analysis can shed light on the relationship between gaze directions and meaning negotiation outcomes in video SCMC (see Chapter 6).

On the practical side, it is suggested that educational interventionists should 'think hard about how they can better communicate their work to the practice community, rather than leaving such communications to third parties' (Pressley, Graham and Harris, 2006, p.14). Therefore, in the current study, during the data collection stage, the researcher communicated carefully with the online teachers who oversaw the educational intervention to make sure that they were fully informed of the whole research process and the task instructions they needed to offer to participants (see Chapter 4).

3.4 Chapter conclusion

In conclusion, this chapter presented the philosophical and methodological choices that underpin this research. This study follows a pragmatic philosophical stance as it permits the use of a wide range of methods to answer the two research questions. The study combines both inductive and deductive approaches to examine existing frameworks and explore further developments. Mixed methods are used to analyse the data, including quantitative statistical analysis and qualitative multimodal analysis. This study is a piece of educational intervention research, and the focus is on how the intervention works rather than on examining the effectiveness of the intervention. The specific data collection procedures will be described in Chapter 4.

4 Data Collection Methods and Procedures

Chapter 3 has justified the pragmatist philosophical stance, the combination of inductive and deductive reasoning method, the choice of mixed-method approach and the educational intervention as the main research strategy. This chapter moves on to describe the research design, data collection methods and stages. First, a brief introduction of the specific research context will be presented, including access and ethical issues. This will be followed by a description of the research design and data collection procedures, and an explanation of why some adjustments are made and applied to the main study. Then, the main data collection stages will be described step-by-step in Section 4.3.

4.1 Research context, access, and ethics

4.1.1 The synchronous audio/video SCMC in the research context

All the data in this study were collected from the online education department in a prestigious university in China. The research context was suitable for the current study mainly for the following two reasons. First, the online education department had online provision which pioneered online teaching in China. The department had developed its own synchronous audio/video conferencing platform to provide both blended and completely online teaching. Both teachers and students were familiar with teaching and learning online within this platform. Second, it was not easy to obtain research access in Chinese universities, especially when the study involves an educational intervention. Third, the department had many students who wanted to improve their oral English, which was suitable for this research design. Considering the nature of the study and the practical and technical needs for data collection, this university

department was a suitable and practically feasible choice for the research context of the current study.

Since the affordances and the limitations of the video SCMC platform have important effects on how people communicate with each other and how meaning is negotiated (e.g. Stockwell, 2010; Yanguas, 2010; Hampel and Stickler, 2012; Wang and Tian, 2013), it is necessary to introduce the functionalities of the technical platform. Figure 9 shows the interface of the self-developed video conferencing platform. It consists of presentation slides, online teacher's video image, class attendants' information, students' video images, students' text-chat area and some control buttons. The online teacher has the overall control of the system and can give the access of audio or video SCMC to certain students so that learners can have oral and visual interactions with online teachers. In the video SCMC, the system can support up to four people sharing their webcam video images and using their audio channel at the same time.



Figure 9: The interface of the video conferencing platform

4.1.2 Access and ethics

The HREC approval for this project (HREC/2015/2151/Li) was obtained from the Open University's Human Research Ethics Committee in March 2016. The research context changed from a private online teaching provider to the online education department of a higher education institution in China, and the research topic became more focused on meaning negotiation. These changes were reported to the head of HREC and a further approval was obtained before the pilot study in November 2016. The researcher strictly adhered to the following guidelines throughout the whole research process:

- (a) OU Ethics Principles for Research Involving Human Participants;
- (b) OU Ethics Principles for Research at the Open University;
- (c) The ethical guidelines published by the British Educational Research Association (BERA).

The researcher was given access to collect data for both the pilot study and the main study. All four participants in the pilot study and eight participants in the main study, together with the two online teachers, participated in this research project voluntarily. All participants read the information sheet (see Appendix 7) which contained specific information about the research project and signed a consent form (see Appendix 8) which asked for their permission for the researcher to use their performances in the video/audio SCMC English classes and the recordings of their interviews as data for this Ph.D. research project.

Since this study has a strong focus on how multiple modes are used to negotiate meaning in video SCMC, the analysis would inevitably use the screenshots of participants video images and all multimodal information and semiotic resources need to be included. In the information sheet, participants were informed that their 'performance in video/audio conferencing classes will be video recorded as research data for analysis and presentation'. Participants' privacy was protected with maximum efforts including anonymised participation, safe storage of data at the Open University and participants' rights to withdraw their participation at any time of the study.

4.2 Research design and data collection procedures

4.2.1 Pilot study

A pilot study was carried out prior to the main study data collection to test the feasibility of the data collection procedures. Two dyads participated in the pilot study. Two online teachers from the department facilitated the task sessions by offering them the prescribed task instructions written by the researcher and some feedback to students according to their task performance. The data collection procedures for the pilot study at the online education department of the university included (1) an online induction and pre-task vocabulary test for all four participants; (2) online task sessions by two dyads doing two spot-the-difference tasks and two problem-solving tasks (one in audio and one in video SCMC respectively) and (3) the individual online video stimulated recall interviews through QQ video conferencing. All of these data were collected through screen video recordings. The total amount of these data was 189 minutes (3h 9min) of screen video recordings. Each video stimulated recall interview lasted around 80 minutes, around 325 minutes (5h 25min) in total.

After the pilot data collection and the initial analysis of the results, a number of improvements were proposed for the main study including: (1) adding a mock IELTS speaking test before the task sessions to test participants' oral proficiency; (2) adding two opinion gap tasks (one in audio and one in video SCMC) to familiarise participants with online oral interactions and with their peers before the spot-the-differences and problem-solving tasks; (3) adjusting the task instructions by telling students there is no time limit for doing the tasks; (4) conducting the video stimulated recall interviews in a face-to-face setting as soon as possible after all the task sessions were completed.

The results of the pilot study were helpful for the main data collection in many ways. First, the pilot study confirmed that the research methods were generally suitable and feasible in the context. Second, improvements were summarised and implemented later in the main study. Third, a set of routine (a list of practical things to do step by step) was developed for collecting data from the task sessions and interviews. The following sections (Section 2 to 4) will focus on the main study regarding participants, research design and rationale and data collection procedures.

4.2.2 Participants in the main study

Section 2 introduces the background information of the student participants and the two online teachers.

4.2.2.1 Participants recruitment and pairing

The participants were recruited by the staff at the university's online education department, without an involvement from the researcher. A general introduction of the research project was drafted by the researcher and published in the department student forum as the participant invitation notice a week before the scheduled first session. Participants were selected according to the following criteria:

- 1) the student is willing to participate in the research project by themselves;
- 2) the student has a laptop/desktop computer, an earphone and a microphone;
- 3) once recruited, the student should be committed and able to attend all the sessions and interviews on time;
- 4) students have at least half a year of online learning experience;
- 5) among those who meet all the above requirements, priorities should be given to those with higher English proficiency.

Eight participants were selected according to the criteria above. The participants were ranked according to their scores in the previous term's oral English final exam from the first to the eighth. Since meaning negotiation is more likely to happen between a more proficient speaker and a less proficient speaker (Long, 1996), the participants were paired up in a way to ensure that any two participants in one dyad had as much different levels of proficiency as possible. Specifically, the pairing is as follows: the first and the fifth, the second and the sixth, the third and the seventh, the fourth and the eighth.

4.2.2.2 Participants' proficiency and pairing

All eight participants were female learners aged between 24 to 38. Most of the participants were white-collar workers in full-time employment and do online learning in their spare time. As has been presented above, a mock IELTS speaking test was added in the main study to test participants' proficiency. The speaking test took the form of an online interview through video SCMC. The students' performances were marked by three experienced university teachers according to the official IELTS speaking test marking criteria. The results in Table 13 show that their proficiency level is around B2 according to the Common European Framework Reference (CEFR) criteria. More details of the mock test will be described later in Section 4.2.4.1.2. The mock IELTS speaking scores of each dyad show that the pairings are generally suitable since there was a proficiency difference between the two interlocutors within each dyad, although the differences within each dyad vary from 0.5 (e.g. Dyad 3) to 1.5 points (e.g. Dyad 1). Since the pairing was done according to the participants' speaking exam score from the previous term, the result and ranking from the mock IELTS test did not strictly match those from the previous speaking exam. However, the mock speaking exam before the data collection stage was still useful in a number of ways. On one hand, it offered a general idea of all participants' proficiency, which was informative for the researcher to design tasks and lexical items according to their level. On the other hand, this result could also be helpful in the data analysis and findings stage for the researcher to analyse whether their English language proficiency affects their performances in meaning negotiation episodes.

Dyad	Student A	Student B
Dyad 1	7	5.5
Dyad 2	6.5	5
Dyad 3	5.5	6
Dyad 4	5.5	6.5

Table 13: Participants' scores in the mock IELTS speaking test

4.2.2.3 Online teachers

As has been justified in the Methodology chapter, this study takes a middle way in axiology, hoping to record students' natural and undisturbed performances in video SCMC task interactions. Therefore, two online teachers, Professor W and Lecturer L from the department, were invited to facilitate the task sessions instead of the researcher. This arrangement allows the researcher to be distant from the participants and gives the researcher time to focus on collecting the screen video recording data.

At the start of data collection, specific information about the research was fully explained to the online teachers, such as the theoretical background of the research, the task-based pedagogy, and the specific data collection procedures and schedules. The main role of the two online teachers was to give task instructions and make sure that students clearly understand what to do for the tasks. They also helped students when they came across some technical difficulties. At the end of each task, the teachers offered students some feedback on the outcome of their tasks. The two online teachers both had a postgraduate degree in language teaching from UK universities and had a deep understanding of the task-based language teaching pedagogy. They also both had more than five years of online teaching experience in the university. They were competent online teachers to deliver the task sessions designed for this project.

4.2.3 Research design and rationale in the main study

4.2.3.1 Data collection techniques and rationale

Since the aim of the study is to examine meaning negotiation routines in audio and video SCMC and to explore the role of multimodality in video SCMC, the key data needed is the recordings of participants' meaning negotiation episodes in audio and video SCMC. These data can be collected through participants doing linguistic tasks with embedded target lexical items. This method has been used in almost all meaning negotiation studies in a range of different research contexts, such as Varonis and Gass (1985), Pica et al. (1993), Smith (2001, 2003), Yanguas (2010), Wang and Tian (2013). This study also followed this research design in general. To elicit students' meaning negotiation interactions and make comparisons of audio and video SCMC, the study design involved two sets of similar tasks for students to do through both audio and video SCMC. Their interactions were video-recorded as the key data for analysing meaning negotiation patterns.

In this study, all online teachers and students were requested to speak English only, so that the meaning would be negotiated through various interactions rather than direct translation into Chinese. It should be acknowledged that this requirement would limit the opportunity of exploring how first language (or translanguaging) was used in meaning negotiation by Chinese students, but it was not the central focus of the study.

4.2.3.2 Choice of tasks and rationale

Smith (2001) used two types of information gap tasks in his study about meaning negotiation routines through text-based SCMC: story-telling tasks and problem-solving tasks. This section justifies how this study adapts his task design for the current study.

The first task type used by Smith (2001, 2003) is the story-telling tasks. There were altogether six pictures in the correct order that make up a complete story. Each student was given three pictures (not in the correct order). They needed to describe their own pictures to each other and together come up with a reasonable order of the three pictures for a complete story. The target lexical items were embedded in the six pictures. This task type itself was very hard because it involved describing all the six pictures, negotiating meaning of some new words, and also putting different pictures into the correct order to make up a reasonable story.

In this study, the storytelling tasks were replaced by spot-the-difference tasks. Considering that participants in this study were intermediate level students, and that they were not very familiar with peer interaction through video/audio SCMC, the study made some adaptations to make the task easier to understand and to complete. First of all, the number of pictures was reduced from six pictures (three for each student) to two pictures (one for each student). Then, instead of making a whole story, this study used two pictures similar to each other but with some minor differences. Each student could only see their own picture. In this way, students still needed to describe the picture to each other to find out the differences between their pictures. The spot-the-difference tasks selected consisted of some target lexical items which might elicit meaning negotiation episodes by each dyad. The tasks were readily-designed spot-the-difference tasks taken from an ESL resource website (<http://bogglesworldesl.com/spotthedifferences.htm>).

The target lexical items in the spot-the-different tasks were easier than the ones in the problem-solving tasks. The spot-the-difference task was also easier for students to understand. Therefore, students were asked to do the spot-the-difference task first to become familiar with peer interactions in synchronous audio/video SCMC English classes before they could do the more involved problem-solving task. The specific tasks can be found in Appendix 3 and 4.

As for the problem-solving task, this study followed Smith's (2001, 2003) task design but uses different lexical items. In Smith's (2001, 2003) problem-solving tasks, two students each had four different items, which were target lexical items for meaning negotiation. They only knew

their own items and did not know their peer's items. First of all, the students were asked to describe their own items to each other with the objective of completing their partial knowledge of what the eight items were. Then, they needed to talk to each other and make some choices together for certain purposes according to the task instruction (e.g. for a flea market sale, or as gifts for their home stay family). This was a good task type because it required students to negotiate the meaning of the target lexical items and use the target lexical items in their latter discussion.

In the problem-solving tasks, the specific choice of words is crucial because it can directly affect students' oral and visual performances in meaning negotiation episodes. First of all, these words should be familiar to Chinese adult learners in their daily life so that students can understand the meaning of their own words (items) clearly. Secondly, these words cannot be too hard or abstract so that students are able to explain the meaning to each other using their existing knowledge of English. Thirdly, the words cannot be too easy because if they know the meaning of these words, there will not be any occurrences of meaning negotiation. Finally, some of the words should be explainable with some body gestures so that the researcher can explore how multiple modes and semiotic resources can be used for meaning negotiation in video SCMC. The final problem-solving tasks are presented in Appendix 5 and 6.

4.2.3.3 Video-stimulated recall interviews and rationale

After the peer interactions during the task sessions, video-stimulated recall interviews (VSRI) are designed to check students' (non-)understanding and their thoughts at certain points of meaning negotiation episodes. This section will offer the rationale for the use of video stimulated recall interviews in this study.

According to Gass and Mackey (2017), stimulated recall is an introspective method in which participants are asked to recall thoughts they had while performing a prior task or while they were participating in a prior event. To assist recall of these thought processes, a stimulus is

used, such as a video recording of the activity. The theoretical assumption behind stimulated recall methodology is that some tangible (video or audio) reminder of the event will stimulate recall of the mental processes in operation during the event itself and will, in essence, aid the participant in mental re-engagement with the original event. As Bloom (1953, p. 161) argued, such stimuli may enable a participant 'to relive an original situation with vividness and accuracy'. An important benefit of stimulated recall methodology, like normal interviews, is that it allows researchers to obtain some valuable 'insider' information about participants' mental or cognitive process, which is hard to be accessed in other ways such as observation. When comparing the stimulated recall method to normal interviews, Bloom (1954) argued that stimulated recall interview has an advantage over a post-hoc interview in that the latter relies heavily on memory without any prompts. Furthermore, compared to think-aloud method, stimulated recall can be easier to conduct because the think-aloud method has high requirements on participants' capability of conducting the task and verbally reporting their ideas simultaneously. Stimulated recall interviews have also been used by researchers specifically in the second language learning field. For example, Ryan and Gass (2012) used video stimulated recall to identify and investigate miscommunication in interactions between NS and NNS at a New Zealand university.

This study aims to explore meaning negotiation routines in audio and video SCMC. So, it is very important for the researcher to know whether a student understands the target lexical item. However, previous studies on meaning negotiation mostly rely on the analysis of transcripts and the researchers' interpretation of students' (non-)understanding without the participants' confirmation (e.g. Wang, 2006; Jun and Jie, 2012). This does not seem to be a particularly valid approach since not all non-understandings and misunderstandings are clearly indicated by students with verbal trace (Ryan and Gass, 2012; Van der Zwaard and Bannink, 2014, 2016). A stimulated recall interview allows students to watch what was happening during their oral interaction and recall their mental processes of how he/she managed or failed (or missed) to negotiate meaning at certain points of interaction in synchronous video/audio SCMC classes. Knowing participants' (non-)understanding or misunderstanding is helpful for the researcher to

do the coding more accurately and to triangulate the findings of the analysis of meaning negotiation routines in synchronous video/audio SCMC. Furthermore, the video stimulated recall can also elicit information about students' cognitive processes during the multimodal meaning negotiation processes. For example, in a successful meaning negotiation, what is the hearer's mental process from non-understanding to understanding? What elements or modes of communication (e.g. some particular words or sentences, or their peer's facial expressions, or body gestures) enhance their understanding of the new lexical item?

A central methodological concern about the validity of stimulated recall is whether this method reveals the processes that participants actually engage in, without their recalls being tainted by their personal analysis and re-ordering of thoughts after the fact (Yinger, 1986; Ericsson and Simon, 1996; Lyle, 2003; Huang, 2014). In other words, it is important to ensure that data generated from stimulated recall can reflect access to direct and precise accounts of previous thought processes and behaviours, without any 'intermediate ordering of reflections' or a posteriori reasoning on the original, previously unordered accounts of intention or thought processes (Lyle, 2003, p. 865; Yinger, 1986). It is widely believed that the greater the delay between event and recall, the greater the potential memory decay (Gass and Mackey, 2017). During the pilot study, due to some practical obstacles, the stimulated recall interviews were conducted three weeks after the task sessions, and some participants admitted that they had already forgot what they were thinking at the point of interaction. Drawing on the literature and the lessons learned from the pilot study, the main study collected video stimulated recall data as soon as possible after the task sessions.

4.2.3.4 A summary of study research design and data collection procedures

Based on the above justification of the use of data collection techniques and the improvements summarised from the pilot study, an updated research design was proposed with three main stages including preparation stage, main task stage, and the post task interview stage, as

presented in Table 14. Each dyad went through all the three stages in the main study. The specific data collection procedures are described in Section 4.

Stages	Session	Content	Form of data
Stage 1: Pre-task Preparation stage	Online session 1	Introduction, pairing, ice-breaking; pre-task vocabulary test (video only)	Screen video recordings (around 1 hour for each session)
	Online session 2	Mock IELTS speaking test; opinion gap tasks (audio and video)	
Stage 2: Main task stage	Online session 3	Spot-the-difference tasks (audio and video)	
	Online session 4	Problem-solving tasks (audio and video)	
Stage 3: Post-task interview stage	Face-to- face interview	Video stimulated recall interview about: a) meaning negotiation episodes b) students' opinions and background	Audio recordings (1.5-2 hours each interview)

Table 14: Research procedure for the main study

4.2.4 Data collection procedures in the main study

4.2.4.1 Task sessions

It should be noted that all the online task sessions and the interviews were all specially designed and independently organised as an educational intervention only for the purpose of this Ph.D. research project. This project was advertised as a free online speaking short-term course to the students that was not related or attached to any modules at the university and students' performances were not marked or in other way related to the students' final assessment results at the university. The participation of the students was only motivated by their interest in this

chance of practising their oral English and it did not add any score for them in the final exam. Also, most students did not know each other before participating in this research project because they were doing different modules in different grades and different online study groups.

4.2.4.1.1 Online session 1: Induction

The first stage aims to collect some pre-task data and to get students familiar with the project. Since the video conferencing system could accommodate at most four students to use audio and video SCMC at the same time, each induction session was attended by two dyads. Professor W. taught Dyad 1 and 3, and Lecturer L. taught Dyad 2 and 4. In the induction session, the online teacher tested the participants' video and audio connection; introduced the aim and the procedures of the project; asked for participants' consent, put students into groups/dyads as was planned by the student manager and gave a specific schedule for each dyad. To get students familiar with each other, the online teacher asked students to do an ice-breaking activity with some information gaps. For example, Student D1A (Dyad 1 Student A) first introduced herself, and D1B had to listen to D1A because after the self-introduction, the online teacher asked D1B some questions about D1A based on what she said. In this way, everyone had to listen to their interlocutor carefully and remember as much information as possible in order to be able to answer the teacher's questions. At the end of the session, all students completed the pre-task vocabulary test which included the target lexical items from their mobile phone.

4.2.4.1.2 Online session 2: Mock IELTS speaking test and opinion gap tasks

The second online session includes a mock IELTS speaking test and two opinion gap tasks through audio and video SCMC respectively. As has been highlighted, this session was added to the data collection after the pilot study to better prepare participants for the following task sessions.

The IELTS speaking test exercise was selected from the *Cambridge IELTS book*. Each mock test took 9-15 minutes, which was similar to the real IELTS speaking tests. The students' performance was video recorded. The two online teachers and the researcher marked all eight participants' mock tests independently according to the official IELTS marking criteria. The students' final marks were calculated as the average score among the three markers (see Table 13 in Section 4.2.2.2). The final scores were sent to each individual student as informal feedback, which is not in any way connected to their assessment at the university. As has been stated earlier, the scores confirmed that the pairings were generally suitable because there were differences in the language proficiency levels between the two interlocutors within each dyad.

The second part of the preparation session is the opinion gap tasks (see Appendix 2). There are two opinion gap tasks; one in video and the other in audio SCMC. To elicit more oral interactions between peers, the researcher designed two easy topics: food and shopping, which were both very familiar to Chinese people. The students were also given some specific questions about the topics in case they did not know what to talk about. After the task, the students were asked to summarise their peer's opinions on certain questions that had been discussed. Therefore, they needed to listen to their peers carefully in order to fill in the opinion gaps and answer the questions raised by the online teacher. The main aim of this opinion gap task was not to elicit meaning negotiation interactions, but to train participants to become more familiar with peer interaction through audio and video SCMC. Participants' performances in these two opinion gap tasks were not included in the data analysis and findings.

4.2.4.1.3 Online sessions 3 and 4: Spot-the-difference tasks and problem-solving tasks

All task materials (including a piece of information sheet, a consent form, four task sheets) had been printed and sent to students before the online sessions. During each task session, students were asked to put their task sheet on the desk so that they looked down at the task

sheet. At the beginning of each task session for each dyad, the online teacher asked students to do a quick gaze direction test. In this test, each dyad was asked to look at their peer's video image, teacher's video image and their own video image for three seconds respectively. This test helped the researcher to identify students' gaze directions during meaning negotiation episodes in video SCMC. The use of hard copies of task sheets had important influence on students' gaze directions and their use of multiple modes and semiotic resources to negotiate meaning (see Chapter 6 and Chapter 7 for more details).

Two improvements were made in the main study, particularly about the task instructions. First, at the beginning of each task session, the online teacher highlighted that there was no time limit for the tasks. This new instruction was added because participants in the pilot study often did the tasks in a hurry and missed some opportunities for meaning negotiation. Moreover, for the main study, students were also told that there were no colour differences between the two pictures in the spot-the-difference task, for which students in the pilot study spent a lot of time. These two changes in the task instruction aimed to maximise students' chances of meaning negotiation in a reasonable amount of time.

In total, the eight online task sessions add up to 407 minutes (6 hours 47 minutes) of screen video recordings. These data are the major source for analysing meaning negotiation interactions in audio and video SCMC.

As in the pilot study, throughout the whole task session period, the researcher was present only to make sure everything went smoothly and to collect data of screen video recordings. Before all task sessions started, all the task sheets were sent to each participant. During each task session, the name 'assistant' was used by the researcher to participate in each task session and help students and teachers with their technical issues. The interface recorded from the researcher's laptop was the same as any other students' interface. This means the video recordings could be used as valid stimuli for students during the stimulated recall interview.

Most online sessions went on smoothly and were fully recorded, except for one small problem in Session 2 for Dyad 2 when D2A had some technical issues and had to use her mobile phone to complete the video SCMC tasks instead of the webcam. But the session was screen video recorded and could still be used for data analysis. Admittedly, this change of device had some influence on the data analysis and findings. More details of this technical issue will be discussed in Section 7.2.4.

4.2.4.2 Interviews

4.2.4.2.1 Video stimulated recall interview

As discussed earlier, one of the main improvements in the main study data collection was that the video stimulated recall interview were done face-to-face rather than online. Another improvement in the main study was that students attended the video stimulated recall interview as soon as possible. Since all participants were from Beijing, they were invited to come to the university within three days after their last online session for the face-to-face interview. The interviews were carried out in a quiet conference room at the university with the researcher and the interviewee inside, so all the interviews went smoothly without any external interruptions. The interviews were audio recorded as data for further analysis. The interviews usually took 1-1.5 hours. Altogether, the researcher had 638 minutes (10 hours 38 minutes) of audio interview recordings. English was the main language for all interviews. Only when interviewees felt that they could not fully express themselves in English, they used Chinese.

A number of techniques were used to make interviewees feel relaxed so that they could offer honest answers about their thoughts and feelings of their past interactions. Before interviewing each student, it was emphasised that it was completely fine if they did not remember something. And if they felt they could not express themselves very well in English, they should feel free to answer the researcher in Chinese. When carrying out the video stimulated recall interviews, the researcher followed practical suggestions on how to improve the validity of VSRI data. For

example, when showing students video recordings, the researcher presented the video from the beginning of an utterance and ended at the end of utterance. When playing meaning negotiation related episodes, the researcher showed a full and extended conversation before and after the meaning negotiation episode to offer more contextual information for students to recall what was going on. Furthermore, the researcher highlighted past tense in her questions to remind participants that they should try to remember what was happening there and then rather than talking about their current feeling or ideas. In summary, the researcher made full and appropriate efforts in her attempts to obtain answers that were as honest as possible from the respondent's perspective.

4.2.4.2.2 Additional interview questions

After video stimulated recall questions, some more questions were asked in the form of general semi-structured interview. The questions were related to students' perceptions of these online classes and their general background experience of language learning. Specifically, the interview questions included:

- 1) students' opinions and preferences for peer interaction through audio and video SCMC;
- 2) students' opinions on the student-centred task based language teaching method;
- 3) students' opinions on technology-related issues;
- 4) students' previous language learning experience, their self-estimated English proficiency, the confidence for their English ability;
- 5) how students deal with new words in daily conversation or during their daily reading, listening exercises;
- 6) students' basic personal information, their job, and the chances of using English for their jobs or in their daily life.

The above questions were asked *after* the video stimulated recall sessions because the video recordings would in general help them to recall what was happening and how they felt about it. These questions were not about memory for facts, but about students' opinions, feelings, and

attitudes about these online task sessions in general, which could also be to some extent stimulated by watching some pieces of video recordings.

The video recordings of online sessions offered data about what students did for meaning negotiation. The video stimulated recall interview data offered context-specific reasons about why participants performed the way they did. The additional interview questions provided the researcher with more personal insights into the choices participants made in meaning negotiation interactions with their peers. In other words, the stimulated recall interview questions focused on a negotiated lexical item, while the general interview questions explored reasons of participants' performances in much wider yet still relevant aspects. The different sources of data were used for triangulation and thus strengthening the final research findings.

4.3 Chapter conclusion

In summary, this chapter described all the contextual information related to the study, justified the specific research design and data collection techniques, and recorded the data collection procedures in detail. The next chapters will focus on analysing the data from different perspectives.

5 Interaction Analysis and Findings

The previous chapter described the data collection procedures used in this study. This chapter will move on to data analysis. This thesis presents data analysis underpinned by three theoretical approaches, which are diverse but complementary perspectives on the data under scrutiny. They include interaction analysis, the statistical analysis of gaze and a multimodal analysis. This chapter will focus on the interaction analysis of meaning negotiation routines in audio SCMC. This chapter will first justify the use of interaction analysis, introduce the analytical method itself and present the coding scheme. Then, specific meaning negotiation episodes will be analysed following the coding scheme. Finally, findings from the interaction analysis will be summarised and reported.

To clearly and easily describe the role of interlocutors' meaning negotiation episodes, this study follows the procedures established by Varonis and Gass (1985). However, their terminology has occasionally been modified, not least to accommodate the fact that the present discussion periodically encompasses text-based exchanges. Consequently, using 'the speaker' to refer to the interlocutor who triggers the non-understanding and 'the hearer' to refer to the one who has the non-understanding and initiates the negotiation by asking questions is not wholly appropriate. Instead, such terms as 'the initiator' and 'the respondent' are employed in the three data analysis chapters and the discussion chapter.

5.1 Rationale and coding scheme

5.1.1 Rationale for interaction analysis

As has been identified by the literature review, one of the research gaps in existing studies is that frameworks developed for face-to-face communication (Varonis and Gass, 1985) and text-based SCMC (Smith, 2003) are used to analyse meaning negotiation routines in audio and

video SCMC. However, researchers are aware that the different affordances of SCMC environments have an important influence on meaning negotiation routines. This gap suggests that it is necessary to examine and explore 'How do students negotiate meaning in audio SCMC and in video SCMC?', which is the first research question of this thesis. This chapter focuses exclusively on audio SCMC.

The aim of this chapter is two-fold. On the one hand, it will examine whether meaning negotiation episodes in audio SCMC in the current study conform to the existing frameworks. This follows a deductive approach. On the other hand, the researcher will also explore if any new patterns or stages emerge from the data following an inductive analytical method. The two processes are seen as complementary in leading to inductive theory development.

5.1.2 Interaction analysis method and coding scheme

Interaction analysis is a well-established analytical method for examining meaning negotiation routines and has been widely used in different contexts. It uses existing meaning negotiation frameworks (Varonis and Gass, 1985; or Smith, 2003) as the basis for its coding scheme. In the current study, four dyads produce ten meaning negotiation episodes in audio SCMC during the spot-the-difference and problem-solving tasks. These episodes are transcribed and analysed following an interactionist approach (Ellis and Barkhuizen, 2005, pp. 165-196).

Smith's (2003) model of meaning negotiation routines in a text-based SCMC environment is used as an initial framework for coding speech turns in the transcriptions. Figure 10 presents this data analysis framework. Varonis and Gass (1985) first develops a model for analysing the patterns of meaning negotiation between non-native speakers. According to this, negotiation episodes are responses to instances of non-understanding, as opposed to misunderstanding. The model consists of two main parts: a trigger and a resolution which involves three phases. The trigger (T) is an utterance that causes non-understanding in the hearer. Then, the hearer

signals non-understanding through an indicator (I). A response (R) phase occurs when the speaker remedies the non-understanding. The last phase occurs when the hearer produces a reaction to the response (RR). Smith (2003) proposes an expanded framework, adding the confirmation (C) and reconfirmation (RC) stages after RR. Smith (2003) identifies three types of confirmation, including simple confirmation, reaffirmation (with new information/input) and comprehension check (e.g. 'Got it?'). The final phase in the expanded model is reconfirmation, which usually takes the form of a minimal reconfirmation (e.g. 'OK' or 'Yes'), or a simple appreciation (e.g. 'Thanks'). Smith (2003) justifies the expanded stages in terms of participants' greater need for explicitness in text-based written SCMC interactions than in face-to-face interactions.

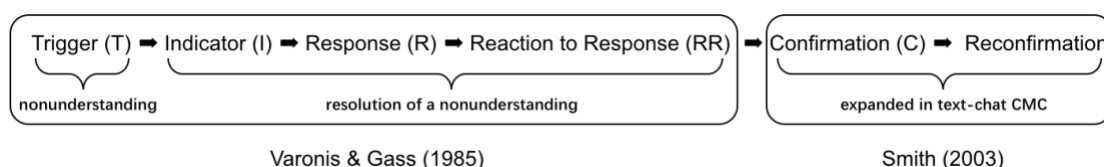


Figure 10: Theoretical frameworks for analysing meaning negotiation routines

Each stage is then subcategorized according to the types of trigger, indicator, response or reaction to response that has been identified. The subcategories are coded using Table 15, summarized by Smith (2003). Specific definitions of each subcategory can be found in Section 2.2.4. In meaning negotiation episodes, interlocutors often perform task appropriate responses (TAR) at different stages, rather than only at the reaction to response stage. These are all coded as 'TAR' in the following analysis.

Trigger	Indicator	Response	Reaction to Response
Lexical	Global	Minimal	Minimal
Syntactic	Local	RT + Lexical	Metalinguistic talk
Discourse	Inferential	Rephrase/	Task appropriate response (TAR)
Content		Elaboration	Testing Deductions

Table 15: Subcategories of negotiation stages by Smith (2003, p. 43)

In addition, data from stimulated recall interviews have been transcribed and are here used to verify the findings from the analysis of participants' oral interactions in meaning negotiation instances, especially to confirm interlocutors' (non-)understanding. Quoted extracts from these interviews are italicized to distinguish them from quotations from the exchanges.

5.2 Interaction analysis

5.2.1 Conforming to previous frameworks of meaning negotiation

The analysis shows that all the ten successful meaning negotiation episodes involve the four basic stages of meaning negotiation (T, I, R, RR) proposed by Varonis and Gass (1985). The following meaning negotiation episode (MNE) shows how the meaning negotiation patterns in my data broadly match their framework.

5.2.1.1 Conform to Varonis and Gass (1985) model

Audio MNE Extract 1 is from Dyad 2's audio interaction for Task 4, the spot-the-difference task (See Appendix 5).

Audio MNE Extract 1: D2 T4 'couch' episode

Turn	Participant	Speech	Negotiation stage
1	D2A	ehm, OK, let's go on, hmm then, there's a couch in the middle of the picture, right? a green couch	T
2	D2B	green coach?	I
3	D2A	couch, a sofa, sofa	R
4	D2B	oh, sofa, yes, sofa, a green sofa in the middle of the picture	RR

5	D2A	yeah	TAR
6	D2B	and a little boy	TAR

The two interlocutors are negotiating the meaning of the word 'couch'. First of all, D2A was describing her picture with the word 'couch', which triggered a lexical non-understanding for D2B. D2B indicated the problem by repeating the trigger with a rising intonation but misunderstood and mispronounced the word as 'coach'. During the interview, D2A explained that D2B (in Turn 2) sounded 'a bit confused' and also mispronounced the word 'couch', which was why she knew D2B did not understand it. D2B's attempt to repeat the trigger suggests that it was a local indicator because she clearly pointed out the cause of her non-understanding. Then, in the response stage (Turn 3), D2A repeated the trigger by emphasizing the correct pronunciation and explained it by rephrasing 'couch' with a synonym 'sofa'. Turn 4 contained D2B's reaction to her interlocutor's response in which she showed her understanding of the word by repeating D2A's description in Turn 1, but with the word 'sofa' instead of 'couch'. This turn closed this meaning negotiation episode and brought the conversation back to task appropriate discussion. D2A quickly confirmed D2B's repetition (Turn 5), and D2B moved on to their task discussion in Turn 6.

Audio MNE Extract 1 demonstrates an interaction that perfectly matches the meaning negotiation patterns identified by Varonis and Gass (1985) without any confirmation. However, in most cases in my study, the patterns offer a closer fit with Smith's (2003) extended framework because there tends to be a confirmation and reconfirmation after the reaction to response stage.

5.2.1.2 Conform to Smith 's (2003) model

Audio MNE Extract 2 is from Dyad 2's audio interaction in carrying out Task 5, a problem-solving task where two participants were asked to select four gifts out of 8 to give to members of their homestay family.

Audio MNE Extract 2: D2 T5 'perfume' episode

Turn	Participant	Speech	Negotiation stage
1	D2A	ehm, OK, OK, next one is Chanel (Researcher's comment: D2A's pronunciation of this word always sound like 'channel') perfume, very expensive, Chanel perfume, perfume	T
2	D2B	I, I I know the brand of Cha Chanel, and perfume, what is perfume?	I
3	D2A	perfume, perfume, the liquid that when you spray on on yourself, it will make you smell good, perfume	R
4	D2B	oh, I know that, I know that	RR
5	D2A	OK?	C
6	D2B	I know that many women like this	RC
7	D2A	yeah, right, yeah, right	/

First, D2A introduced her item, 'perfume', and it seemed that she expected D2B might not understand this word, so she added some explanation 'very expensive' and repeated the word three times even without D2B's indicator. D2B did not understand 'perfume', just as D2A expected, but managed to clearly indicate the lexical trigger 'perfume' with a question. In Turn 3, D2A responded to D2B's question with a clear elaboration of the word 'perfume' and repeated it again. D2B then immediately understood what was meant by 'perfume' and signalled this by exclaiming 'oh' with falling intonation and quickly repeated 'I know that' as her reaction to D2A's response. But D2A required more confirmation, and to seek it, she asked 'OK' with rising intonation in Turn 5. Then D2B reacted to D2A's question by adding her own understanding of the word 'many women like this'. By then, D2A was finally convinced that D2B had managed to guess the meaning correctly. With the 'yeah right' repeated twice, D2A closed the meaning negotiation routine. They could now move on to the task-related discussion.

During her video stimulated recall interview, D2B confirmed that she did not know the word 'perfume' before and managed to guess the meaning because of three pieces of important information given by D2A: the brand 'Chanel', liquid, and it smells good.

5.2.2 Two new stages: Confirmation of Trigger (CT) and Clarification of Indicator (CI)

In addition to the stages derived from existing frameworks, new stages were identified in meaning negotiation episodes in audio SCMC from the data. The trigger and the indicator stages were discovered not always to be directly followed by the problem-solving stage incorporating response (R) and reaction to response (RR). Instead, participants in audio SCMC interactions tended to confirm the trigger (CT) and the indicator (CI) before moving on to resolving their non-understanding. For example, during the confirmation of trigger (CT) stage, the initiator (the interlocutor who knows the meaning of the word) would usually repeat the trigger with rising intonation to ask the respondent if this is the problem. With this confirmation request, the respondent (who does not know the meaning of the word) would usually confirm the indicator (CI) by clarifying what exactly the problem is. It suggests that only after both speakers understand the problem clearly, they move on to the stage of resolving the non-understanding.

The three examples below will demonstrate the occurrence of CI and/or CT at different stages of meaning negotiation: (1) CT and CI before the resolution stage, (2) CT and CI during the resolution stage, and (3) CI after the resolution stage. Categorizing the use of CT and CI according to their position in a meaning negotiation routine is helpful for the subsequent discussion of the reasons for these phenomena (see Section 5.3 in this chapter).

5.2.2.1 CT and CI before the resolution stage

Audio MNE Extract 3 is from Dyad 2's audio interaction in Task 5, the problem-solving task where two participants were asked to select four gifts out of eight to give to members of their homestay family.

Audio MNE Extract 3: D2 T5 'razor' episode

Turn	Participant	Speech	Negotiation Stage
1	D2A	OK, I will tell you what I've got, four items, they are a razor, do you know razor?	T
2	D2B	razor, sorry	I
3	D2A	yeah, razor	CT
4	D2B	sorry can you explain? razor	CI
5	D2A	it is for, it is used by a man to shave his face, shaving, you know?	R
6	D2B	oh, I (..) no (.) I know that, I know that	RR
7	D2A	you know that? it is, it is used to get rid of the moustache or ...	C
8	D2B	yeah, I I know that, I know that	RC
9	D2A	OK	C
10	D2B	ehm, men, men often use it	RC
11	D2A	haha, OK, next one ...	TAR

D2A began by stating her intention to list all four of her items to D2B, but after saying the first item, 'razor', she performed a comprehension check to see whether D2B understood this word. In the stimulated recall interview, when asked why she had done that, D2A explained, 'because through communicating with her, during several classes, I know she did not master a lot of vocabulary so, I think she probably doesn't know what is a razor [sic]'. Hearing the comprehension check, D2B uttered the trigger by saying 'sorry' as a negative response to it. However, in her reply (Turn 3), D2A only repeated the trigger, and offered no further explanation. So, this turn was used more as a confirmation of the trigger rather than a response because it contained no attempt to resolve the problem or to explain the meaning. Then (in Turn 4), D2B explicitly asked D2A to explain the word 'razor', which confirmed that she did not understand the lexical trigger. In this turn, D2B confirmed her indicator and clarified the issue. Only then did both interlocutors arrive at a consensus that they needed to work out the meaning of 'razor' for D2B's benefit.

Having understood that D2B's difficulty was not with the pronunciation but the meaning of the word 'razor', in Turn 5, D2A started resolving the non-understanding by explaining the use of a razor, followed by a comprehension check 'you know?'. In her reaction to D2A's response (Turn 6), D2B first hesitated and initially said 'no'. In the interview, she admitted that this was because she 'didn't remember' the word 'shaving' in D2A's explanation, which was why she said 'no'. But then she quickly changed her mind and said, 'I know that' twice to confirm to D2A her understanding as she could 'guess the sentence because D2A first said 'cleaning the face'.

Although D2B said she understood the word, D2A revealed at the interview that she was thinking, 'I'm not quite sure if she really understands what it is, I tried to give her some further references to make her know clearly what it is'. In Turn 7, D2A did not use the word 'shave', instead, she paraphrased it by saying 'get rid of the moustache', she even said 'or' but then paused because, as she later explained, she was 'trying to find an easier alternative for moustache'. But D2B interrupted D2A's further explanation at this point (Turn 8) and reacted to her with a clear and strong confirmation 'yeah, I know that' with another repetition to stress that she understood the meaning of the word 'razor'.

However, it seemed that D2A was still not fully convinced that D2B really understood 'razor' as her response to D2B's reconfirmation was a minimal 'OK'. Then D2B (in Turn 10) added her own explanation by saying 'men often use it', suggesting that she managed to guess the meaning of 'razor' correctly based on D2A's explanation. D2A's laugh in Turn 11 showed that she was finally convinced that D2B understood 'razor' correctly and became more relaxed. She then closed this meaning negotiation episode with an 'OK' and carried on their task-related interaction by signalling 'next one'.

Sometimes, participants might have already moved to resolve the non-understanding but without clearly understanding what exactly the problem was. Once participants realized that confusion persisted about the issue, they tended to come back to confirm the trigger (CT) and

the indicator (CI) to clarify the issue before moving on again to resolving the problem. The following two examples illustrate the meaning negotiation routines that happen in such cases.

5.2.2.2 CT and CI during resolution stage

Audio MNE Extract 4 comes from the audio interaction by Dyad 4 for Task 4, the spot-the-difference task, where two participants had different pictures and had to describe their pictures to each other and work together to identify the differences.

Audio MNE Extract 4: D4 T4 'drawer' episode

Turn	Participant	Speech	Negotiation stage
1	D4A	there is a driver in the desk	T
2	D4B	hmm?	I
3	D4A	it's the driver with a lock in the desk.	R
4	D4B	IN the desk?	TAR
5	D4A	under the desk or something...do you have a, is there a driver in your picture?	CT
6	D4B	driver? what kind of driver? drive what?	CI
7	D4A	hmm...	R
8	D4B	car driver?	CI
9	D4A	抽屉 [drawer]	R
10	D4B	drawer.	RR
11	D4A	drawer.	C
12	D4B	yeah, yeah, there is a drawer.	RC, TAR

At first, D4A wanted to indicate the presence of a drawer, but she pronounced the word as 'driver', which triggered a non-understanding for D4B. In Turn 2, D4B indicated her non-understanding with 'hmm', using rising intonation. D4A replied in Turn 3 by expanding her previous sentence and trying to locate the drawer in the picture, but she still pronounced 'drawer' as 'driver'. This attempt to explain the drawer showed that D4A was already moving to the explanation stage because she was still not aware that her mispronunciation had resulted

in non-understanding for her peer. In Turn 4, D4B reacted to D4A's response with a question 'in the desk?' with a clear stress on the preposition 'IN'. This suggested that D4A's explanation had confused her, rather than leading to understanding. It could be seen that in turns 3 and 4, both interlocutors moved to resolve the non-understanding, although they had not yet reached a shared understanding of the nature of the communication breakdown.

In Turn 5, D4A first changed the preposition 'in' into 'under' but revealed at the interview that she was 'not sure'. Then, in the second part of the turn, she appeared to realize that the main problem for D4B was not the preposition, but the word 'drawer', so she tried to confirm the trigger (CT) with D4B by asking if there is a 'drawer' (pronounced as 'driver') in her picture. In this turn, D4A finally realized that the problem might be 'driver/drawer', so she returned to confirming the trigger with D4B. However, her mispronunciation made D4B more confused. Consequently, in Turn 6, D4B uttered three consecutive questions 'driver?', 'what kind of driver?', 'drive what?' to directly point out her non-understanding, by insistently demanding a clear answer. This stage was devoted to confirmation of the indicator, by means of repeated clarification requests. Only when the nature of the source of non-understanding was established could D4A and D4B move on to the resolution phase. During the stimulated recall interview, D4B confirmed that she was 'feeling impatient at this point because of the confusion caused by D4A'.

In Turn 7, D4A wanted to explain the word but did not manage to say anything. As she recalled in the interview, she 'didn't know also how to describe the driver/drawer, because my pronunciation is not good'. D4B might have realized that D4A was not able to explain the word, so in Turn 8, D4B offered a guess, 'car driver?'. In an attempt to confirm the indicator and to clarify the problem caused by D4A, D4B resorted to a strategy referred to by Smith (2003) as 'testing a deduction' (p. 44).

In Turn 9, D4A chose to use Chinese to reply to D4B's question. This was her second attempt to resolve the non-understanding. D4B quickly understood the Chinese term and pronounced

the word correctly. D4A repeated D4B's pronunciation, which was a form of modified output. D4B confirmed the existence of a drawer to her and moved on to their task-related response.

This example demonstrates how pronunciation could trigger non-understanding in audio SCMC interactions in a way that was simply not possible in text-based SCMC interactions and represents a fundamental difference between the two modes. Equally, difficulty in the perception of an utterance could be yet another source of non-understanding. In this case, however, D4B was listening carefully and heard D4A clearly, but she could not make sense of it as she thought D4A was saying 'driver'. The following example is also related to pronunciation.

5.2.2.3 CI after the resolution stage

Audio MNE Extract 5 is the 'stationery' episode from Dyad 2 Task 3 where the two interlocutors were asked to choose four gifts out of eight options for four different members in their homestay family.

Audio MNE Extract 5: D2 T3 'stationery' episode

Turn	Participant	Speech	Negotiation Stage
1	D2A	ehm, the third one is, stationery, stationery, you know, like pencil, paper, notebook, erasers and scissors, are used, used for, studying ... do you know that?	T
2	D2B	em, could you please spell, spell this word?	I
3	D2A	OK, I will ex, explain it to you, station, stationery includes pencils, and notebook, scissors, erasers, the tools you use during your study, when you need to write something, you need a pencil, and then write on a notebook, right?	R
4	D2B	yes	RR
5	D2A	things you use during your study, have you got it?	C
6	D2B	em, how to pronounce it please? could you please pronounce it again? stationery, right?	CI

7	D2A	stationery, stationery	R
8	D2B	stationery, stationery, yes, I know, I I can guess, what it is	RR
9	D2A	OK, it's pencils, pens, and notebooks, scissors, erasers, got it?	C
10	D2B	oh, I OK, I can gu, I can guess it	RC

At the interview, D2B said, of the term 'stationery', 'I can't guess what it means', so in Turn 2, she asked 'her [D2A] to spell it for me'. However, D2A did not comply with her request but instead explained the meaning of the word in turns 3 and 5. D2B later revealed that she could 'imagine (guess) general idea of its meaning' but 'was still not very sure what it is really'. D2B now seemed to attribute the problem to pronunciation because after D2A's explanations in Turn 3 and 5, she came back to confirm the indicator (CI) in Turn 6 and stressed twice that she wanted to know the pronunciation of the word. After D2A pronounced the word clearly (Turn 7), D2B was finally satisfied and confirmed that she could guess the meaning in Turn 8.

A possible reason why she insisted on asking for the pronunciation of the word could be that she wanted to infer the meaning of the word from its pronunciation. But the word 'stationery' was problematic in this respect. It sounded like 'station' as in 'bus station', but its meaning has nothing to do with 'bus station'. Therefore, D2B might have found it hard to connect the meaning of the word to its pronunciation. This example confirmed that the respondent's perception of even an accurately pronounced item may play an important role in negotiation for meaning.

The above analysis presented three meaning negotiation routines in detail. Table 16 is a summary of all ten meaning negotiation episodes in relation to the CT and CI stages and the sources of non-understanding. Seven out of ten MNEs had CT and CI stages in their meaning negotiation routines across all three dyads (Dyad 1 did not succeed in any of their MNEs). As for causes of non-understanding, five episodes were only caused by meaning, two by pronunciation and another three by both meaning and pronunciation.

Dyad	Lexical item	CT and CI stages	Reasons for non-understanding
D2	couch	N/A	meaning
D2	cube	CT, CI	meaning
D2	carrot	CT	pronunciation
D2	Rubic's cube	CT, CI	meaning and pronunciation
D2	razor	CT, CI	meaning
D2	perfume	N/A	meaning
D2	stationery	CI	meaning and pronunciation
D2	skateboard	CT, CI	meaning
D3	toaster	CT, CI	meaning and pronunciation
D4	drawer	CT, CI	pronunciation

Table 16: A summary of MNEs in audio SCMC

5.2.3 Technical issues

The above findings are based on the analysis of ten successful meaning negotiation episodes in audio interactions. Eight out of the ten episodes came from Dyad 2, who were particularly good at meaning negotiation. The other two episodes came from dyads 3 and 4. In fact, twelve successful meaning negotiation episodes were found in all audio tasks. But two of them were caused by technical issues where one participant could not hear the other clearly, so their interactions were full of noise, interruptions, confusion and misunderstanding. Although these two episodes were not included in the analysis, it should be acknowledged and highlighted that in synchronous audio SCMC environments, communication breakdowns could be triggered not only by linguistic non-understanding but also by technical issues, which is a major difference from text-based SCMC in Smith (2003).

5.3 Findings

5.3.1 Three possible routines

As can be seen from the above analysis, the two new stages CT and CI, appeared in meaning negotiation routines at three possible points: 1) immediately after the indicator and before any resolution or explanation, 2) during the resolution stage (R or RR), and 3) after completing the resolution stage and following the confirmation request (Figure 11). These three possible routines are drawn from the meaning negotiation episodes identified in the data and are exemplified in the three extracts. But it is important to discuss why these three examples generate different pathways of meaning negotiation.

In the 'razor' example (see Section 5.2.2.1, Audio MNE Extract 3), the source of non-understanding was the meaning of the word 'razor'. In the CT stage, the initiator confirmed the pronunciation of the word to the respondent. In this case, the pronunciation was not the cause of the non-understanding. During the CI stage, the respondent clearly requested the initiator to explain the meaning of the word. Here, the meaning negotiation routine follows pathway 1 in Figure 11.

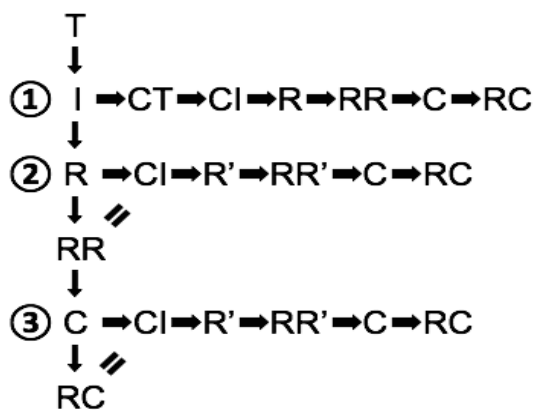


Figure 11: Negotiation for meaning pathways as exemplified in three audio SCMC interactions

However, in the 'drawer' example (see Section 5.2.2.2, Audio MNE Extract 4), the source of the trigger of non-understanding was not the meaning of the word 'drawer' but its pronunciation as 'driver'. The initiator, D4A, struggled with her explanation. Her interlocutor, D4B, engaged in three successive attempts at resolution by (1) exploring global meaning, (2) seeking clarification of a local indicator, and (3) testing a deduction. Finally, the focus was transferred to D4A's

mispronunciation of the word 'drawer' as 'driver'. This was finally revealed as the actual trigger of non-understanding, but it took several turns devoted to repeating and seeking to confirm the trigger to figure this out. Only when the initiator had recourse to the L1 equivalent of 'drawer' did the respondent realize that her non-understanding was caused by the initiator's mispronunciation.

In the model proposed for Pathway 2 in Figure 11, the RR stage and CI stage are connected by an equals sign to indicate that the clarification and confirmation of a non-understanding may be embedded either in the RR (respondent's reply to the initiator's explanation) or (as in Pathway 3 in Figure 11) in the RC (reconfirmation) stages of a meaning negotiation routine

In Audio MNE Extract 5, the 'stationery' example (see Section 5.2.2.3, Audio MNE Extract 4), the source of the respondent's difficulty was both meaning and pronunciation. Therefore, when the initiator was explaining the meaning of the word, the respondent did not interrupt but listens to it. After the explanation, when the initiator performed a confirmation request (Turn 5), the respondent confirmed her understanding of the meaning of the word (Turn 8), but not before raising another question (CI) about its pronunciation (Turn 6).

These three routines are different essentially because the nature of the spoken medium permits different kinds of non-understanding. Potential non-understanding could be caused by the meaning of the word, or its pronunciation, or both. This finding is different from that in written interactions in text-based SCMC environments where the only possible cause of non-understanding is meaning.

5.3.2 An expanded meaning negotiation routine in audio SCMC

The three possible routines reveal different layers of potential non-understanding due to the specific technological affordances of the audio SCMC environment, which is why a further

confirmation of the trigger and the indicator take place. Despite the different pathways, it can be seen that CT and CI stages were always followed by a resolution stage with R and RR. In other words, in audio SCMC, it was not possible to move to resolve non-understanding unless both interlocutors clearly identified what exactly triggered it. The respondent tended to find an appropriate point at which to clarify the issue and asked the initiator to resolve the non-understanding. Thereupon, an expanded meaning negotiation routine can be concluded as in Figure 12. The whole meaning negotiation process in audio SCMC includes up to eight stages with five different aims: (1) triggering non-understanding (T); (2) identifying non-understanding (I); (3) clarifying the non-understanding (CT and CI); (4) resolving non-understanding (R and RR); and (5) confirming understanding (C and RC).

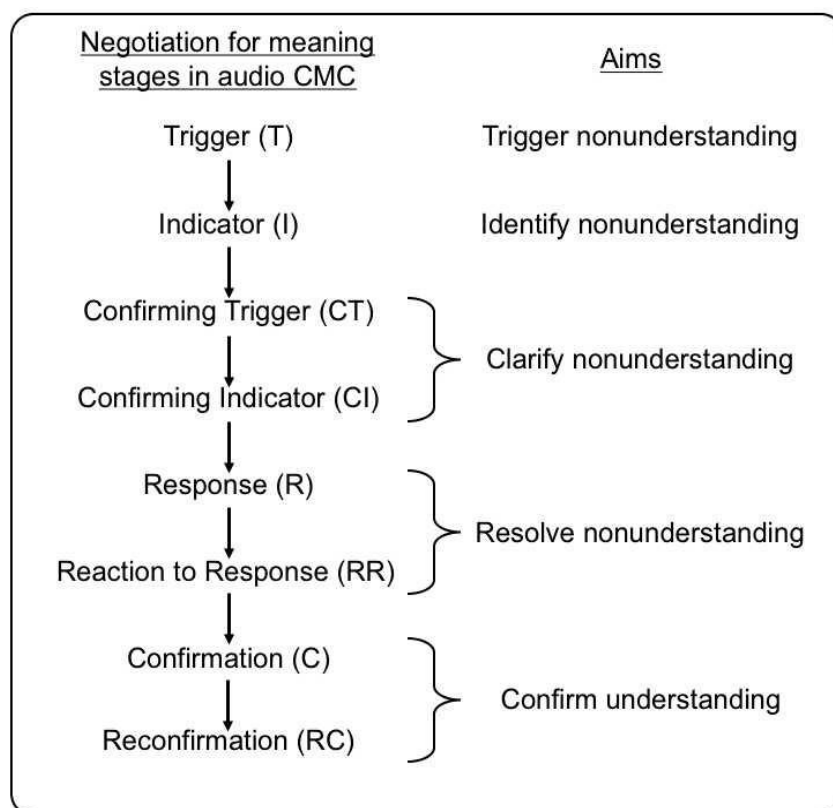


Figure 12: A negotiation for meaning routine for audio SCMC interactions: An expanded model

Compared to the text-based model by Smith (2003), this expanded routine highlights the clarification of non-understanding. The proposed stages in meaning negotiation (CT and CI) can help interlocutors to understand the nature of the non-understanding in the audio SCMC context. In this sense, synchronous audio communication may require an even higher level of explicitness than text-based SCMC, as claimed by Smith (2003). The essential difference is the phonological dimension, which does not exist in text-based SCMC, but can trigger many non-understandings in speech-based interactions in either audio or video conferencing.

5.4 Chapter conclusion

This chapter followed the interactionist approach, analysed five meaning negotiation episodes in audio SCMC and presented illustrative examples. The key findings of the analysis are as follows. First, meaning negotiation routines in audio SCMC conformed to the existing frameworks of Varonis and Gass (1985) and Smith (2003). Second, two new stages, confirmation of trigger (CT) and clarification of indicator (CI), were identified in audio SCMC negotiated interactions. Third, three variants of the expanded routine were identified depending on the type(s) of trigger. Based on the above three findings, an expanded meaning negotiation routine for audio SCMC was proposed.

The next two chapters will focus on analysing gaze in meaning negotiation routines and the role of multimodality in video SCMC.

6 Gaze Analysis and Findings

The previous chapter has analysed meaning negotiation routines in audio SCMC. This chapter moves on to explore the role of gaze in meaning negotiation in video SCMC. This will contribute to answering the second research question: 'What roles do multiple modes and semiotic resources play in meaning negotiation episodes in video SCMC?'. This chapter first justifies why this analysis is necessary and outlines the coding scheme that is used. Then it moves on to explain how gaze directions can be identified and confirmed using different sources of data, and what episodes are included in the analysis. Next, the coding process and the statistical analysis will be presented. Finally, the findings of statistical regression analysis will be reported and explained.

6.1 Rationale, coding scheme, and coding method

6.1.1 Rationale for a statistical gaze analysis

The reasons for conducting a statistical gaze analysis are as follows. First, existing SCMC studies suggest that some students do not look at their peer's video image during meaning negotiation in video SCMC (Lee, 2006; Wang and Tian, 2013; Guo and Möllering, 2016). In many cases, by not doing so, they miss important multimodal information from their peers during interactions. On the other hand, those students who often looked at their peer's video image during negotiated interactions seemed to be more successful at meaning negotiation than those who seldom did so (Wang and Tian, 2013). But this argument needs statistical proof. This generates an initial hypothesis that there might be a potential positive relationship between the time participants spent looking at their peer's video image and successful meaning negotiation results.

Moreover, it seems that there is disagreement on the role of the visual mode in video SCMC in the literature. Some studies argue that video can be distracting for students when they are trying to focus on the language during task interactions (e.g. Lee, 2006; Van der Zwaard and Bannink, 2014, 2016). However, other studies have reported positive effects of video for second language learning in SCMC environments (e.g. Wang, 2006; Wang and Tian, 2013; Guichon and Cohen, 2014). More research needs to be carried out to examine the role of the visual mode in video SCMC.

Researchers who have undertaken gaze analysis (Develotte et al., 2010; Satar, 2013; Lamy and Flewitt, 2011) have tended to use qualitative methods to explore patterns of interaction, rather than focusing exclusively on meaning negotiation episodes. Instead, this chapter will propose a quantitative approach to measuring the importance of the visual mode in meaning negotiation episodes in video SCMC. It aims to fill a research gap by using inferential statistical analysis to examine the role of gaze in such episodes.

To present the features of gaze and other relevant multimodal features, this chapter has to use many screenshots of the video SCMC task interactions. To protect participants' and the online teachers' privacy with maximum effort while still presenting sufficient data for the gaze analysis, some pieces of information on the screenshots, such as their names and other personal information are covered.

6.1.2 The coding scheme

Since the objective of gaze analysis is to explore the role of gaze in meaning negotiation in video SCMC, the two factors chosen for the statistical analysis are the amount of time students look at their peer's video image on the screen and the number of successful meaning negotiation episodes within each dyadic interaction. The latter can be easily counted. It is more challenging to calculate the amount of time students spent looking in different directions as

these change frequently. The first step is to come up with a coding scheme for students' gaze directions. Based on initial scrutiny of the video recordings of all 37 meaning negotiation episodes, two gaze directions appeared highly related to the research question about the roles of multiple modes and semiotic resources in MNEs in video SCMC. They were looking at the peer's video image on the screen; and looking at the task sheet on the desk. These two directions occurred most frequently. Prolonged detailed study led to the identification of two further gaze directions, making four in all. The final coding scheme for gaze directions comprises: (1) gaze directed at peer's video image on the screen; (2) gaze directed at the task sheet on the desk; (3) gaze in other directions and (4) unidentifiable gaze directions.

The coding scheme used in this study is similar to the one developed by Lamy and Flewitt (2011), who categorize gaze according to the part of the video SCMC interface interlocutors focus on (see Section 2.4.2.3.2). But since the paper is written in French, a language that the researcher does not speak, it was not possible for the researcher to make any further comments or adaptations based on their coding scheme. The final coding scheme of the current study was developed according to the research question and the data, as was explained in the previous paragraph. Other existing gaze direction categorizations developed by Develotte, Guichon and Vincent (2010) and Satar (2013) are not employed in this study because the focus of the current study is on the role of multiple semiotic resources in meaning negotiation episodes in video SCMC, while the two existing models focus on online teaching strategies and social presence.

Another widely discussed gaze direction in literature (e.g. Sindoni, 2014) is the impossibility of mutual eye contact during video SCMC, as has been introduced in the literature review (see Section 2.4.2.3). Mutual eye contact is impossible in video SCMC because the webcam and the peer's video image are not in the same place in the video SCMC interface. When one is looking at their peer's video image, he/she cannot focus on the webcam, vice versa. In the current study, mutual eye contact is not identified as a key gaze direction mainly for the following two reasons. On one hand, participants' the social presence and relational

management are not directly related to the research questions. On the other hand, in video SCMC where the peer's video image is much smaller than the actual size, and their peer's gaze directions are far less clear than in face-to-face communication, eye contact may play a less important role. Interlocutors tended to use a wide range of semiotic resources (such as head movements, hand gestures, facial expressions) to express themselves and to understand their peer, rather than attempting to make some impossible mutual eye contact.

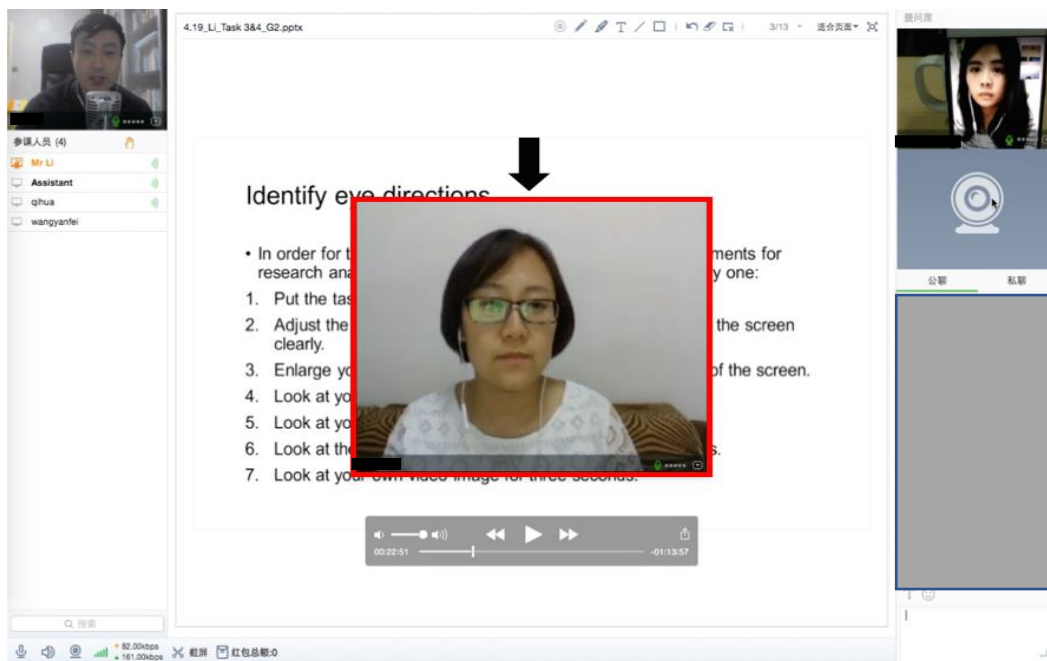
The following paragraphs will offer definitions and examples of the four codes used in this study and explain how gaze directions can be identified and triangulated with different sources of data.

6.1.2.1 Gaze directed at Peer's video image

This code refers to those points when a participant is looking at her peer's video image, however momentary. According to the interface of the video conferencing system, the default position of the teacher's video image is at the top left-hand corner of the screen, and the students' video frames are in the top right-hand corner (Screenshot 1). When a student wants to enlarge either the teacher or a peer's video image, they can double click on it, and it will appear in the middle of the screen in a bigger frame (Screenshot 2). In this system, the size and the position of participants' video images in the interface cannot be customised, meaning that these are the only two possible places for interlocutors' video image. This has important effects on participants' gaze directions and offers possibilities for the researcher to be able to identify their gaze directions.



Screenshot 1: Peer's video on the top right corner of the screen



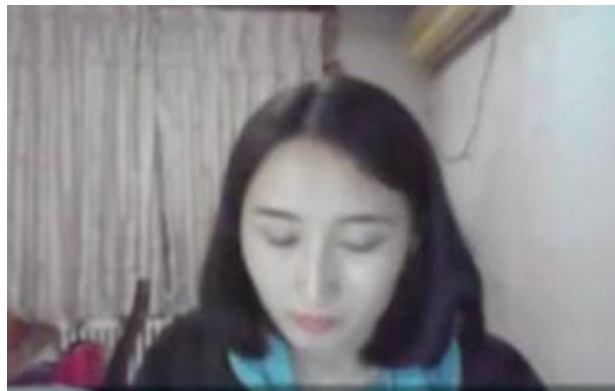
Screenshot 2: Peer's enlarged video frame in the middle of the screen

6.1.2.2 Gaze directed at the task sheet

Another main gaze direction for participants is looking down at the task sheet (Screenshot 3 and 4). The task sheets were given to students in hard copy so that they would not use the electronic version on their screen, thus obscuring the SCMC interface. This also made it more straightforward to distinguish between gaze at their peer's video image and gaze at the task sheet.



Screenshot 3: D1A looking down at the task sheet



Screenshot 4: D1B looking down at the task sheet

6.1.2.3 Gaze in other directions

Gaze at the peer's video image and at the task sheet were the two main gaze directions observed in the sample, but students also looked at other things. For example, some students looked at other parts of the screen including their own video images, or the teacher's video image, or the task instructions on the screen; some students used an online dictionary to look up new words, so they minimised the video conferencing interface for a while. In other cases, their eyes did not focus on anything in particular. For example, they might roll their eyes or look up when they are thinking. However, these cases happen only occasionally, and it was hard to interpret gazes that happened without any noticeable patterns. Since this chapter focuses on the relationship between the time students spent looking at their peer's video image and their success in meaning negotiation, gazes in other directions are not directly related to the research aim. Therefore, all these cases were coded as 'other directions'.

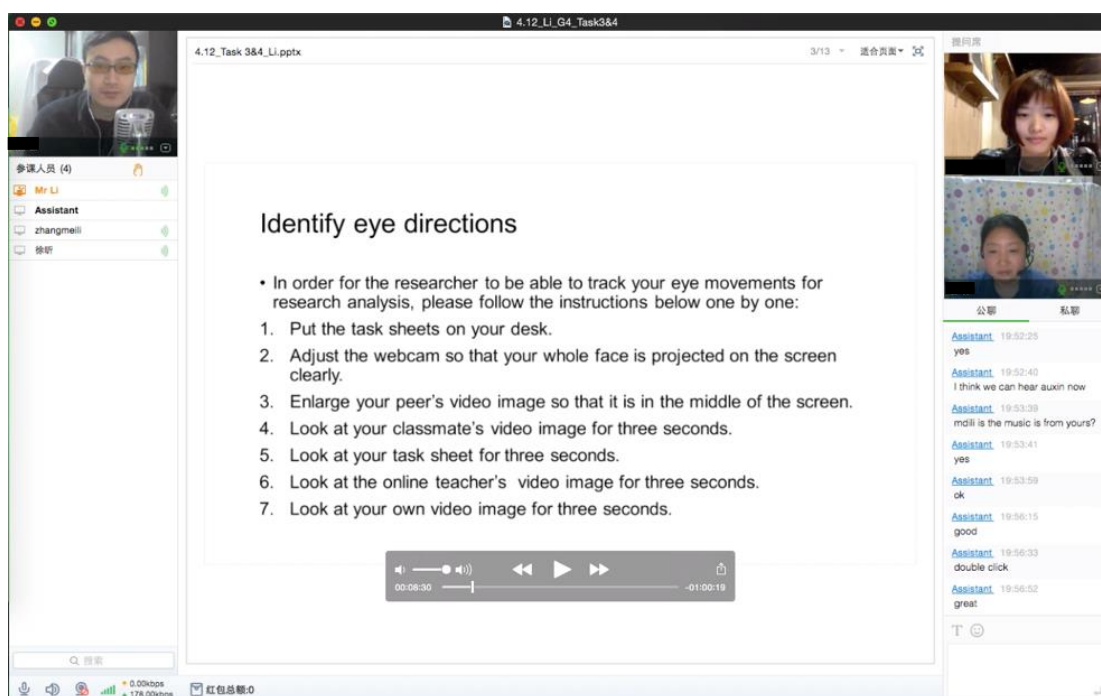
6.1.2.4 Unidentifiable gaze directions

In very extreme cases, some students' eyes were out of the video frame for a short period of time. These cases were coded as 'unidentifiable' since the researcher could not judge the direction of the participants' gaze from the video recording. This category is different from the previous code 'other directions' because in 'other', students' eyes were still in the video frame and the participant was neither looking at their peer nor looking at their task sheet. Distinguishing between 'other directions' and 'unidentifiable directions' could improve the validity of this approach to gaze analysis by offering a full picture of students' gaze directions during meaning negotiation episodes in video SCMC.

6.1.3 How to identify gaze

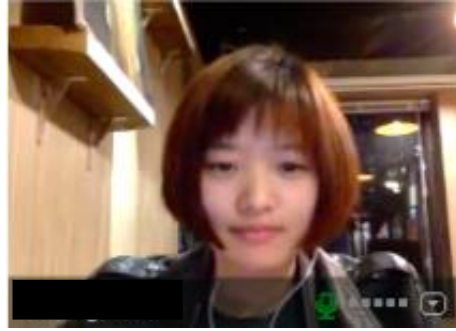
6.1.3.1 Pre-task gaze direction test

Before each task, the students were asked to carry out a gaze direction identification test (Screenshot 5). This process was overseen by the online teacher. Students in the test were asked to put their task sheet on the desk, adjust the webcam, enlarge the peer's video image and then look at their peer's video image, teacher's video image and their own video image for three seconds respectively. This step was taken before data collection because it had been anticipated that gaze might be one of the factors that influenced meaning negotiation. Students' gaze directions during the test offered the researcher a reference point for what a student's gaze looked like when she was looking at her peer's video.



Screenshot 5: The gaze direction identification test screenshot for D4_T3

Screenshots 6 and 7 are the screenshots of Student D4A's gaze directed at her peer's video image during the test and task interactions. Screenshots 8 and 9 are the screenshots of Student D4B's gaze directed at her task sheet during the test and task interactions. Comparison between students' gaze during the test and task interactions enables an initial judgment about students' gaze directions.



Screenshot 6: D4A_T3: Gaze directed at peer's video_test



Screenshot 7: D4A_T3: Gaze directed at peer's video_task



Screenshot 8: D4B_T3: Gaze directed at task sheet_test



Screenshot 9: D4B_T3: Gaze directed at task sheet_task

6.1.3.2 Persistent gaze in one direction

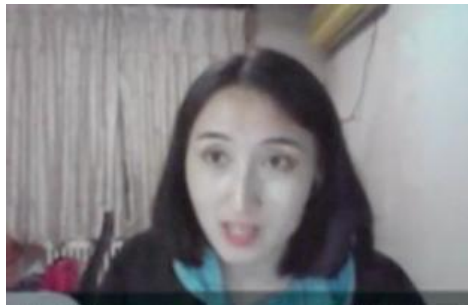
Following the application of the coding scheme, this study deals predominantly with gaze directed at two objects, the peer's video image and the task sheet. Gaze at the task sheet could be readily identified because students were asked to put the task sheet on the desk, and they had to look down while looking at the task sheet. The main difficulty for the researcher was how to distinguish, when a student was looking up at the screen, whether the student was looking at her peer's video image or at other parts of the screen.

Although all students were asked to take the gaze direction identification test, they were still able to make changes during their task interactions. For example, many students close the peer's video window in the middle of the screen, and their peer's video window moved automatically to the default location in the top right-hand corner of the screen. In this case, the researcher was unable to identify their gaze direction in task interactions by comparing them with the student's gaze in the test. But if a student was constantly and repeatedly looking in the same direction, especially when talking to their peer during negotiated interactions and the gaze direction was either on the middle of the screen or the right top corner, then it was highly likely that it was directed at their peer's video image. This method is based on an underlying assumption that people tend to look at the person they are talking to. In other words, it was

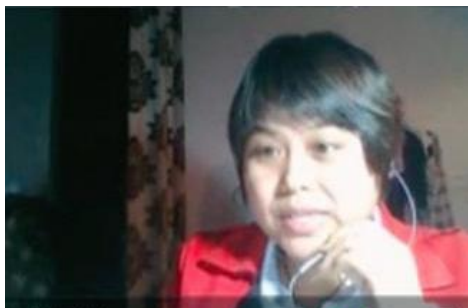
almost impossible for a student constantly and repeatedly to focus on other things on the screen while talking to their peer.

For example, Screenshot 10 and 11 show Student D1A's and D3A's gaze screenshot while they were talking to their peers. The video recordings show that they looked in this particular direction repeatedly, especially while talking to their peers. Their gaze directions were also towards the top right-hand corner of the screen, as compared to the direction in the pre-task gaze identification test. Therefore, it is overwhelmingly likely that their gaze was directed at their peer's video image.

It is acknowledged that the above assumption for this coding method may not always hold. Therefore, confirmation of gaze directions by participants is needed in the video stimulated recall interviews, which is presented in Section 3.3.



Screenshot 10: D1B's gaze directed at peer's video while talking



Screenshot 11: D3A's gaze directed at peer's video while talking

This approach could be applied even when the participants encountered technical problems, such as D2A, who could not see D2B's video on her own screen. To solve the technical issue, the researcher, who attended the session as a recording assistant, used her mobile phone as a 'third party' to connect D2A and D2B so that both students could see each other's image during their video SCMC interactions. The task session went on smoothly in this way, but the students' screen video recordings were slightly different from others, thus causing some difficulty for the gaze coding process. However, after watching this dyad's interactions repeatedly, the researcher discovered that D2A frequently adopted a particular gaze direction especially when she was talking to D2B. Screenshot 12 shows four screenshots of D2A's gaze while talking to D2B in four different meaning negotiation episodes. The consistent pattern offered a strong indication that D2A's gaze in all four shots was towards D2B's onscreen video image.



Screenshot 12: D2A's gaze while talking to D2B in different meaning negotiation episodes

6.1.3.3 Video stimulated recall interview

The above two methods could identify almost all occurrences of gaze in the sample. But it was still possible that if a student was looking at the top right-hand corner of the screen, she might be looking at her own video image rather than her peer's. This was because the two participants' video frames cover a relatively small area of the screen and were displayed vertically adjacent to each other (see Screenshot 5). To help disambiguate, a third data source, derived from video stimulated recall interviews (VSRI), was used.

The video stimulated recall interviews were conducted within two days of the last video SCMC session. Questions were asked about participant's gaze during negotiated interactions (see Appendix 9). For example, 'Do you care how you look in the camera?', 'Did you look at your own video image and/or your peer's video image very often during task interactions?', 'What information (if any) can you obtain by looking at your peer's video?'. Students' answers to these questions shed further light on the ways in which their gaze was directed and the reasons for their choices of gaze directions.

When asked whether they look at their own video image during task interactions, all students said 'no' or 'not often'. Gaze VSRI Extracts 1, 2 and 3 contain students' answers to this question. D2B stressed that when she was completely engaged in the conversation, she forgot the presence of the camera as she was concentrating on the task interactions.

Gaze VSRI Extract 1: D2B about her own presence in the webcam

Researcher: OK, that's good, do you care a lot about how you look in the camera?

D2B: ehm, to be honest, at the beginning of the talking, maybe I can I care about the looking of my in the picture, when we were very engaged in our talking/conversation, I already forget and ignore how I look in the camera, I was completely focused on the talking, in the beginning, I may look at myself from the camera and make my hair, but in the end, I didn't care how I look from the camera.

Gaze VSRI Extract 2: D1B about her own presence in the webcam

Researcher: OK, do you care how you look in front of the camera?

D1B: ehm, ehm, not really, haha, I didn't care this too much

Gaze VSRI Extract 3: D2A about her own presence in the webcam

Researcher: and did you look at your own video camera very often?

D2A: my own?

Researcher: aha

D2A: no I didn't

Researcher: you didn't, OK, and when you were talking to her, in her video images in the middle of the screen?

D2A: it's not even middle, it's just like this, when I was talking to her, I was trying to see her facial expressions, so I didn't focus too much on my video

When asked what information, if any, they could gain from looking at the peer's video image, students answered that they could gain a certain amount, mainly through interpreting their facial expressions and gestures. Gaze VSRI Extracts 4 and 5 demonstrate students' preference for video-based communication because they could see their peer's facial expressions and gestures, guess their peer's feelings/attitudes, and judge whether or not they understood what was said, all of which ensured that communication run more smoothly.

Gaze VSRI Extract 4: D1B the benefits of video SCMC

D1B: the benefits of video is [sic], when you are communication[ing], you can guess what is your partner's feeling or attitude according to her facial expression or her hand gestures

Gaze VSRI Extract 5: D4B the benefits of video SCMC

D4B: I prefer video.

Researcher: why?

D4B: because I can, we can have eye contacts for better communication, gestures, smile face, and facial expressions video, hmm, I am able to find out, according to her facial expression, if she understand me or not, if she has a problem or not, and if she wants to go on talking or not, and her attitude towards me, when she talks, yes, in this way, communication is better because it's more smooth, clearer.

During the stimulated recall interview with D2A (Gaze VSRI Extract 6), who encountered technical issues and had to use another device to look at her peer's video, the researcher further confirmed the reasons for the direction her gaze during the negotiated interactions. D2A revealed that she was mainly looking at her laptop because it showed her peer's video. She also mentioned that if she had not been obliged to hold the mobile phone with one hand, she might have used more gestures to assist in negotiating meaning.

Gaze VSRI Extract 6: D2A's confirmation of gaze directions

Researcher: yes, OK, OK, and then it was camera, when you were doing video conferencing task, there was a video camera and there was a task sheet, was it hard for you to take care of both

D2A: no, because after I had first rough understanding of task sheet, I don't need to focus all the time on the task sheet, I can see her video naturally, I don't need to focus on the task sheet

Researcher: OK,OK, OK... ahh.. yea... in the second day you watched her video from your laptop and she watched you from your phone, did you hold your phone with your hand?

D2A: yea, with my laptop...

Researcher: so one hand was holding phone, and then did you think that limited you from doing some hand gestures?

D2A: yes, yes, sure

Researcher: OK, OK, right

D2A: if we had a better condition, you know, we can see each other, and I can use my hands, maybe I can use more hand gestures, to make it more, to make it easier for her to understand

Researcher: OK, OK, OK, OK, which camera did you focus more on?

D2A: hmm, the laptop, the laptop

Researcher: the laptop

D2A: yea, because the phone is only for her to see me, so I do not need to focus on the phone

Before each video stimulated recall interview, the researcher watched the video recordings of the online task interactions several times to detect any potential meaning negotiation episodes which merited to the asking of questions and to discover any apparent anomalies or confusing episodes (in terms of students' facial expressions, gestures, gaze) that needed further clarification from participants. For example, the researcher was confused when D1B was looking at the screen with a wide range of up and down eye movements, together with some hand movements accompanied by the sound of typing. During the interaction about 'stationery', D4B looked down but seemingly not in the same direction as she often did while looking at her task sheet. The researcher observed these anomalies while preparing questions for VSRI. Therefore, the interview participants were shown these episodes and were asked what they were doing at that moment. Student D1B confirmed that she had at that point minimised the SCMC system window and was searching the internet to look up the unknown word (Gaze VSRI Extract 7). Student D4A confirmed that she was using her mobile phone dictionary to look up the word 'suite' but did not manage to find it (Gaze VSRI Extract 8).

Gaze VSRI Extract 7: D1B's confirmation of gaze direction

Researcher: so here, you got it from?

D1B: yeah, I looked it up

Researcher: from the dictionary? [online dictionary]

D1B: yeah

Researcher: so you managed to spell it correctly and then found the Chinese meaning

D1B: yeah

Gaze VSRI Extract 8: D4A's confirmation of gaze direction during the 'stationery' MNE

Researcher: what were you doing then?

D4A: look up in the dictionary that I want to describe 一套文具 [a suite of stationeries], but I didn't find it, so I used 'suit'

Researcher: ah, OK, you looked it up in your phone?

D4A: yeah, in my phone

Analysis of the data from video stimulated recall interviews demonstrates students' preference for video communication, confirms their gaze during the negotiated interactions, and clarifies some potential problems and anomalies.

To summarise, three approaches were combined to identify participants' gaze directions during negotiated interactions. First, the researcher gained a general idea of the gaze directions through a pre-task gaze identification test. Then, if students were constantly looking at a particular area of the screen and doing so repeatedly, especially while talking to their peer, their gaze was highly likely to be focused on their peer's video image. Finally, students confirmed this interpretation of their gaze directions and clarified some incongruities during the video stimulated recall interview. Used in combination, these three methods offer strong evidence of a student's gaze direction, and they are especially useful for helping to establish whether a student was looking at their peer's video image or not.

Section 6.1 explained the reasons for undertaking gaze analysis, offered a simple and effective coding scheme and a clear account of each code, and described a triangulated approach to identifying and confirming students' gaze direction. The collection process has generated sufficiently convincing data to lay a solid foundation for gaze analysis. Section 6.2 will focus on reporting how the data was coded.

6.2 The coding process

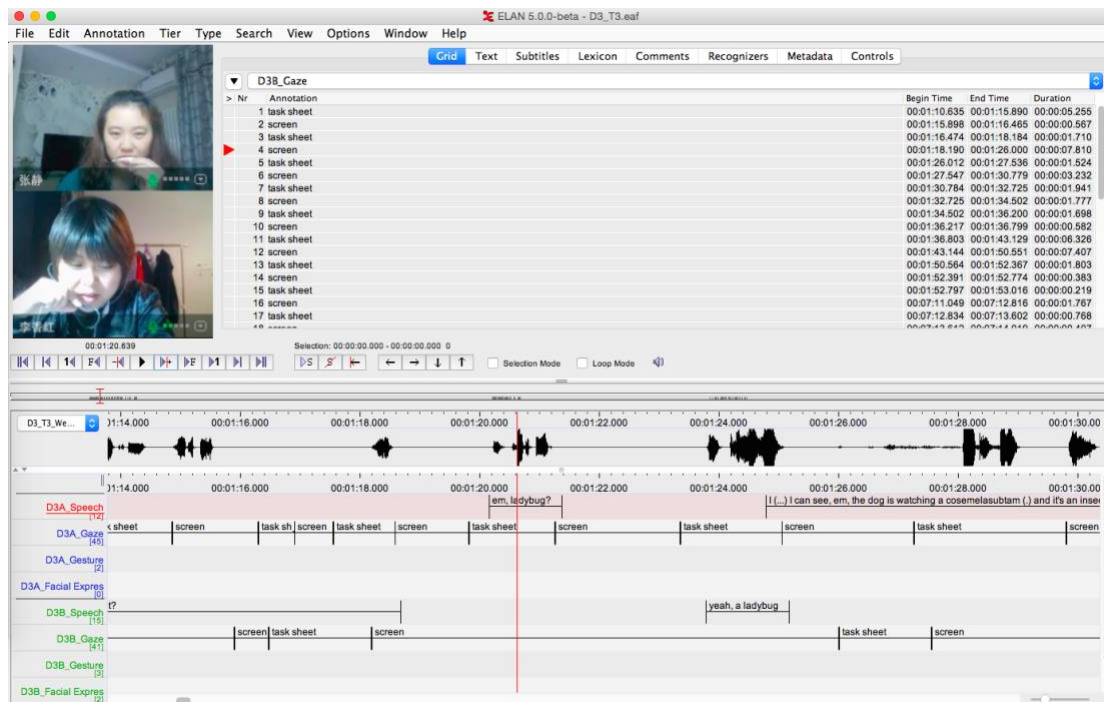
6.2.1 What episodes are included?

Before coding and analysing students' gaze, it is important to first define what episodes are included in the analysis because this decision can directly influence the result of the analysis. In this study, 37 meaning negotiation episodes (MNEs) were identified. These include both successful ones (the respondent reached a correct understanding) and unsuccessful ones (the respondent did not understand the meaning of the negotiated word) which may be incomplete. Analysing all episodes offers a fuller picture of students' gaze and its relationship to meaning negotiation results.

One anomaly worth mentioning is that Dyad 2 did Task 6 instead of Task 5 because on the day of the class, one student could not find the Task 5 sheet. This has no significant influence on the analysis because both Task 5 and Task 6 are of the same type and have a similar level of difficulty. The only difference is the specific lexical items during negotiated interactions.

6.2.2 Coding with ELAN

ELAN is computer software, a professional tool to manually and semi-automatically annotate and transcribe audio or video recordings. It has a tier-based data model that supports multi-level, multi-participant annotation of time-based media ('ELAN_software', 2020). In this study, four tiers were used for each component of multimodal communication: speech, gaze, gesture and facial expression. For ease of coding, the analysis focused only on the two participants' video frames instead of the whole video conferencing interaction. Screenshot 13 shows the screenshot of the specific ELAN annotation interface used in this study to carry out multimodal annotation and code participants' gaze directions. The gaze analysis focused on the coding of gaze in meaning negotiation episodes.



Screenshot 13: The interface of ELAN multimodal annotation process

The coding process involved watching the video in ELAN frame by frame, then selecting one of the four available codes for the direction of gaze and attributing it to the relevant section of the video. Each frame had a duration of 0.013 seconds. This ensured that coding could be extremely fine-grained. The starting and ending time of each coded gaze and the duration for these gazes were recorded and exported to Excel for statistical analysis.

6.3 Statistical gaze analysis and findings

6.3.1 The overview of data

In total, the gaze analysis included codes for 1010 gaze movements by four dyads in all meaning negotiation episodes in video SCMC. The overall coded time was one hour, one minute, and 3.585 seconds, translated into 3663.585 seconds. Table 17 summarises each dyad's gaze data during negotiation for meaning episodes.

The overall time of task interaction for each dyad is presented in the last column. During the task instruction, the teacher emphasized that there were no time limits for completing these tasks. Therefore, students were under no time pressure. From the table, it could be seen that different dyads used different amount of time in negotiated interactions. For example, Dyad 2 spent 2262 seconds (37 minutes 42 seconds) in Task 6 because there were some technical issues and other practical issues (e.g. one participant could not find the task sheet), while Dyad 1 only spent 407 seconds (6 minutes 7 seconds). The data reflected students' natural performances in video SCMC task interactions. The different lengths of task interaction and negotiation for meaning were fully controlled by interlocutors. Such differences might be caused by many factors including students' attitudes towards or patience in meaning negotiation, technical issues, the physical environment they were in during the task interaction and their personal arrangements after the online session. For example, one participant was in a noisy public cafe, and another had to look after her child during the online session, all of which might have affected their time spent on task interactions. Therefore, it was not possible to try to make sense of why each dyad spent different amount of time on different tasks and try to relate this task interaction time to their competence in negotiation for meaning.

On average, during interactions devoted to the negotiation of meaning, students spent more than half the time (53.91%) looking at their peer's video image (PVI), while 38.58% of their time was occupied by looking at the task sheet (TS), and only 7.26% of the time looking in other directions (OD). The 'unidentifiable' (UI) category took up only 0.25% of their time and had no discernible influence on the result of the analysis.

Dyad_Task	PVI (s)	PVI%	TS (s)	TS%	OD (s)	OD%	UI (s)	UI%	NfM time (s)	Task time (s)
D1A_T3	26.973	40.60%	37.846	56.97%	1.613	2.43%	0.000	0.00%	66.432	407
D1B_T3	14.848	21.97%	40.972	60.62%	11.772	17.42%	0.000	0.00%	67.592	
Sum_D1_T3	41.821	31.20%	78.818	58.81%	13.385	9.99%	0.000	0.00%	134.024	
D1A_T5	22.988	27.99%	50.128	61.03%	7.772	9.46%	1.245	1.52%	82.133	1029
D1B_T5	22.686	27.53%	49.840	60.48%	9.883	11.99%	0.000	0.00%	82.409	
Sum_D1_T5	45.674	27.76%	99.968	60.76%	17.655	10.73%	1.245	0.76%	164.542	
D2A_T3	94.843	68.52%	38.064	27.50%	5.519	3.99%	0.000	0.00%	138.426	1282
D2B_T3	36.479	26.29%	84.115	60.63%	18.149	13.08%	0.000	0.00%	138.743	
Sum_D2_T3	131.322	47.38%	122.179	44.08%	23.668	8.54%	0.000	0.00%	277.169	
D2A_T6	646.524	72.49%	122.067	13.69%	85.598	9.60%	37.728	4.23%	891.917	2262
D2B_T6	554.407	62.08%	262.147	29.35%	76.488	8.56%	0.000	0.00%	893.042	
Sum_D2_T6	1200.931	67.28%	384.214	21.53%	162.086	9.08%	37.728	2.11%	1784.959	
D3A_T3	51.513	50.68%	50.125	49.32%	0.000	0.00%	0.000	0.00%	101.638	987
D3B_T3	36.902	36.30%	63.512	62.48%	1.235	1.21%	0.000	0.00%	101.649	
Sum_D3_T3	88.415	43.49%	113.637	55.90%	1.235	0.61%	0.000	0.00%	203.287	
D3A_T5	225.856	91.47%	12.370	5.01%	8.695	3.52%	0.000	0.00%	246.921	1149
D3B_T5	143.773	58.33%	86.933	35.27%	15.774	6.40%	0.000	0.00%	246.480	
Sum_D3_T5	369.629	74.91%	99.303	20.13%	24.469	4.96%	0.000	0.00%	493.401	
D4A_T3	15.913	7.08%	206.086	91.75%	2.627	1.17%	0.000	0.00%	224.626	1257
D4B_T3	13.937	6.19%	211.066	93.81%	0.000	0.00%	0.000	0.00%	225.003	
Sum_D4_T3	29.850	6.64%	417.152	92.78%	2.627	0.58%	0.000	0.00%	449.629	
D4A_T5	7.866	10.06%	50.556	64.63%	19.798	25.31%	0.000	0.00%	78.220	961
D4B_T5	29.610	37.79%	47.734	60.92%	1.010	1.29%	0.000	0.00%	78.354	
Sum_D4_T5	37.476	23.94%	98.290	62.78%	20.808	13.29%	0.000	0.00%	156.574	
Sum	1945.118	53.09%	1413.561	38.58%	265.933	7.26%	38.973	1.06%	3663.585	9334

Table 17: Gaze information in meaning negotiation episodes in video SCMC

6.3.1.1 Gaze time on the peer's video

In 37 meaning negotiation episodes, on average, participants spent slightly more than half of the time (53.91%) studying their peer's video image on the screen. But the percentage varied largely for different participants in different tasks. In particular, in Task 5, D3A spent 91.47% of the time looking at her peer's video image, while in Task 3, the figure for D4B was only 6.19%. There also existed significant differences between the time spent looking at their peer's video image both within dyads and between different dyads. Comparing different dyads, it could be seen that Dyad 2 (22 minutes 12seconds; 57.33%) and Dyad 3 (7 minutes 38 seconds; 59.2%) spent more time overall looking at each other's video image than Dyad 1 (1 minute 26 seconds; 29.48%) and Dyad 4 (1 minute 7 seconds; 15.29%) in both tasks. Within each dyad, there was a consistent pattern that Student A spent more time looking at her partner's video image than Student B in both tasks, with the only exception being that D4B outperformed D4A on this measure in Task 5.

In terms of task type, for all four dyads, the overall time each dyad spent looking at each other's video image was longer in the problem-solving task (Task 5 or Task 6) than in the spot-the-difference task (Task 3). For example, Dyad 2 spent over 20 minutes looking at each other's video image in Task 6, a problem-solving task, but only 2 minutes 11 seconds in Task 3, a spot-the-difference task. This could be explained in terms of the fact that the spot-the-difference task sheet (which was a picture full of small details that can be potential differences) had more intensive information than the problem-solving task sheet (which only had four items for students to describe and choose from).

6.3.1.2 Gaze directed at task sheet

On average, four dyads spent 38.58% of their time looking at the task sheet in all 37 meaning negotiation episodes. Since gaze at the task sheet and at peer's video image were the two most frequently used gaze directions, those dyads who spent less time looking at the peer's

video image, such as Dyad 1 and Dyad 4, tended to spend more time on the task sheet. For example, Dyad 4 in Task 3 spent more than 92.78% of their time looking down at their task sheet. For the same task, the figure for Dyad 2 was less than half of that (44.08%). With regard to task type, it could be seen that three out of four dyads (the exception being Dyad 1) spent a higher percentage of time looking at the task sheet in Task 3 (a spot-the-difference task) than Task 5 or 6 (both problem-solving tasks). For example, Dyad 4 spent 6 minutes 57 seconds looking at the spot-the-difference task sheet while only 1 minute 38 seconds on the problem-solving task sheet. This result is consistent with the result in Section 1.1.

6.3.1.3 Gaze in other directions

Despite the different amounts of time spent looking at their peer's video image and the task sheet, most participants spent less than 10% of their time looking elsewhere. Only two students, D1B and D4A spent more than 10% of their time looking in other directions. This could be accounted for in terms of their regular consultation of online dictionaries. The specific scenario will be studied in further details in the multimodal analysis chapter.

6.3.2 Statistical analysis

6.3.2.1 Variables

The goal of this gaze analysis is to explore the role of video in SCMC negotiated interactions. Specifically, in carrying out a regression analysis, the aim is to establish to what extent looking at an interlocutor's video image in meaning negotiation episodes can contribute to the success of meaning negotiation. Therefore, the X variable, which, in regression analysis, is the predictor or explanatory variable, is the amount of time each dyad spent looking at each other's video image during MNEs. It is important to stress that the time is the sum of the time spent in doing this by both students in the dyad. This is because the interaction is mutual, and the interactants cannot be separated. When one student is looking at her peer's video, she can see all of her

peer's non-verbal behaviour, including gaze, facial expressions, gestures, and so on. It does not matter if one student is looking at her peer's image and the other is not. For example, if Student A was looking at the B's image but Student B was not looking at Student A. Student A could still see Student B's modes of communication other than gaze, such as nodding, smiling, frowning, sitting forward. All of this multimodal information offered evidence for Student A to make judgements on whether Student B understood what she was talking about. In cases when both students were looking at their peer's video image, they could both see each other's multimodal information, which offered them a further indication of each other's (non-)understanding of the negotiated lexical item.

All the multimodal elements including gaze, facial expressions and gestures may potentially help students to decide how to move forward with the meaning negotiation process. For example, if the initiator saw a frown on her peer's face or observed her suddenly bending forward towards the screen, she might decide to offer a lengthier explanation of the negotiated item. But if the initiator saw a nod or a smile from the respondent's video, she might decide either to ask for confirmation or directly move on to the task-related discussion. In contrast to a qualitative multimodal analysis, which examines in detail a number of specific cases in which students made use of multimodal information in meaning negotiation, a regression analysis could offer a statistical calculation of the extent to which looking at a peer's video image correlates with success in meaning negotiation.

	A	B	C	D	E	F	G	H
1	Dyad	word	successful		Dyad	No. of MNE	No. of Success	Percentage
2	D1_T3	torch	no		D1_T3	5	0	0.00%
3		frickle	no		D1_T5	2	0	0.00%
4		sailing model	no		D2_T3	2	2	100.00%
5		glass of salt	no		D2_T6	9	6	66.67%
6		ladybug	no		D3_T3	3	0	0.00%
7	D1_T5	stationary	no		D3_T5	6	3	50.00%
8		razor	no		D4_T3	8	2	25.00%
9	D2_T3	trash can	yes		D4_T5	2	2	100.00%
10		toy rocket	yes		Sum	37	15	40.54%
11	D2_T6	portable	no					
12		blender	no					
13		nail polish	yes					
14		portable clothes rack	yes					
15		blender	no					
16		toaster	yes					
17		piggie	yes					
18		boardgame	yes					
19		phone accessories	yes					
20	D3_T3	ladybug	no					
21		tin	no					
22		dustbin	no					
23	D3_T5	bouuet	yes					
24		razor	yes					
25		magnifying glass	no					
26		cosmetics	yes					
27		rubik's cube	no					
28		magnifying glass	no					
29	D4_T3	study	yes					
30		ladybug	yes					
31		lampshade	no					
32		salt shaker	no					
33		sandwich	no					
34		can	no					
35		brim	no					
36		sneakers	no					
37	D4_T5	bouquet	yes					
38		suite	yes					

Table 18: The Y variable: the number of successful meaning negotiation episodes

The Y variable in this analysis is the number of successful meaning negotiation episodes. Table 18 lists all 37 meaning negotiation episodes in eight video SCMC tasks by four dyads and the results of these meaning negotiation episodes. In this study, a successful meaning negotiation refers to the meaning negotiation episodes in which the respondent (the student who at the outset did not know the meaning of the lexical item) managed to arrive at the correct meaning of the lexical item through negotiating with their peer. Using this criterion, 15 of our 37 meaning negotiation episodes were successful, including eight from Dyad 2, three from Dyad 3, four from Dyad 4, and none from Dyad 1.

6.3.2.2 Regression analysis result

The above paragraphs explained what the X and the Y variables represent. The final data for these two variables are summarised in Table 19.

Dyad_Task	Time of gaze directed at peer's video image (by second) (the X variable)	No. of successful MNE (the Y variable)
D1_T3	41.821	0
D1_T5	45.674	0
D2_T3	131.322	2
D2_T6	1200.931	6
D3_T3	88.415	0
D3_T5	369.629	3
D4_T3	29.85	2
D4_T5	37.476	2

Table 19: Final data for regression analysis of gaze directions and meaning negotiation success

With this set of data as input, Excel is used to calculate the correlation coefficient between the two variables, generate a trend line and equation (see Figure 13), and conduct a regression analysis (Table 20). The following paragraphs will report the statistical results and explain their meaning in this context.

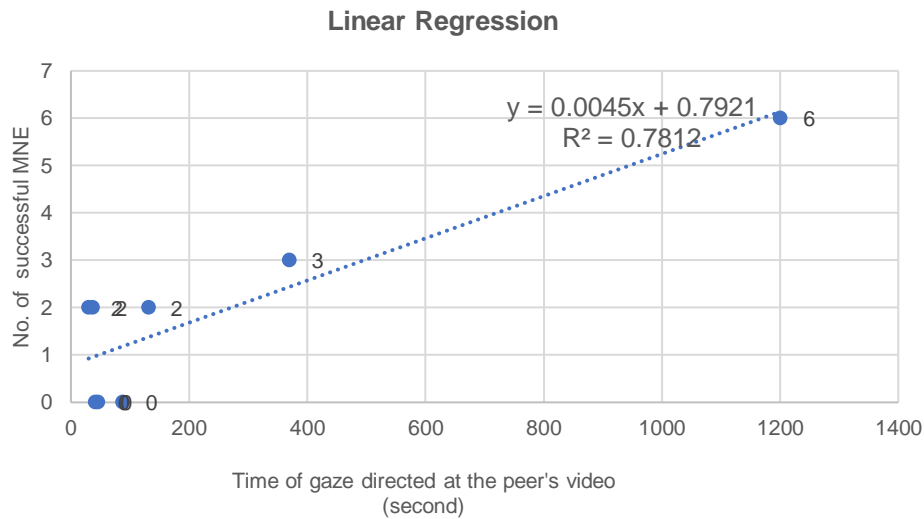


Figure 13: Linear regression between the time of gaze directed at peer's video image

SUMMARY OUTPUT								
<i>Regression Statistics</i>								
Multiple R	0.88386819							
R Square	0.78122299							
Adjusted R Square	0.74476015							
Standard Error	1.0260918							
Observations	8							
<i>ANOVA</i>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	1	22.5578137	22.5578137	21.4251845	0.003582438			
Residual	6	6.3171863	1.05286438					
Total	7	28.875						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	0.79206701	0.43167667	1.83486177	0.11620264	-0.26420774	1.84834177	-0.26420774	1.84834177
X Variable 1	0.00445395	0.00096224	4.62873466	0.00358244	0.002099437	0.00680847	0.002099437	0.00680847

Table 20: The regression analysis result between gaze time spent on peer's video image and the number of successful MNEs

6.3.2.2.1 Correlation coefficient

The correlation coefficient measures the degree to which two variables are linearly related. The range of the correlation coefficient can be from -1 to 1, with 1 indicating a completely positive

linear relationship and -1 indicator a completely negative linear relationship. It is also called the Multiple R in the regression analysis. In the present analysis, the correlation coefficient is 0.88, which is a relatively strong correlation. This indicates that the time spent by students looking at their peer's video image is closely related to the number of successful meaning negotiation episodes. The next step of regression analysis can further specify the relationships between these two variables from a statistical perspective.

6.3.2.2.2 R square (R^2)

In regression analysis, R-squared measures the extent to which change in the X variable contributes to change in the Y variable. In this case, the change of time of gaze directed at the peer's video image contributes to 78% of the change of the number of successful MNEs. R squared serves to predict the likelihood of future events falling within the predicted outcomes. So, another way of explaining the meaning of R squared is that, if we conduct the same experiment (video SCMC tasks) in a bigger sample (with more dyads) in the same population, there is a 78% chance that the number of successful MNE can be predicted by this model (the equation/trend line in Figure 13). This is also a relatively strong prediction in statistical terms.

6.3.2.2.3 P-value

Another important measure is the p-value, which is used to determine the statistical significance of a hypothesis test. In other words, it indicates to what extent the result of the regression occurs randomly. The smaller the p-value is, the more significant the result is. The accepted threshold for determining statistical significance is 0.05. If the p-value is smaller than 0.05, it means we can say with 95% confidence that the regression analysis result is statistically significant. In this study, the p-value is 0.0035, much smaller than 0.05, and even smaller than 0.01. This means that the result of the regression analysis is statistically significant because we can say with 99% of confidence that the result did not occur randomly. This p-value can also be used to predict the result of future events. In this case, it means if we repeat the experiment (video SCMC

tasks), it is highly likely (with 99% confidence) that we can obtain a similar result in the regression analysis. In other words, the chances of the sample data occurring randomly are extremely low (less than 0.01).

6.3.3 Other possible factors

The above analysis demonstrated the strong correlation between the time spent on looking at the peer's video image and the success in negotiation for meaning. But further evidence is needed to clarify which of the two factors is the cause and which is the result. This problem can be partly answered by students' comments on video SCMC during the stimulated recall interview. For example, D4B said she preferred video to audio SCMC because she could see her peer's 'gestures, smile face, and facial expressions' and tell if her peer understood her or not. Similarly, D1B commented, '*the benefits of video is [sic], when you are communicating, you can guess what is your partner's feeling or attitude according to her facial expression or her hand gestures*' (see Section 6.1.3.3). The interview result shows that most participants preferred video to audio SCMC because when looking at their peer's video image, they were able to obtain more multimodal information, which could promote their meaning negotiation process and make the communication move on smoothly. The video stimulated recall interview offers some qualitative evidence for the statistical result and confirms time spent on looking at the peer's video image is the cause of success in meaning negotiation, rather than the other way around.

Another possible factor affecting the success of meaning negotiation could be interlocutors' linguistic proficiency. It could be assumed that students with a higher level of proficiency could afford to focus more on the peer's video image and succeed in meaning negotiation in video SCMC. Table 21 lists each dyad's score in the mock IELTS speaking test and the number of successful MNEs in video SCMC. Although Dyad 1 has a higher IELTS overall score as well as a higher difference among the two interlocutors, they did not manage to negotiate meaning in video SCMC. However, the other three dyads all succeeded in some MNEs. This table

illustrates there does not seem to exist any statistically significant relationship between students' oral proficiency and their success in meaning negotiation.

Dyad	Student A	Student B	No. of successful MNE
Dyad 1	7	5.5	0
Dyad 2	6.5	5	8
Dyad 3	5.5	6	3
Dyad 4	5.5	6.5	4

Table 21: The mock IELTS speaking test scores and the number of successful MNE in video SCMC

The previous section focuses exclusively on how gaze time spent on the peer's video image may affect the success in negotiation for meaning. What has not been discussed is whether there exists any statistically significant relationship between the time spent on the task sheet and the success in meaning negotiation. Table 22 lists the time spent on looking at the task sheet by each dyad in each task and the number of successful MNEs they achieved. A regression analysis (Table 23) is carried out to explore the statistical relationship between these two factors. For the result to be statistically significant, the p-value should be less than 0.05. However, the p-value in this regression analysis is 0.1732933, much bigger than 0.05. Therefore, there doesn't exist any statistically significant relationship between the two factors. In addition, the low correlation coefficient (0.578852) and R square (0.335082) also illustrate that the gaze time at the task sheet has no statistical relationship with the number of successful MNEs.

Dyad_Task	Time looking at TS (s)	No. of successful MNE
D1_T3	78.818	0
D1_T5	99.968	0
D2_T3	122.179	2
D2_T6	384.214	6
D3_T3	113.637	0
D3_T5	99.303	3
D4_T3	417.152	2
D4_T5	98.29	2

Table 22: Gaze time on the task sheet and the number of successful MNEs

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.578862							
R Square	0.335082							
Adjusted R Square	0.202098							
Standard Error	1.818129							
Observations	7							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	8.329175102	8.329175	2.519721	0.17329316			
Residual	5	16.52796776	3.305594					
Total	6	24.85714286						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.583168	1.199026184	0.486368	0.647274	-2.49902692	3.665363	-2.499027	3.665363
X Variable 1	0.00818	0.005153024	1.587363	0.173293	-0.00506655	0.021426	-0.005067	0.021426

Table 23: The regression analysis result between gaze time spent on peer's video image and the number of successful MNEs

6.3.4 Final statistical findings

Therefore, based on the above statistical analysis results, it could be concluded that the more time students spent looking at their peer's video image during negotiated interactions, the more successful they were in meaning negotiations. Although the sample data was limited to only eight video tasks, the regression analysis is statistically significant and could be used to predict the results of future/repeated experiments. It should be noted that the result is based on gaze analysis in meaning negotiation episodes only, rather than in the full video SCMC interactions, as this is the focus of the research question.

The X variable in this analysis is the actual amount of time (how many seconds) each dyad spent looking at each other's video image. It is not the ratio or percentage of time spent looking at the peer's image as opposed to the overall time for meaning negotiation (which also includes time spent on looking at task sheets and in other directions). This means that it does not matter how much time students spent looking at the task sheet or in other directions. As long as they spent more time looking at their peer's video during negotiated interactions, it was likely that

they would be more successful in negotiating for meaning. In fact, the same regression analysis was carried out to examine the relationship between the time spent looking at the task sheet and the success of meaning negotiation. But no statistically significant result was found between these two variables.

Based on students' actual gaze movements during the negotiated interactions, these statistical findings in relation to gaze analysis provide a partial answer to the second research question: 'What roles do multiple modes and semiotic resources play in meaning negotiation episodes in video SCMC?'. The statistical findings are triangulated with another source of data: students' responses to stimulated recall interviews about the use of video in SCMC interactions, as briefly reported earlier. The combination of both qualitative and quantitative analysis offered a comprehensive understanding of the role of the visual channel in video SCMC interactions.

This result is possibly one of the first attempts to manually code students' gaze in video SCMC and to conduct a statistical analysis to explore the role of video in negotiated SCMC interactions. Therefore, it needs further examination in a variety of research contexts and with different task designs. The gaze coding and analysis procedures were so demanding and time-consuming that it is almost impossible to repeat this experiment and analysis in a substantially bigger sample. An alternative might involve using eye-tracking devices and analysis software. In this study, eye-tracking methods were not used mainly because of the lack of these technical devices in the research context and the limited scale of a doctoral research project.

6.4 Chapter conclusion

This chapter first offered a rationale for the use of gaze analysis in this study and introduced the coding scheme and method. Then, the coding process was presented, and the statistical findings were reported and analysed. The key finding of the chapter is that the researcher identified a statistically significant positive correlation between the amount of time interlocutors

spent looking at their peers' video image and the number of successful meaning negotiation episodes in video SCMC. In summary, the more time each dyad spent looking at each other's video image, the more likely it was that they were successful in meaning negotiation. This finding offers statistical evidence of the important role played by the visual image in multimodal video SCMC. To complete the answer to research question 2, Chapter 8 will offer a multimodal analysis of four meaning negotiation episodes in video SCMC.

7 Multimodal Analysis and Findings

7.1 Data analysis methods

The previous chapter proved that the visual communication plays an important role in meaning negotiation in video SCMC. Complementing the findings of the statistical analysis of gaze in Chapter 6, the present chapter will continue to answer the second research question in this dissertation: 'What roles do multiple modes and semiotic resources play in meaning negotiation episodes in video SCMC?'. The answer will be achieved in this context by analysing the multimodal interactions of specific meaning negotiation episodes.

This chapter starts by offering the rationale for- and presenting the methods used in multimodal analysis. Then the multimodal analysis of four meaning negotiation episodes will be presented. Data from participants' answers to video stimulated recall interviews (VSRI) will be used to explore their thoughts at certain points of negotiated interactions. Next, in the findings section, various levels of multimodal communicative competence will be identified and the relationships between different modes and semiotic resources will be illustrated with examples.

7.1.1 Rationale for the multimodal analysis

As was pointed out in the literature review, existing meaning negotiation studies in video SCMC agree on the importance of multiple modes of semiotic resources but have not used multimodal analysis to support the argument (e.g. Wang, 2006; Lee, 2006; Wang and Tian, 2013; Guo and Möllering, 2016). Meanwhile, those who have carried out a multimodal analysis have not been particularly or extensively focused on meaning negotiation episodes (Develotte, et al., 2010; Satar, 2015, 2016; Guichon and Wigham, 2016; Cohen and Wigham 2019). Therefore, this research gap needs to be filled by evidence from a multimodal analysis of meaning negotiation episodes in video SCMC. Gaze analysis has demonstrated that there exists a statistically

significant positive correlation between the amount of time interlocutors spend looking at their peers and the success of meaning negotiation. It can be deduced that participants gain important information necessary for the success of meaning negotiation through looking at their peer's video image. This visual information also affects participants' verbal communication and their own multimodal performances. Therefore, it is necessary to look into how participants make use of different modes and semiotic resources to negotiate in video SCMC. A multimodal analysis of some specific meaning negotiation episodes will offer insights into the role of multiple modes and semiotic resources in video negotiated interactions.

7.1.2 What episodes are included for the multimodal analysis?

Although there is a large amount of literature studying meaning negotiation in different contexts, almost all of it focuses on complete meaning negotiation routines as part of successful meaning negotiation episodes. Very few publications deal with unsuccessful meaning negotiations, although not all meaning negotiation episodes are successful. However, in the data, there are a substantial number of unsuccessful or incomplete meaning negotiation episodes, where the communication breakdown, or the non-understanding, was unresolved but the participants moved on with their conversation. Therefore, the researcher seeks to look into a broader range of meaning negotiation episodes.

In this thesis, it is acknowledged that meaning negotiation episodes also include those incomplete routines, as long as they contain a clear communication breakdown or a non-understanding. A successful meaning negotiation episode means the non-understanding is resolved at the end of the interaction, usually through a complete meaning negotiation routine. An unsuccessful meaning negotiation episode can be either (a) a complete meaning negotiation routine which nonetheless results in a misunderstanding or (b) an incomplete meaning negotiation routine (e.g. one which lacks an explicit indicator, or the remaining meaning negotiation stages). A key element for identifying the second type of unsuccessful meaning negotiation episode is that one interlocutor is aware of the non-understanding and that

there is adequate evidence in the interaction itself, or from a subsequent stimulated recall interview, to prove the non-understanding.

Multimodal analysis, together with video stimulated recall interview data and comparison of successful and unsuccessful meaning negotiation episodes, can shed light on what actually makes the difference in the negotiation process in the video SCMC environment. Is it the students' attitude towards meaning negotiation? Or is it technical limitations? Or is it task design issues? Knowing the answers to these questions can help researchers and online teachers to understand what methods can be useful for promoting successful meaning negotiation and SLA in video SCMC environments?

7.1.3 The critical incident approach

There are 37 meaning negotiation episodes (including successful and unsuccessful ones) in the video task interactions of all four dyads. Since multimodal analysis is extremely fine-grained, it is impossible within the scope of a thesis to analyse and present the multimodal analysis of all 37 episodes. Therefore, critical incident technique is used to identify one case of meaning negotiation episodes from each dyad for the multimodal analysis.

In this particular study, critical incidents are defined as follows. Episodes are selected for analysis if students are explicit about their use of the visual mode or talk about the use of modality during meaning negotiation. This is because these episodes show students' conscious awareness of the modalities available to them, as well as their capacity for choosing appropriate modalities to facilitate their negotiated interaction. In such cases, students' attention tends to focus on particular modes. These episodes are critical in enabling the researcher to examine how multimodality is purposefully used by participants to facilitate the meaning negotiation process. These critical incidents are easy to identify because they involve explicit reference by participants to the mode(s) they use.

However, it is equally important to examine those episodes where multimodal elements are neglected by interlocutors during negotiation in video SCMC. For example, an interlocutor may appear confused, indicating her non-understanding, but the peer may not notice it since she was not looking at the screen. Or a student makes gestures related to meaning negotiation but these are not clearly visible on the screen. Examples like these are important for the study as they show how opportunities for successful meaning negotiation might be missed because of inadequate use of multimodal resources. Clearly, it is hard to imagine or assume what would have happened if multimodal resources were properly used. But by comparing and contrasting such episodes with successful meaning negotiation episode where multimodal resources were used, the analysis may generate insights into the difference that multimodality makes in meaning negotiation episodes in video SCMC.

These selection criteria can help identify those episodes where multimodality makes a difference to meaning negotiation, thus discovering the role of the multimodality in video SCMC. Following the criteria, one episode was selected from each dyad. Since a statistical analysis of gaze has found that the more time an interlocutor spends looking at their peer's video image ('screen time' in short), the more likely they are to succeed in meaning negotiation, the current multimodal analysis scrutinises in detail four selected episodes according to the screen time they occupy, from shortest to longest (see Table 24). The sequence in which they are dealt with will be as follows: (1) Dyad 4 Task 3 saltshaker; (2) Dyad 1 Task 5 razor; (3) Dyad 3 Task 5 magnifying glass and (4) Dyad 1 Task 6 nail polish.

Dyad	Task and lexical item	Individual screen time	Average screen time	Meaning negotiation outcome
D1A	Task 5 Razor	23s	23s	Unsuccessful
D1B		23s		
D2A	Task 6 Nail polish	10min 46s	10min	Successful
D2B		9min 14s		
D3A	Task 5 Magnifying glass	3min 45s	3min 4s	Unsuccessful
D3B		2min 23s		
D4A	Task 3 Saltshaker	16s	15s	Unsuccessful
D4B		14s		

Table 24: Screen time for each dyad in their selected meaning negotiation episodes

7.1.4 Multimodal analysis and transcripts

In multimodal analysis, it is necessary for the researcher to include a multimodal transcript. Traditionally, in conversation analysis, transcription is considered to be the first step or the basis for the analysis. However, in multimodal analysis, there are so many modes and semiotic resources that it is impossible to include all the information in the transcript. Therefore, researchers in the field of multimodal studies in SCMC tend to use multimodal annotation software such as ELAN or Atlas.ti to display and analyse the multimodal data. In this case, multimodal transcription is no longer used as a basis for the analysis, instead, a streamlined multimodal transcript is used by researchers to demonstrate the outcomes of the multimodal analyses to the readers (Develotte, et al, 2010; Guichon and Cohen, 2014; Satar, 2016; Lee et al., 2019). This thesis follows accepted practice in the field.

Despite general agreement on the use of annotation software for multimodal analysis in the field, there are no well-established ways of presenting multimodal transcripts. Researchers still need to develop the way that best suits their own study according to their specific research context, research design and research questions.

In this study, the multimodal analysis of meaning negotiation episodes followed 3 steps: annotation, analysis, and transcript production. To begin with, the researcher followed the tier-based approach in ELAN to annotate participants' performances in multiple modes and with different semiotic resources in video SCMC. Four tiers were used to annotate each participant's performances in meaning negotiation episodes, including speech, gaze directions, gestures, and facial expressions. In the 'speech' tier, students' verbal speech was transcribed following a verbatim approach. The 'gaze' tier included four codes: peer's video image, task sheet, other directions, and unidentifiable gaze directions, as has been described in Section 6.1.2. Moreover, other multimodal elements such as body and hand gestures, and facial expressions such as smiles and frowns were also described with written words in detail in the 'gesture' and 'facial expressions' tiers respectively.

Next, in the analysis stage, two key topics were analysed with the help of ELAN annotations: the meaning negotiation patterns and the multimodal interactions. On one hand, the stages of each meaning negotiation episode were analysed based on the models by Varonis and Gass (1985) and Smith (2003). The author also kept an open mind in identifying any new stages in meaning negotiation in video SCMC. This analysis could contribute to answering the first research question on how students negotiate meaning in video SCMC. On the other hand, the author repeatedly scrutinised all the tiers of annotations both horizontally and vertically to identify any inter-relationships among participants' performances in multiple modes and with different semiotic resources. In other words, the researcher aimed to identify how multiple modes and semiotic resources were used by participants in different combinations to express themselves multimodally. For example, the researcher would pay attention to where a participant was looking at and what gestures or facial expressions she performed while explaining the meaning of a word to her peer. This analysis could provide answers to the second research question about the roles multiple modes and semiotic resources play in meaning negotiation episodes in video SCMC.

Finally, a transcript for each MNE was produced based on the annotations in different tiers and the analysis of the meaning negotiation stages and the multimodal interactions. The aim of the transcript is for the author to clearly present the multimodal interactions to the readers, rather than for the analysis. Some screenshots were also made at this stage to demonstrate some important visual elements during the MNEs, such as gestures and facial expressions used by participants to negotiate meaning of the unknown words. By combining the transcripts and the screenshots, the researcher was able to offer a clear representation of the multimodal interactions in the paper-based thesis.

This three-step method of multimodal analysis was chosen because it could offer a systematic approach for the researcher to note down all the relevant detailed information in multimodal meaning negotiation episodes. In other words, this approach could enhance the validity of the multimodal analysis. Meanwhile, it was also effective and efficient in helping the researcher to

clearly analyse the inter-relationships among different modes and semiotic resources used in multimodal interactions, which was the most challenging yet important task in the multimodal analysis. In addition, as has been demonstrated above, the use of multimodal annotation software could help the researcher to produce line-by-line multimodal transcripts in which all the salient multimodal information is fully provided to the readers.

7.2 Multimodal analysis

7.2.1 Dyad 4 Task 3: 'saltshaker' episode

This meaning negotiation episode is from Dyad 4's attempt to carry out Task 3. In this task both students were given a similar picture, but with seven minor differences. Students were asked to describe their picture to each other in detail and establish the differences between the two pictures (see Video MNE Extract 1).

Video MNE Extract 1: Dyad 4 Task 3: 'saltshaker' episode

Turn	Conversation	Meaning negotiation stage	Gaze direction	Multimodal components
1	D4B: ehm, there's a salt shaker on the chair (.) salt shaker made of glass in my picture	Trigger, explanation,	both task sheet	screenshot 14
2	D4A: butterfly in my picture (technical noise)	Task related discussion	both task sheet	/
3	D4B: I did not hear you, last sentence	Negotiation of modality	D4A: task sheet D4B: peer's video image	screenshot 15.1
4	D4A: is there, ehm, is there a butterfly on the chair in your picture?	Task related discussion	D4A: glance at peer's video image, mostly task sheet D4B: task sheet	screenshot 15.2 and 15.3

5	D4B: no, my one is a salt shaker	Trigger	both task sheet	/
6	D4A: maybe it's the fourth	Task related discussion	both task sheet	/
7	D4B: maybe it's the fourth difference	Task related discussion	both task sheet	D4A: D4B: taking notes on the task sheet

7.2.1.1 Turn-by-turn multimodal analysis of the episode

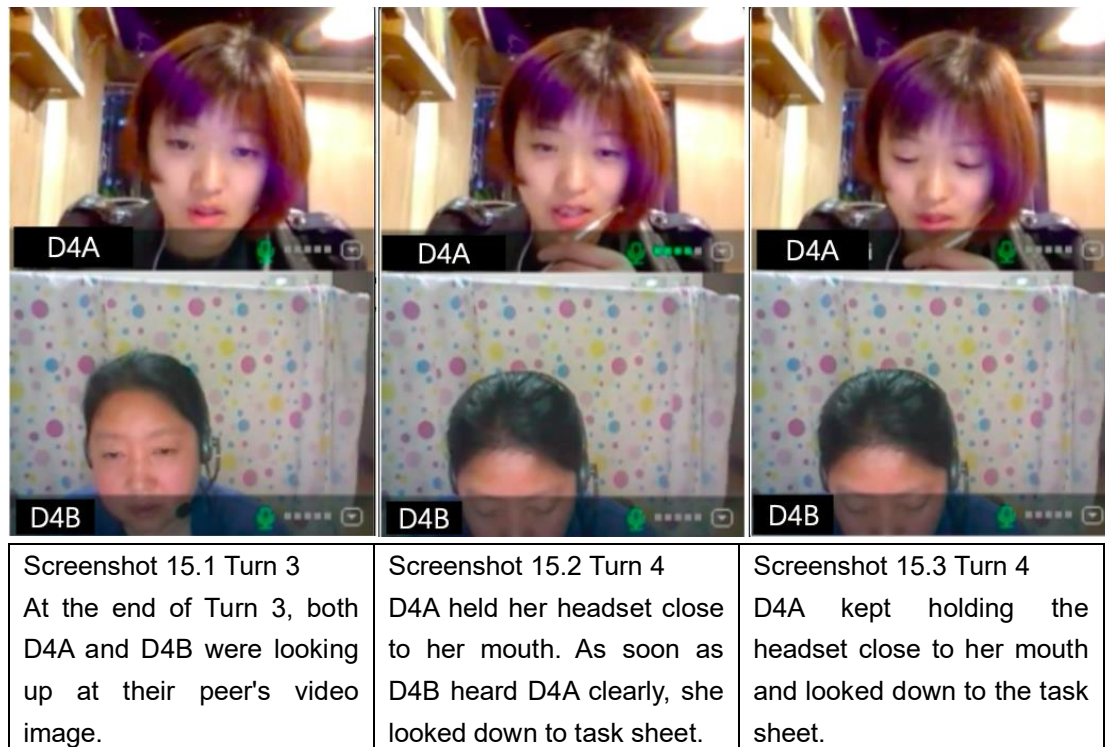
In this episode, Student D4B first introduced a trigger, a 'salt shaker', in Turn 1, after a short pause, she repeated the trigger 'salt shaker' and offered a short explanation 'made of glass'. She ended her sentence with 'in my picture', which demonstrated that her focus was on the task sheet. In fact, it was not hard to use gestures to explain the meaning of a 'salt shaker'. And it seemed that D4B wanted to explain the meaning of the word. But she was totally committed to looking down at her task sheet, and as a result, she showed no attempt to use the video channel to explain the meaning of the word to her peer. Furthermore, her camera was positioned in such a way that her face occupied only the lower two thirds of the frame and the rest of her body was out of frame, which made it almost impossible for her partners, D4A, to see her facial expression, or gestures clearly (Screenshot 14).



Multimodal components: D4B was saying 'salt shaker', made of glass, while looking down at her task sheet. D4A was listening to D4B and also looking down at the task sheet. D4B's camera could only capture her head. D4A was listening to D4B and looking down at her task sheet.

Screenshot 14: D4_T3_salt shaker_Turn 1

D4A did not follow D4B's identification of a 'salt shaker'; instead, in Turn 2, she said there was a 'butterfly in my picture'. But there was some noise during Turn 2 due to technical interference, so in Turn 3, D4B explicitly pointed out that she hadn't heard D4A clearly and added 'last sentence' to further clarify which part she did not hear. Meanwhile, D4B looked up at her peer's video image while reporting the technical issue. Otherwise, both D4A and D4B were mostly looking at their task sheet during this task.



Screenshot 4: D4_T3_salt shaker_Turn 3&4

Soon after D4B's complaint about audibility, at the end of Turn 3 and the beginning of Turn 4, D4A reacted by looking up at D4B's video image (Screenshot 15.1). This reaction suggested that both students were aware of the multimodal resources available to them and were able to quickly look for an alternative modality to continue their communication. Both students quickly switched to video to avoid the problem in audio. Meanwhile, in Turn 4, D4A held her headset microphone close to her mouth, trying to repeat her previous question 'is there?' (Screenshot 15.2). As soon as D4A said 'is there' and D4B heard her clearly, which meant that the audio problem had been solved, D4B quickly started to look down and concentrate on her task sheet again (Screenshot 15.2). Then, D4A also looked down at her tasks sheet while finishing her previous sentence 'is there a butterfly on the chair in your picture?' (Screenshot 15.3). D4A also ended her question with an explicit reference to the task sheet, which shows that they had both reverted to their original discussion right after the technical issue was resolved.

In Turn 5, D4B replied 'no' and stressed the trigger again 'my one is a salt shaker'. There was also an implicit reference to the task sheet as 'my one' in this context clearly referred to 'my

picture'. It is also worth noting that during Turn 5, both D4A and D4B were constantly looking down at their task sheets; neither looked at the other's video image at any time. This was so in all Dyad 4's meaning negotiation episodes, as was evidenced from the gaze analysis (See Table 17 in Section 6.3.1).

In Turn 6, D4A again ignored the trigger 'salt shaker' and directly moved on to the task related interaction, suggesting that this is 'the fourth difference'. During the video stimulated recall interview (Extract 1.1), D4A told the researcher that she did not know the meaning of 'salt shaker', but she was more focusing on completing the task (spotting the differences), than resolving her non-understanding. Later in the interview, when asked about what she usually did when coming across a new word during daily conversations or reading, she expressed her preference to looking up the word in the dictionary than asking for the meaning to her interlocutor.

Video MNE VSRI Extract 1.1: D4A salt shaker

Researcher: yeah, so here when she said something on the chair, did you get what she said? D4A: I didn't know what she said it, but in my picture there is no that thing she said, cause if it is, I will know what it is
--

In Turn 7, D4B agreed with D4A about the 'fourth difference' and ended the meaning negotiation episode, without D4A's indicator of non-understanding and the subsequent stages in meaning negotiation routines. In summary, this was an incomplete and unsuccessful meaning negotiation episode

7.2.1.2 Attitudes towards meaning negotiation and modality: evidence from interviews

The analysis of this meaning negotiation episode seems to suggest that both interlocutors in Dyad 4 were not particularly competent in using the video channel to facilitate their meaning

negotiation process. During the interviews participants were asked about their general attitudes towards meaning negotiation in their daily English use, and their comment on the video/audio SCMC English class experience. Their answers went some way to explaining why they performed the way they did during the meaning negotiation episodes in video SCMC tasks.

In this discussion of the term 'salt shaker', it seems that D4A was not keen on negotiating the meaning of the unknown word. When asked about what she would do if she met an unknown word at work, when talking to native speakers, she said she would ask them to write it down so that she could look it up in the dictionary; or she would ask her colleagues. What she would not do is directly ask the native speaker to explain the meaning of the word (Video MNE VSRI Extract 1.2). So it seems that D4A was (a) not used to negotiating meaning and (b) held negative attitudes towards meaning negotiation. It seems that the dictionary is a natural enemy of meaning negotiation. Those students who were used to using dictionaries whenever they came across new words did not seem to be good or competent at negotiating meaning in oral conversations. This lack of competence in negotiating meaning was partly the reason why D4A did not indicate her non-understanding, thus failing to negotiate meaning in this episode.

Video MNE VSRI Extract 1.2 VSRI: D4A's attitude towards meaning negotiation at work

Researcher: so when you talk to a native speaker for example in your job, and then sometimes they may say something you don't understand, and then in that case will you ask them?
D4A: of course can you write it down for me?
Researcher: so you would ask them to write it down for you?
D4A: yeah so I can search the dictionary or ask my other colleagues
Researcher: would you ask them to explain it in the English?
D4A: no
Researcher: okay so you prefer to write it down and look it up in the dictionary
D4A: yes

When asked about her preferred mode of communication, D4A expressed her strong preference for audio SCMC, as compared to video SCMC, because video SCMC makes her feel nervous (Video MNE VSRI Extract 1.3). She also explained that she intentionally did not look at her peer's video image to tell whether her peer understood her or not because she was not aware that this would provide her with more information. Especially when doing tasks, D4A paid her full attention to the task sheet, and not once did she attempt to look at her peer's video image. D4A's answers during the interview showed a negative attitude towards video SCMC and revealed her low level of competence in using multimodal video SCMC.

Video MNE VSRI Extract 1.3 VSRI: D4A's attitude towards audio/video SCMC

Researcher: the first question I may ask, so we all these tasks, always one in audio and one in video, so compared video with audio, which one do you prefer?

D4A: I prefer audio,

Researcher: audio 是只有录音的, 没有视频的 [only audio, no video]

D4A: 对 [yes]

Researcher: OK, OK, why is that?

D4A: I think if we don't look at each other, so maybe we don't know this

Researcher: ehm ... 就是你喜欢的是视频还是音频? [I mean do you prefer video or audio]

D4A: 音频, 就是如果我们不是那种互相看着的话, 可能就不会那么紧张 [audio, the one without looking at each other, maybe I feel less nervous]

Researcher: ah, OK, so you feel nervous when you have video

D4A: yeah

Researcher: OK, when you were having the video communication, did you look at D4B's video image?

D4A: yes

Researcher: yes?

D4A: yes

Researcher: and did you look at her to tell whether she understands you or not?

D4A: I didn't

Researcher: you didn't?

D4A: yeah

Researcher: OK, OK, 就是你不太会去看她的视频来判断她是不是明白你的意思? [so you mean you didn't look at her video image to see if she understood you or not]

D4A: yeah, just listen to her, haha

Researcher: 为什么呢? [why?]

D4A: 额, 没有意识到 [I wasn't aware of that]

Researcher: OK, right, so was it hard for you to take care of both the camera and the task sheet in the video? is it hard for you?

D4A: not so hard

Researcher: not so hard?

D4A: 嗯嗯 [yes, right]

Researcher: right, OK, but you still prefer audio?

D4A: yeah

Researcher: 我的意思是说, 在视频的时候, 你一边要看 task, 一边还要看她的视频, 这两个事情同时做会不会有一点难? 还是你觉得很简单, 没有问题? [I mean, during video tasks, On the one hand, you need to look at the task sheet, on the other hand, you need to look at her video image. Is it hard for you to take care of both sides? or do you think it is easy, no problem?]

D4A: 没有问题, 就是我的心思就是没有有意识地去看她的视频, 我的心思都放在 sheet 上 [no problem, I mean I intentionally didn't look at her video image, I am completely focused on the task sheet]

Researcher: OK, 就是这两个 Task 都是这样 [is that so for both tasks?]

D4A: 对 [yeah]

Researcher: 就是你的心思全都放在 Task sheet 上 [completely focus on the task sheet]

D4A: 对 [yeah]

Researcher: OK, OK, 明白, alright, 所以那视频和音频对你来说区别不太大? [OK, so as I understand, audio and video SCMC were not too different for you?]

D4A: 对 [yes]

While D4A had a strong preference for audio SCMC, D4B, firmly believed that video SCMC is '*more efficient*' (Video MNE VSRI Extract 1.4) and '*clearer*'. This is because in video SCMC, D4B could see her peer's gestures, smiles, facial expressions, from which she could tell whether D4A understood her or not, as well as whether D4A wanted to continue talking or not. This extract shows that D4B was fully aware of the visual mode and the affordances of video channel. However, during their task interaction, D4B did not seem to make good use of the visual channel. As a viewer in front of the screen, she seldom looked at her peer's video image focusing instead on the task sheet or her notes. As a presenter in front of the webcam, she did not set her webcam up properly; as a result, most of the space captured by the webcam contained little useful information for her peer in terms of meaning negotiation. In fact, the top third of the frame was only the background and her face occupied the lower two thirds of the frame. The angle of the camera was also not very good because it mainly captured the top of D4B's head, meaning that her facial expressions were not clear and her gestures and posture below her shoulders remained unseen. She also made no intentional use of gestures to show the meaning of the word she wanted to explain. Therefore, it seems that although D4B was highly aware of the affordances of the video SCMC, in practice she failed to make good use of multimodal resources to facilitate the meaning negotiation process. These factors also contributed to the unsuccessful result by Dyad 4 in the video SCMC negotiated interactions.

Video MNE VSRI Extract 1.4 VSRI: D4B's attitude towards audio/video SCMC

Researcher: so we did two tasks, with one task in audio one in video, so compared video to audio, which one do you prefer?

D4B: I prefer video.

Researcher: why?

D4B: because I can, we can have eye contacts for better communication, gestures, smile face, and 表情, [facial expressions]

Researcher: okay, facial expressions.

D4B: 嗯嗯 [yes, right]

Researcher: when you were having video-conferencing, did you look at her video image?

D4B: sometimes.

Researcher: sometimes.

D4B: when we talk about the picture, I just concentrate on seeing picture.

Researcher: 那就是相比于 video 和 audio, 有哪些具体的你觉得 video 你能得到哪些更具体的信息。[so compare video and audio, what information in specific can you get through video conferencing?]

D4B: video, 嗯.....我就是可以知道, 根据对方的表情可以判断出来, 她是明白了还是不明白, 她有问问题, 还是没有问问题, 然后她是愿意继续谈还是不愿意继续谈, 然后, 对, 她对我的态度是什么样子的, 说话的时候, 然后, 嗯, 对对对, 这样的话, 交流起来是更好的, 更清楚更明白。[video, hmm, I am able to find out, judging by her facial expression, if she understand [sic] or not, if she has a problem or not, and if she wants to go on talking or not, and her attitude towards me, when she talks, yes, in this way, communication is more smooth, clear, free of misunderstanding.]

Researcher: 所以你觉得视频交流会更清楚。[so you think video conferencing makes communication clearer?]

D4B: 对, 对, 而且更省一些时间, 更有效率。[yes, yes, and also time-saving, more efficient.]

Researcher: so was it hard for you to, especially when you had the Spot-the-difference task, you need to both look at the task sheet, and also sometimes you may want to look at the camera, was it hard for you to take care of both?

D4B: I think I only look at the pictures, seldom look at D4A.

Researcher: okay, okay.

D4B: because I think the task is for us to just find out differences.

Researcher: how about the other two tasks? Spot-the-difference task, so you didn't have to look at the task, always.

D4B: but I need to look up, look at the note I have prepared, haha.

7.2.1.3 Comprehensive analysis of Dyad 4

Despite the different attitudes towards audio and video SCMC displayed by D4A and D4B, they performed very similarly in terms of their gaze focus during the task interactions. They both mostly looked down at their task sheet or notes and very seldom looked at their peer's video image, and it was their intentional choice, according to the interview. Their gaze choices could be caused by many factors, including nervousness, the incompetence to use the video SCMC tool, not wanting to interrupt their peers, or too much information on the task sheet.

In this study, the task sheet was another important semiotic resource in the SCMC environment. The sheet contained important information relating to their task completion and was frequently referred to by the participants. In Dyad 4's meaning negotiation episode, in almost every single turn, either D4A or D4B would refer to the task sheet explicitly such as 'in my picture', 'in your picture', or implicitly, such as 'my one', which actually meant 'my picture'. In this way, they both constantly communicated to each other that they were focusing on their task sheets, rather than on the video image. According to the interview, both students intentionally chose to look at their task sheets instead of their peer's video image. During the VSRI for D4A, she even agreed that there was not much difference between audio and video SCMC for her, because she did not use the visual mode anyway.

When students looked at their peer's video image, they could gain useful visual information. Gaze directions were an important measure of the extent to which a dyad made use of the visual mode. Dyad 4 spent on average 92.78% of time looking at the task sheet in spot-the-difference tasks, and 62.78% in problem-solving tasks. It could be argued that the spot-the-difference task sheet was information intensive, so students frequently had to look at their task sheets. But during the problem-solving tasks, there were only four items in each task sheet, and when students were negotiating meaning for a specific item, it was not necessary to constantly look at their task sheet; a conclusion supported by the following focus data: Dyad 3

spent only 20.13% of their time and Dyad 2 only 21.53% of their time looking at the task sheet, in problem-solving tasks.

The above analysis shows that the two members of Dyad 4 were not competent in using multiple modes and semiotic resources during meaning negotiation. They tended to intentionally ignore the visual mode and fully focused on task completion by communicating through the audio mode. Therefore, it seems that although video was available during the video SCMC, whether to use it or not was each student's individual choice and decision. Teachers or researchers should not assume that participants could competently or even willingly make use of the multiple modes and semiotic resources in video SCMC. While audio communication is necessary, visual communication is a participant's own choice, because listening is compulsory (as long as the audio mode works properly) but watching is optional.

7.2.2 Dyad 1 Task 5: 'razor' episode

The second meaning negotiation episode to be analysed involves Dyad 1 completing Task 5, a problem-solving task where each student had pictures of four different gifts, and they were asked to agree on a choice of four gifts out of eight to give to the members of their homestay family in the UK (see Video MNE Extract 2). This episode shows how D1B pretended to understand the word 'razor' using multiple modes.

Video MNE Extract 2: Dyad 1 Task 5, the 'razor' episode

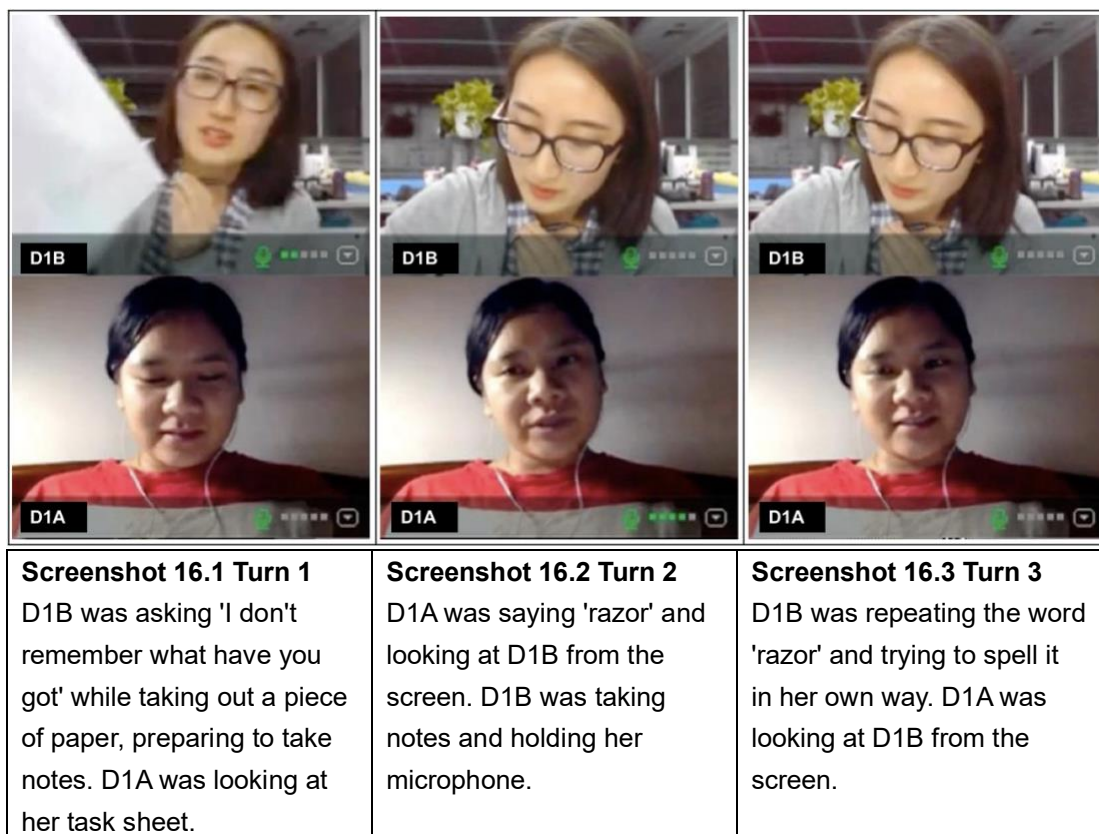
Turn	Conversation	Meaning negotiation stages	Gaze directions	Multimodal elements
1	D1B: I don't remember what have you [sic] got? a perfume?	Task related interaction	D1A: task sheet D1B: note	Screenshot 16.1
2	D1A: a razor, razor,	Trigger	D1A: peer's video image D1B: note	Screenshot 16.2

3	D1B: razor,	Repeating trigger (no indicator)	D1A: task sheet D1B: note	Screenshot 16.3
4	D1A: razor ... and then perfume	Repeating trigger	D1A: peer's video image D1B: note	
5	(task interaction unrelated to razor) D1A: so so it seems, so (..) so it seems, ehm (...) razor (...) is the best choice for the father	Trigger	both: peer's video image	Screenshot 17.1-17.3
6	D1B: the father?	No indicator	both: peer's video image	Screenshot 17.3
7	D1A: father, for the father, yeah, razor	Task related interaction	both: task sheet	Screenshot 18.1
8	D1B: I think it is, maybe	Task related interaction	both: task sheet	Screenshot 18.2-18.3
9	D1A: a razor	Task related interaction	both: task sheet	Screenshot 19.1
10	D1B: ehm, yeah, let's decide the razor for father	Task related interaction	both: moving between the task sheet and peer's video image	Screenshot 19.2-19.3

7.2.2.1 Turn-by-turn multimodal analysis of the episode

Before this interaction, D1A had listed all her 4 items but D1B did not completely follow her. This explained why, in Turn 1, D1B asked what D1A had got, and queried 'perfume?', the only one she could remember. While asking the question, D1B took out a piece of paper, preparing to take notes (see Screenshot 16.1). In Turn 2, D1A replied 'a razor', and repeated 'razor', which was the first item on her task sheet. Meanwhile, D1B had already started taking notes (Screenshot 16.2). This was a trigger (of non-understanding) for D1B, as she revealed in her stimulated recall interview. In Turn 3, D1B repeated the word 'razor' and was trying to spell it according to its pronunciation and write it down but she did not ask the meaning of the word to D1A (Screenshot 16.3). According to the stimulated recall interview (Video MNE VSRI Extract 2.1), D1B confirmed that she did not understand the meaning of razor. She was trying to spell

the word in her own way and hoped that D1A could further describe the word without her explicitly asking for an explanation. Unfortunately, D1A did not offer any description, and D1B eventually looked up the word razor in an online dictionary during their subsequent task interaction.



Screenshot 16: D1_T5_ razor_Turn 1-3

Video MNE VSRI Extract 2.1: D1B's thoughts during the 'razor' episode

D1B: razor?
 Researcher: yeah
 D1B: ehm
 Researcher: did you know this word before or?
 D1B: ehm razor, when D1A say razor, I didn't know what she means actually, in that time, I was confused, but I try to guess, I try to guess after she ...
 Researcher: after she describe it?
 D1B: yeah, after she describe it
 Researcher: here when she said razor, did you get it?

D1B: no

Researcher: no?

D1B: no, in that time I didn't get it

Researcher: but were you taking notes?

D1B: no, I just try to spell it in my own way, haha (both laughing)

Researcher: so you have your own spelling? how did you spell it, if you remember?

D1B: razor, I ... just r, a, s, or z, e, r

Researcher: yeah, OK

D1B: OK

Researcher: but then how did you guessed [sic] in the end, the word razor?

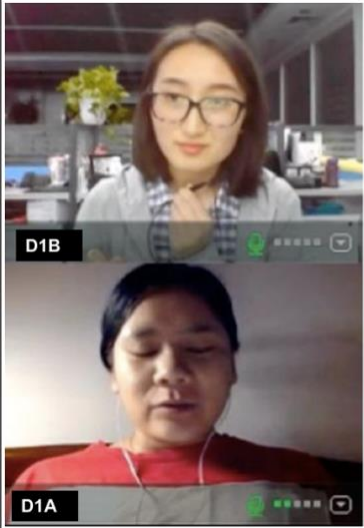


D1B: razor? I checked it, in in online, yeah

Researcher: during the class?

D1B: yeah

Researcher: OK, OK

After Turn 4, they engaged in task related interaction unrelated to the word 'razor'. Then D1A mentioned 'razor' again in Turn 5 (Screenshot 17.1), explaining that she thought it was '*the best choice for the father*'. D1B was looking at D1A's video image throughout the whole of Turn 5, and suddenly smiled as soon as D1A said 'razor' (Screenshot 17.2). And D1A saw D1B's smiling reaction during the short pause right after she said 'razor' (Screenshot 17.3). At this point, D1B still did not know the meaning of 'razor' but she smiled in reaction to D1A and asked 'for the father?', a task related question which was vaguely related to its meaning. But again, D1B did not explicitly indicate her non-understanding and D1A did not offer any further description or explanation.

		
<p>Screenshot 17.1 Turn 5 Before D1A said 'razor'. D1B was listening to D1A and looking at her video image.</p>	<p>Screenshot 17.2 Turn 5 D1A just pronounced the word 'razor' and D1B already started smiling.</p>	<p>Screenshot 17.3 Turn 5 & 6 D1A was smiling and looking at D1B's video image. D1B was also looking at D1A and possibly saw her smile.</p>

Screenshot 17: D1_T5_ razor_Turn 5

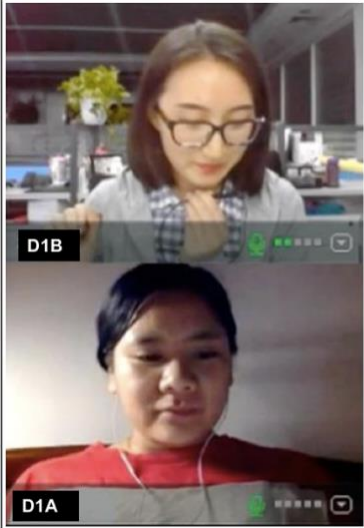
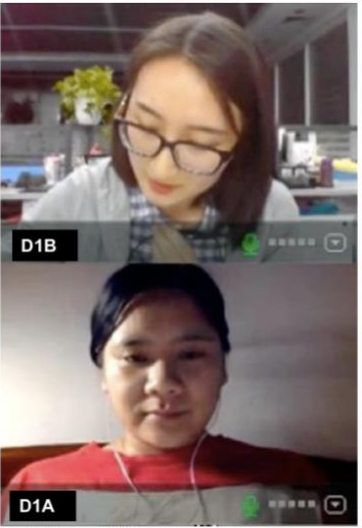
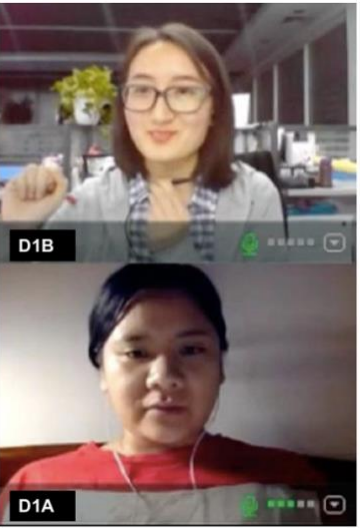
In Turn 7, when D1A confirmed '*it's for the father*', D1B smiled again (Screenshot 18.1) while D1A was looking at D1B's video image. At the beginning of Turn 8, D1B nodded twice (Screenshot 18.2 and 18.3) while offering her agreement '*I think it is*', but then she added '*maybe*', suggesting that she was not sure what she agreed on. Such a response indicated that she still did not understand the word 'razor'.



<p>Screenshot 18.1 Turn 7 D1A confirmed 'for the father, yeah' and was looking at D1B's video image. D1B was smiling and looking at D1A.</p>	<p>Screenshot 18.2 Turn 8 D1B nodded while saying 'I think it is' during Turn 8. D1A was looking at her task sheet.</p>	<p>Screenshot 18.3 Turn 8 D1B nodded while saying 'I think it is' during Turn 8. D1A was looking at her task sheet.</p>
---	--	--

Screenshot 18: D1_T5_razor_Turn 7 and 8

In Turn 9, D1A repeated 'a razor' and D1B was taking out her pen, preparing to take notes (Screenshot 19.1). In Turn 10, she offered a stronger agreement saying '*yeah, let's decide the razor for the father*', while leaning forward and taking notes (Screenshot 19.2). At the end of this turn, D1B finished writing, looked up to her peer's video image and smiled into the camera, when D1A was looking at D1B's video image (Screenshot 19.3).

		
<p>Screenshot 19.1 Turn 9 D1A repeated 'a razor', while looking at her task sheet. D1B was taking out her pen, preparing to take notes.</p>	<p>Screenshot 19.2 Turn 10 D1B was saying 'Yeah, let's decide the razor for the father', while leaning forward and taking notes.</p>	<p>Screenshot 19.3 Turn 10 At the end of the turn, D1A looked up at D1A's video and smiled after taking notes. D1A was looking at D1B and possibly saw her smile.</p>

Screenshot 19: D1_T5_razor_Turn 9 and 10

This moment was the end of their discussion about the word 'razor'. According to D1B's hesitating words 'maybe' and the video stimulated recall interview, D1B did not guess the meaning of the word 'razor' throughout the whole episode. She confirmed that she learned the meaning of the word after the meaning negotiation episode when she looked the word up from the online dictionary.

7.2.2.2 Attitudes towards meaning negotiation and modality: evidence from interviews

D1B's attitude towards meaning negotiation and audio/video SCMC

The stimulated recall interview confirmed D1A's reluctance to negotiation for meaning. D1B explained that this reluctance was because she '*did not want to interrupt her (peer)*' (Video MNE VSRI Extract 2.2). When asked about her attitudes towards interrupting people and asking

for the meaning of an unknown word, D1B reported that she was '*afraid to do that*' and chose to '*give it a chance to listen*' and hope that she '*can figure out what it is*'. When she could not guess the meaning, as in this 'razor' episode, she tried to spell it in her own way, looked it up in the online dictionary and eventually understood the meaning from the Chinese translation on the online dictionary.

Video MNE VSRI Extract 2.2: D1B's attitude towards meaning negotiation

D1B: ehm ... I just ... I think it doesn't matter on this all the picture, and I don't want to interrupt her to continue to describe

Researcher: OK, so do you think interrupting people in the middle and asking what do you mean by this mean is not [polite]?

D1B: ehm, actually, I am afraid to do that, yeah, I just, ehm in my own, I said it's a shorter ability to know this part, I just give it a chance to ... to listen, and after that maybe I can figure out what it is

Researcher: OK, so you think if you don't interrupt her here, you may try to guess in the later conversation what is it [sic]

D1B: yeah, yeah

Researcher: good, OK, let's go on

Despite D1A's reluctance to meaning negotiation, she seemed to be competent in using multimodal resources to achieve what she wanted in the SCMC. She tried her best to avoid explicitly indicating her non-understanding. For example, she often looked at her peer's video image, and used nods, smiles to show her 'understanding' and 'agreement'. Her camera was set at a very good angle so that her face could be seen clearly. D1B's competence in making use of multiple modes to express herself was also paralleled by her attitude towards video SCMC during the stimulated recall interview (Extract 2.3). When asked about whether she preferred audio or video SCMC, D1B chose video because they '*can see each other's face*' and '*it's more like we're communicating*'. She demonstrated a good understanding of the benefits of video SCMC over audio SCMC when she said she could tell her peer's feelings or

attitudes according to her facial expression and hand gestures and body movements. She felt more distant from her interlocutor during audio SCMC.

Video MNE VSRI Extract 2.3: D1B Attitudes towards audio and video SCMC

Researcher: so first of all we did each task, Spot-the-difference and problem-solving task in both audio and in video, and compare audio with video, which one do you prefer?

D1B: ehm, ehm, I prefer, the video

Researcher: why?

D1B: I guess we can see each other, and we can see each other's face and make the conversation as well

Researcher: ehm, is there any additional information from her, for example, like what?

D1B: ehm, I can see the ehm

Researcher: you can use Chinese if you like

D1B: ehm, 怎么说呢,就是,更像是在交流 [how to say it, it's, it's more like we are communicating]

Researcher: OK, 那有没有什么你觉得是视频的优点和缺点、问题? [so what do you think are the advantages, disadvantages or problems of video conferencing?]

D1B: 视频的优点可能就是在交流方面你可以通过对方的表情或者手势去猜测他当时是一个什么样的心情和态度 [the benefit of video is, when you are communication, you can guess what is your partner's feeling or attitude according to her facial expression or her hand gestures]

Researcher: 嗯 [OK]

D1B: 音频的话可能这方面的信息会少一点, 交流起来可能对方的距离感也会远一点 [in audio, there's fewer information like this, and the partner may feel more distant during the communication]

In conclusion, D1B's performance in the 'razor' episode illustrated her competence in making use of multimodal resources. The main reason why she did not indicate her non-understanding was that she did not want to interrupt her peer and that she was 'afraid of' asking questions to her peer. Therefore, it could be concluded that competence in making use of multiple modes

and semiotic resources was not the only factor contributing to the success of meaning negotiation in video SCMC. The willingness to negotiate meaning also played an important role.

D1A's attitude towards meaning negotiation and audio/video SCMC

In their task interactions, D1A also did not seem to be interested in negotiation of meaning. In the 'razor' example, as long as they had agreed to choose the razor for the father as a gift, the task was done. Without D1B's explicit indication of non-understanding, D1A did not offer further descriptions or explanations of the word 'razor' to her peer. Similarly, in Task 3, the spot-the-difference task, when D1B said she had a 'model ship' instead of a 'model rocket' in her picture, D1A did not understand what D1B meant and did not ask for any further explanation. During the stimulated recall interview (see Video MNE VSRI Extract 2.4), D1A confirmed her non-understanding and explained that as long as they spotted the difference, '*the task is done here*'.

However, when asked about what she would do when coming across a non-understanding during a daily conversation with a native speaker, D1A very firmly said that she would ask for the explanation directly. She did not feel embarrassed to ask when talking to native speakers. Therefore, it seems that D1A still holds a positive attitude towards negotiation for meaning, although in this research project she was mainly focusing on task completion. D1A's lack of interest in meaning negotiation could be caused by the task design, as negotiation for meaning was not the requirement for the task completion.

Video MNE VSRI Extract 2.4: D1A's attitude towards meaning negotiation

Researcher: yes, ehm here you were describing about the boy, and then D1B said instead of the model rocket she has a model ship, did you get it?

D1A: not really,

Researcher: not really, OK, so why did you not continue asking her

D1A: because the task is done here, Spot-the-difference

Researcher: OK

D1A: no matter what specific toy she had in her picture, it doesn't matter, all that matters is that it's different,

Researcher: OK, so it's very important for you to complete the task?

D1A: yeah,

Researcher: and you have talked to some native speakers?

D1A: yes,

Researcher: and did you come across some new words what they say and you don't understand?

D1A: not very often,

Researcher: not very often?

D1A: yes

Researcher: OK, if there are something, would you prefer to ask him or her directly or would you prefer to note it down and look it up afterwards?

D1A: I'd ask them

Researcher: you would?

D1A: ehm

Researcher: OK, you would, do you think it would be kind of embarrassing or face-threatening

D1A: no

Researcher: not at all?

D1A: no, they are native speakers (laughing),

Researcher: yeah

When it comes to the preference for video or audio communication, D1A expressed her strong preference for video SCMC during the very first induction session. According to the interview (Video MEN VSRI Extract 2.5) it was important for her to be able to see her peer's video image. D1A felt nervous in audio communication because she was not sure if the message could be delivered as well as it would be in video communication when they could see each other's faces and could use gestures. D1A was also aware of and competent in presenting herself in front of the webcam for her peer to see her. For example, she used her facial expressions to show she

felt sorry when she had to ask D1B to repeat all her items in the task sheet as D1A had not remembered D1B's items.

Video MNE VSRI Extract 2.5: D1A's attitude towards audio and video SCMC

Researcher: OK, right, good, so that's all the audio and video parts we need you to review and I have some other more general questions about the whole online class, so first of all, I want to compare the audio and video channel, I remember in the very first class, you were saying you really don't like audio (both laughing),

D1A: I don't like audio, yes

Researcher: why is that?

D1A: because ehm I don't know, it just always makes me nervous when I can't see the other person's face, it makes me nervous

Researcher: OK

D1A: and I think it makes me nervous because I wasn't sure if my message would be delivered as well as it did when we were video conferencing, when I can see other people's face, because we have body language to substitute ...

Researcher: OK, so if you analyse or if you really think about your performance or your talk through audio and video, what kind of additional information you can get from video that you can't get from audio?

D1A: ehm, I think it helps to understand the whole situation, like when I feel sorry when I asked her to repeat and repeat and repeat, by looking at me, she will be able to tell that I'm sorry, if it is audio conferencing, no, she wouldn't be able to tell

7.2.2.3 Summary

So far, it can be seen that neither Dyad 1 nor Dyad 4 (the 'salt shaker' episode) were well attuned to negotiating meaning. The difference is, while Dyad 4 did not demonstrate a clear awareness of (D4A) and ability (D4A and D4B) to use multimodality in meaning negotiation, Dyad 1 clearly possessed more competence in this negotiation task. However, D1B

220

intentionally used her multimodal communicative competence to avoid meaning negotiation, which is why, in this case, they did not manage to arrive at a correct understanding of the lexical items they were intended to discuss.

The analysis of these two dyads' meaning negotiation episodes illustrates that both students' attitudes towards meaning negotiation, and their multimodal communicative competence levels, could largely influence their success in negotiating meaning in a video SCMC. The next two episodes (Section 3 and 4) will demonstrate what happens when students had positive attitudes towards meaning negotiation in multimodal video SCMC.

7.2.3 Dyad 3 Task 5: 'magnifying glass' episode

Video MNE Extract 3 shows Dyad 3 undertaking Task 5, in which the two participants were required to choose 4 items for their homestay family members. In this meaning negotiation episode, Student D3B was trying to explain the meaning of the expression 'magnifying glass' to D3A.

Video MNE Extract 3: D3 T5, the 'magnifying glass' episode

Turn	Conversation	Meaning negotiation stage	Gaze directions	Multimodal components
1	D3B: alright, just now I've told you that there's a picture of a bouquet, and other things are Rubik's cube, and cos, ehm, cosmetics, and a magnifying glass	Trigger	D3A: mostly peer's video image D3B: mostly task sheet	/
2	D3A: a pair of glasses?	Indicator	both peer's video image	Screenshot 20.1-20.3 D3A used hand gesture to show glasses
3	D3B: no no magnifying, magnifying glass ... it is a kind of a glass, but if you use that, it can enlarge something, and you can see the details about that	Reply	D3A: mostly peer's video image D3B: mostly task sheet	Screenshot 21 D3B made a hand gesture of 'enlarging' at the lower left corner of her screen.
4	D3A: OK, I got it	Response to reply	both: peer's video image	D3A slight nod, thumb pointing at herself, at the lower-left corner of the screen
5 D3A: I heard that, I heard you said that ehm (.) your paper is a (.) gla? a pair of glasses?	Confirmation of trigger	D3A: peer's video image D3B: mostly task sheet	/
6	D3B: no, no, magnifying <u>glass</u>	Confirmation of indicator	both: peer's video image	/
7	D3A: magnifying glasses?	Indicator	D3A: peer's video image D3B: mostly task sheet	/
8	D3B: glass (.) yeah (..) it's glass, not glasses	Reply	both: peer's video image	/
9	D3A: glass?	Response to reply	both: peer's video image	/

10	D3B: just glass (.) not glasses	Reply	both: peer's video image	D3A confused facial expression
11	D3A: not a pair of glasses?	Response to reply	both: peer's video image	Screenshot 22. D3A hand gesture showing 'glasses'
12	D3B: no, no, no, no no, not glasses (.) yeah, no, yo, you could not wear (.) with that, yeah (.) it's (.) you see, ehm, it's kind of, how to say that, it's a kind of glass can make the things larger, yeah? when you use that, you can see (.) it (.) larger, bigger	Reply, explanation	D3A: peer's video image D3B: moving between peer's video image and the task sheet	Screenshot 23. D3B slightly shook her head, D3B hand gesture to show 'enlarge' again at the lower-left corner of the screen
13	D3A: I see, I see, but I think thi, this ehm present is suitable for the (.) son, do you agree with me?	Task interaction	both: peer's video image	D3A slightly nodded
14	D3B: Billy? ehm (.) what is that for?	Task interaction	both moving between peer's video image and the task sheet	/
15	D3A: I think (.) for the first time, ehm (.) it's very young, and he ehm (.) has some interests, ehm (.) in ehm, different kinds of things	Task interaction	both: mostly peer's video image, glanced at the task sheet occasionally	/
16	D3B: ehm (.) comparing, comparing ehm, the Rubik's cube, and magnifying glass, I prefer the Rubik's cube, Rubi (...) k's cube, I think it's suitable (.) more suitable than (.) magnifying glass	Task interaction	D3A: peer's video image D3B: mostly task sheet	/

3.1 Turn-by-turn multimodal analysis of the episode

In Turn 1, D3B mentioned three items in her task sheet, including '*Rubik's cube, cosmetics and magnifying glass*', without any explanation of these items, which triggered D3A's non-understanding. So in Turn 2, D3A uttered '*a pair of glasses*' as an indicator of non-understanding. This was a local indicator as she was repeating part of the trigger with rising intonation. Meanwhile, she was staring at her peer's video image and imitating the shape of glasses with her hands around her eyes (Screenshot 20.1-20.3). D3B was looking at D3A's video images while D3A was making gestures.

		
<p>Screenshot 20.1 Turn 2 D3A was trying to ask if the item was 'a pair of glasses?' She was lifting up her hand towards her eyes. D3B was looking at D3A's video image and listening to her.</p>	<p>Screenshot 20.2 Turn 2 D3A was trying to ask if the item was 'a pair of glasses?' She put her hand by her eyes and made a circle shape with her fingers. D3B was looking at D3A's video image and listening to her.</p>	<p>Screenshot 20.3 Turn 2 D3A was trying to ask if the item was 'a pair of glasses?' She put her hand by her eyes and made a circle shape with her fingers. D3B was looking at D3A's video image and listening to her.</p>

Screenshot 20: D3_T5_magnifying glass_Turn 2

Knowing that D3B had not understood, D3A, in Turn 3, first rejected D3A's guess of 'a pair of glasses', repeated the trigger '*magnifying glass*' and started explaining '*it can enlarge*

something, you can see the details about that'. While saying the word 'enlarge' D3B used a hand gesture, but it was in the lower left-hand corner of the screen (Screenshot 21), and it was a very rapid gesture (less than a second), so it was possibly hard to notice.



D3B was saying 'it can enlarge something', while making a hand gesture of 'enlarging' at the lower-left corner of her screen. D3A was looking at and listening to D3B.

Screenshot 21: D3_T5_magnifying glass_Turn 3

After D3B's explanation, D3A replied positively in Turn 4 'OK, I got it', with a slight nod. She also used a rapid hand gesture, with her thumb pointing at herself, but her hand was at the lower left-hand corner of the screen, which was partly blocked by the name tag area of the video conferencing interface. D3B showed no awareness of the need to present gestures clearly in front of the webcam. According to the stimulated recall interview, D3A thought D3B was talking about high-tech virtual reality glasses. So she confirmed her understanding in Turn 4. When D3A was asked why she did not offer any further explanation or issue a clarification request, D3A answered that she had not thought about it at that moment. It was also worth noting that D3A was not reluctant to indicate non-understanding. As she stressed during the interview (Video MNE VSRI Extract 3.1), she was not afraid of asking questions to her peer if she lacked understanding. She did not feel that it was either impolite or embarrassing to ask

questions. But in this case, they moved on because D3A thought she had guessed the correct meaning, although actually it was a misunderstanding.

Video MNE VSRI Extract 3.1: D3A's misunderstanding of 'magnifying glass'

Researcher: OK, so first of all when she said magnifying glass, you asked a pair of glasses?

D3A: yes

Researcher: what were you thinking?

D3A: 可能是现在比较流行的那个有 VR 技术的眼镜 [maybe the popular glasses with Virtual Reality technology]

Researcher: 哦, 你想到了那个 [oh, so you guessed that]

D3A: 对所以我当时说那个礼物的时候, 我说这个礼物比较合适 [that's why when I chose the gift, I said this is a suitable gift]

Researcher: oh, OK

D3A: 就是对一个男孩子来说, 他的求知欲比较强 [I mean for a teenage boy, he should be very interested in this (VR glasses)]

Researcher: oh, 所以你那时候理解的是 VR 眼镜 [so you understood that it was VR glasses (she was referring to)]

D3A: 嗯 [OK]

Researcher: OK

.....

Researcher: 你猜的是 VR 眼镜, 那你为什么不直接跟她问一下确认一下这个 VR 眼镜呢?

[so why didn't you try to confirm with her that it was VR glasses you were thinking about?]

D3A: 没有, 没有想到问 [no, I didn't think about asking her]

Researcher: 那你觉得就是向她提问确认这个事情是 ok 的还是不礼貌或者有点不太好意思的感觉? [so do you think asking her questions to confirm the meaning of the word is fine or do you consider it to be impolite or embarrassing]

D3A: 没有 [no]

Researcher: OK

D3A: 就是没有不好意思的感觉, 有不懂的我会去问, 但是在那个场合下就感觉顺理成章地直接走了, 没有弄明白 [I didn't feel embarrassed, if there's something I don't know, I would ask, but in that situation, it just went on smoothly, although I didn't really get the right meaning.]

According to the stimulated recall interview with D3B, when D3A said she got it, D3B just assumed she really had '*got it*' and therefore did not engage in any further explanations or comprehension checks (Video MNE VSRI Extract 3.2).

Video MNE VSRI Extract 3.2: D3B's failure to check comprehension

Researcher: Yeah here when she said she got it, you just thought that she got it?

D3B: hahahaha, yeah, I think

Because of this misunderstanding, D3A used this phrase again (Turn 5) after further task-related discussion since she thought the virtual reality glasses were the most suitable gift for a teenage boy. In Turn 6, D3B corrected D3A by stressing the singular 'glass' instead of 'glasses'. Although this looked like negotiation of form, this perception was because the two different forms had different meanings. But D3A did not realise this and repeated the wrong plural form again with a rising intonation pattern 'magnifying glasses?' in Turn 7. In Turn 8, D3B stressed the singular form again, and explicitly said it was '*just glass, not glasses*'. During the stimulated recall interview (Video MNE VSRI Extract 2.3.3) D3B explained the reason why she stressed the singular form, but she also said she should have explained the differences more clearly to D3A. In Turn 9, D3A finally realised that D3A emphasised '*glass*', so at this point she repeated it with a raising tone, indicating her non-understanding or confusion. In Turn 10, D3B repeated again '*just glass, not glasses*', which made D3A more confused since she was thinking of the VR glasses. Therefore, in Turn 11, D3A explicitly asked '*not a pair of glasses*', which indicated her previous understanding might be wrong. D3A even used gestures again (Screenshot 22), to mimic a circle near her eyes, just as she did in Turn 2 (Screenshot 20).



<p>Screenshot 22.1 Turn 11 D3A was asking 'not a pair of glasses?' while trying to move her hand towards her eyes. D3B was not looking at D3A's video image.</p>	<p>Screenshot 22.2 Turn 11 D3A was asking 'not a pair of glasses?' while showing a circle with her fingers around her eyes. D3B was looking at D3A's video image.</p>	<p>Screenshot 22.3 Turn 11 D3A was asking 'not a pair of glasses?' while showing a circle with her fingers around her eyes. D3B was looking at D3A's video image.</p>
---	--	--

Screenshot 22: D3_T5_magnifying glass_Turn 11

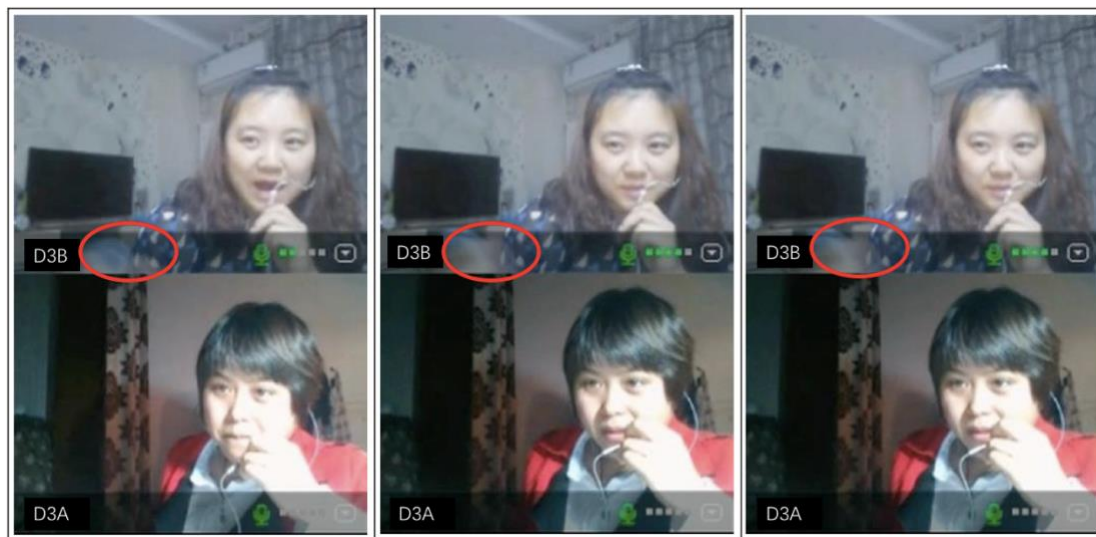
Video MNE VSRI Extract 3.3: D3B's further attempts at clarification

Researcher: so you are emphasising it's glass, not glasses, why?

D3B: I think the glasses means 眼镜 [glasses], and glass just a kind of material, so I think she ehm just now I thought maybe she couldn't understand the word, glasses or glass, I want to focused let her focus on glass, yeah but I should explain, explain that it's a kind of material

Having known that D3A still had not fully understood, D3B in Turn 12 first clarified that it was not a pair of glasses, and further explained that '*you can't wear it*'. She then tried to explain the meaning of 'magnifying glass' again by saying '*you can see it larger, bigger*'. While saying the words 'larger' and 'bigger', D3B used some very quick hand gestures (Screenshot 23), possibly showing an 'enlarging' gesture, similar to the one she did in Turn 3 (Screenshot 21). The hand gestures were so quick and mostly out of the frame that the screenshot could only show a flash in Screenshot 23. In Turn 12, D3B did not show any awareness of the webcam, such as

reminding her peer to look at the screen, showing the gesture in the middle of the webcam, or repeating the gestures for her peer to see.



D3B was saying 'you can see it larger, bigger', while using two very quick hand gestures during 'larger' and 'bigger'. Her hand gestures (in red circles) were too brief and at the bottom of the frame. D3A was staring at D3B's video image and listening to D3B.

Screenshot 23: D3_T5_magnifying glass_Turn 12

In Turn 13, D3A replied '*I see, I see*' to D3B's explanation, and then moved on to the task-related interaction. But from Turn 13 to Turn 16, it is clear that D3A was still thinking about the virtual reality glasses, which indicated that she did not really understand D3B's explanation '*just glass, not glasses*', and her emphasis on '*you can't wear with that*'. Despite the confusing conversation they had for the task-related interaction between Turn 13 to 16, neither of them appeared to consider that they might have a misunderstanding of the meaning of 'magnifying glass'. When asked about whether D3B could tell if D3A understood her during the interaction (Video MNE VSRI Extract 3.4), D3B confirmed that she could not judge '*at that time*', but she did not make any further effort to confirm D3A's understanding.

Video MNE VSRI Extract 3.4: D3B's failure to confirm D3A's understanding

Researcher: were you looking at her to see whether she understood or not?

D3B: just now I was looking at her, I think

Researcher: and then did you think that she understood you?

D3B: hahahaha, I couldn't judge at that time

Researcher: you couldn't judge?

D3B: yeah, yeah

3.2 Attitudes towards multimodality: evidence from interviews

It already seems clear that the participants in Dyad 3 had positive attitudes towards meaning negotiation as both interlocutors were happy with asking each other questions or giving further explanations to clarify the meaning of the word. This section focuses on reporting their attitudes towards the whole experience of doing these tasks through audio and video SCMC environments. According to Video MNE VSRI Extract 3.5 and 3.6, both D3A and D3B stated that they preferred video SCMC because it could provide more information for them to better understand each other. However, both students also confirmed that they did not always look at their peer's video image, especially when they had to focus on the task sheet. Their answers demonstrated their understanding of the additional multimodal information afforded by video SCMC, as compared to audio SCMC. However, they also showed their inability to make full use of the visual mode during their task interaction because of the task sheet.

Video MNE VSRI Extract 3.5: D3B's attitude towards audio and video SCMC

Researcher: so you did these two tasks, for each task you have one in audio and one in video, and comparing audio and video, which one do you prefer?

D3B: if there is no [sic] any problem about the Internet, I prefer audio

Researcher: audio?

D3B: oh, video

Researcher: video? The one with images?

D3B: yes, video, yeah yeah, images, video can help you understand

other people talking, and the body gesture can help you to understand each other, but if there is something about, ehm, sometimes I like video a little bit because I can focus on what she said

Researcher: you mean on audio or video?

D3B: audio, audio

Researcher: audio is only the sound

D3B: yeah yeah, audio yeah, audio

Researcher: and audio can make you concentrate

D3B: yes yes

Researcher: so do you think video can disturb you

D3B: a little bit

Researcher: distract

D3B: yeah yeah a little bit you should focus, you should look at their others' face, you need to get some information about whether she can understand you or not, although I didn't always look at her face, hahaha, because we need to look at the paper often, yeah, so sometimes I couldn't, yeah, look at her face, if without any material, just face-to-face just like us, I would just look at your face this way, hahahaha

Researcher: yeah, OK, yeah

Video MNE VSRI Extract 3.6: D3A's attitude towards audio and video SCMC

Researcher: we have the problem-solving task, and the Spot-the-difference tasks, and each one is done, one in audio and one in video, right? 一个是视频, 一个是音频, [one in audio and one in video], so compared audio to video, which one do you prefer? and why?

D3A: 都可以 [both are fine]

Researcher: 哪一个更喜欢呢? any preference?

D3A: video

Researcher: why?

D3A: because I can, I can see her expressions,

Researcher: you can see her expressions, and does that help your understanding?

D3A: yeah,

Researcher: so do you feel easy or hard to take care of both the video camera and the task sheet? 就是你上课的时候既要看视频, 又要看那个 task sheet, 这个对你来说处理上会不会觉得困难还是觉得不难?

D3A: 还是有点困难的, 有时候你专注于纸上的一些问题, 可能一些表情就忽略了 [it is a bit hard, sometimes when I concentrated on the task sheet, I might ignore her facial expressions]

7.2.4 Dyad 2 Task 6: 'nail polish' episode

Video MNE Extract 4 is from Dyad 2 Task 6, in which each student had a list of four different items. They had to choose four out of the eight different items for fund raising in a flea market in their university. Students needed to explain their own items to each other and then discuss which ones to choose. The episode below is when D2A was trying to explain the word 'nail polish' to her interlocutor D2B.

Video MNE Extract 4: Dyad 2 Task 6, the 'nail polish' episode

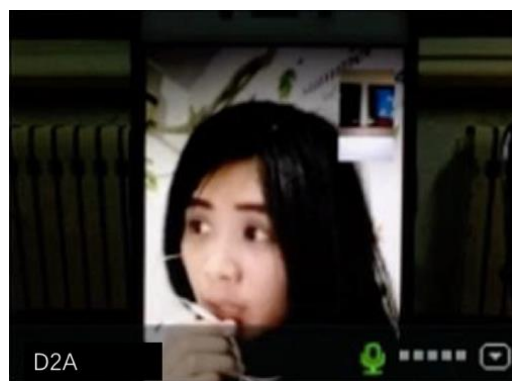
Turn	Conversation	Meaning negotiation stage	Gaze direction	Multimodal elements
1	D2A: ehm, next one is, nail polish, nail ... nail polish, do you know what it is?	Trigger/ Comprehension check	D2A: Peer's video image; D2B: task sheet and the online dictionary on the screen	D2B was looking for the previous word 'blender' in an online dictionary.
2	D2B: nail college?	Repeat, Confirmation of Trigger	D2A: peer's video image D2B: peer's video image and notes	D2B had a frown on her face while trying to take notes.
3	D2A: it's used for ... to make your, to make your nails colourful	Reply	both: mostly peer's video image	D2B touched her nose and scratched her neck
4	D2B: oh, sorry	Response to reply	both: mostly peer's video image	D2B slightly signed and scratched her hair while saying 'sorry'
5	D2A: do you know what it is? ehm, girls like that, nail polish, do you know finger nails? finger nails?	Comprehension check, Repeat Trigger	D2A: peer's video image D2B: peer's video image and notes	Screenshot 27.2 compared with screenshot 27.1 in Turn 7, the researcher slightly adjusted D2A's video image
6	D2B: nail college? ehm, actually, I can't hear you very clearly	Confirmation of trigger, reference to the aural mode	both: mostly peer's video image	Screenshot 26 D2A showed her fingers in front of the camera. D2B moved forward to be closer to the camera but showed no reaction to D2A's hand.
7	D2A: can you see me? can you see my picture?	reference to the visual mode	both: peer's video image	Screenshot 27.1 compared with screenshot 27.2 in Turn 5, the researcher slightly adjusted D2A's video image
8	D2B: yes, I can see you, but I, actually, I can't hear you very clearly, there, ehm there are some noise in, in our pictures	reference to the visual mode	both: peer's video image	/
9	D2A: but I can hear you clearly	reference to the aural mode	both: peer's video image	D2A held her speaker
10	D2B: oh, it'sOKnow	reference to the aural mode	both: peer's video image	/
11	D2A: OK, so do you know what is nail polish?	Comprehension check	D2A: peer's video image D2B: peer's video image and notes	/
12	D2B: nail ... college?	Repeat/ Confirmation of trigger	both: peer's video image	D2B scratched her hair again
13	D2A: polish, polish	Repeat trigger	D2A: her own video image to check if her hand was presented clearly D2B: peer's video image and the task sheet	Screenshot 28: D2A put down the microphone and showed her hand in front of the camera.

14	D2B: ... nail bolish?	Repeat/ Confirmation of trigger	D2A: her own video image to check if her hand was presented clearly D2B: peer's video image	Screenshot 29: D2A was showing her hand in front of the camera and looking at her own video image.
15	D2A: nail, nail, can you see me? do you know what is nails? nails?	Repeat trigger, reference to the visual mode, comprehension check	D2A: moving between her own and peer's video image D2B: peer's video image	Screenshot 29: D2A showed her finger nails right in front of the camera in the centre of her own video image, but D2B seemed to be writing something or taking notes and did not look at D2A's video image.
16	D2B: nail? oh, yes, yes, nail	Response to reply	D2A: moved between her own and peer's video image D2B: peer's video image	Screenshot 30: D2A kept showing her fingers in front of the camera. D2B slightly nodded while saying 'oh, yes'
17	D2A: and nail polish, do you know that?	Repeat trigger, comprehension check, Confirmation	D2A: moved between her own and peer's video image D2B: peer's video image	D2A put down her hand away from the camera
18	D2B: I know that	Reconfirmation	D2A: peer's video image D2B: peer's video image and her note	Screenshot 31: Both D2A and D2B smiled
19	D2A: OK, it's something (overlapping) that make your nails colourful, OK?	Reply to response (explanation)	D2A: peer's video image D2B: peer's video image and her note	D2B kept smiling
20	D2B: nail polish, OK	Reconfirmation	both: mainly peer's video image	/
21	D2A: do you, do you know it now?	Confirmation, comprehension check	both: peer's video image	/
22	D2B: OK, I know what it is, ehm, make the nail, ehm ... if your nail longer, you can use it, make it short, right?	Reconfirmation	both: peer's video image	Screenshot 32 & 33: There's clearly a gap between the video image and the sound. It seems that the video image comes a bit earlier than the sound. D2B showed her finger nails in front of her camera.
23	D2A: no no no no, it makes you, your nails colourful	Reply to reconfirmation, explanation	both: peer's video image	Screenshot 34: D2B showed the hand gesture of cutting nails
24	D2B: oh, I know that! you can, you can use to make your nails colourful, maybe you can use red nail polish, black nail polish, red nail polish, right?	Reply to response	both: peer's video image	Screenshot 35: D2B nodded clearly, touched her hair with her right hand and raised up her left hand.
25	D2A: yes, yes,	Reconfirmation	both: peer's video image	/

7.2.4.1 Multimodal analysis of meaning negotiation episode

7.2.4.1.1 Technical issues in the video channel for Dyad 2

Before any analysis takes place, it is important to understand the technical setting of the interaction, which was slightly different from others. Due to a connection failure, D2B could not see D2A's video image through the video conferencing system from her laptop. The online teacher and the dyad tried different ways to deal with this technical issue. In the end, the researcher had to use the video chat function of WeChat (an instant messaging and video/audio chat app in China) to connect D2A and D2B so that both of them could see each other's video image. In this case, D2A could see D2B's video image from her laptop and Screenshot 24 shows her gaze direction. When looking at her own video image, D2A's gaze direction was towards the researcher's mobile phone camera (Screenshot 25). D2A was holding her mobile phone in one hand so that her video image could be seen by D2B. As for D2B, she could see D2A's video image (through the researcher's mobile phone) and her own video image from her laptop. This information was used to identify gaze directions during the multimodal analysis.



Screenshot 24: D2_T6_technical issue_1



Screenshot 25: D2_T6_technical issue_2

This solution was not perfect because the mediating role of the technology was more complicated and the size of D2A's video image became even smaller. This physical setup also limited the use of gestures by D2A as she had to hold her mobile phone in one hand. The changes posed obvious challenges to the dyad's meaning negotiation and to their use of multimodality during the video exchange. However, this scenario was the only technologically possible solution. This incident also demonstrates how technological issues affected video SCMC. By analysing students' performance under this technical condition, the researcher aims to identify how students made use of their multimodal resources in this imperfect technological environment to negotiate meaning and communicate with each other.

7.2.4.1.2 Turn-by-turn multimodal analysis of the episode

In Turn 1, D2A introduced the words 'nail polish', and repeated the phrase after a short pause. D2B was still looking up a word from their previous conversation and did not look at D2A while she was talking. This was a trigger stage and D2A used a comprehension check to see whether D2B understood the two words. D2B replied to her by repeating the trigger, but with a wrong pronunciation, which sounded like 'nail college'. This was a confirmation of the trigger stage as proposed in the interaction analysis of meaning negotiation routines in audio SCMC (see Section 5.2.2), because D2B was trying to confirm what exactly the trigger was. She also

frowned and touched her nose. D2A subsequently revealed, in the stimulated recall interview, she was looking at her peer's video image and considered D2B's mispronunciation and her 'confused' facial expressions as signs of non-understanding (Video MNE VSRI Extract 4.1). Therefore, in Turn 3, D2A gave D2B an explanation of the meaning of 'nail polish', '*it's used to make your nails colourful*'. In Turn 4, D2B had a slight pause and sighed, then said 'sorry' while scratching her hair. As revealed later in the exchange (Turn 6) and in the VSRI (Video MNE VSRI Extract 4.2), D2B here was hearing extraneous noise in her earphone and did not hear D2A clearly. But she did not point it out in Turn 4.

Video MNE VSRI Extract 4.1: D2A identifies the non-understanding

Researcher: OK, here, when she, when you, she repeated nail polish... and you

D2A: she didn't pronounce the word correctly

Researcher: then how did you know that she didn't understand?

D2A: if she understand [sic] she wouldn't pronounce the word incorrectly

Researcher: OK

D2A: so I need to explain it to her

Researcher: were you also watching her video image?

D2A: yes, I was watching

Researcher: and did you tell that she didn't understand?

D2A: ah, her visual expressions seem to be a little bit confused

Researcher: OK

D2A: yea

Researcher: OK, right, ah... let's move on

Video MNE VSRI Extract 4.2: D2B troubled by extraneous noise

Researcher: so, it was a listening problem

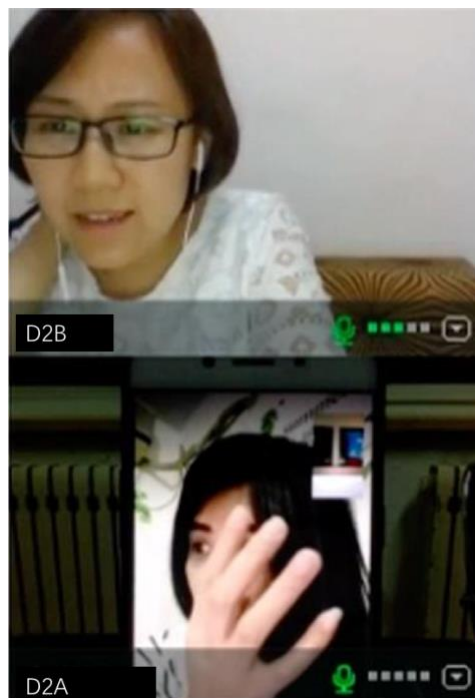
D2B: yes, I can remember, at this time, there are some noise, noise in the background, so I can't hear her very clearly

Researcher: OK

D2B: yes

Researcher: OK, no problem, let's continue (playing the recording)

D2A did not know exactly what problems D2B was experiencing, so in Turn 5, she first used a comprehension check trying to clarify the indicator and offered some further explanation '*girls like it*'. After seeing D2B's hesitation, D2A used another technique, to scale down the trigger from 'nail polish' to 'finger nails', and performed a comprehension check again by asking '*do you know finger nails?*'. In Turn 6, D2B first tried to repeat the trigger but she could not manage to do it so she explained that her actual problem was not the meaning, but the fact that she could not hear D2A very clearly. While D2B was clarifying her problem in Turn 6, D2A thought D2B still did not understand what a nail was, so she showed her hand in front of the camera of her mobile phone for the first time, trying to use the visual image as multimodal input to demonstrate the meaning of finger nails to her interlocutor (see Screenshot 26 and Video MNE VSRI Extract 4.3). Nonetheless, it seemed that D2B did not notice D2A's video image because she showed no reaction to D2A's very obvious hand movements. Later in the conversation when D2B notice D2A's gestures, she responded very clearly and quickly to them.



Screenshot 26: D2_T6_nail polish_Turn 6

Video MNE VSRI Extract 4.3: D2A shows her nails to the camera

Researcher: OK, so why did you do it? how did you think of showing her

D2A: uhm, uhm, I think she didn't know the word nail and I was trying to make her say what is the nail

Researcher: OK

D2A: I was waving my hand

Researcher: OK

D2A: showing her nails

Researcher: OK

D2A: yea

When D2B said she could hear clearly, a natural reaction from D2A might conceivably have been to ask how to adjust the audio mode and solve this issue. But since this was a multimodal environment, the audio channel was not the only way of communication. A video image was available and was possibly more effective for explaining the meaning of the word at this point of their interaction. Judging from D2A's previous hand gestures, it seemed that D2A's primary focus at this point was on the visual mode, while D2B's focus was on the audio mode. So in Turn 7, instead of following her peer's issue about audio mode, D2A asked D2B if she could see her, and repeated the question again '*can you see me? can you see my picture?*'. On the one hand she wanted to ask whether D2B could see her, because they had some technical issues during this task interaction. On the other hand, it was likely that D2A wanted to attract her peer's attention to the video image as the priority at this point of their negotiation, because it would be helpful for D2B's understanding of the words 'finger nails'. A very important point was reached when D2A realised that she and her interlocutor were prioritising different modes, so she tried to direct their use of modes before negotiating the cause and meaning of the non-understanding.



Screen shot 27.1 Turn 7

The researcher moved the mobile phone screen to make D2A's picture clearer

Screen shot 27.2 Turn 5

The previous position of D2A's picture in the interaction

Screenshot 27: D2_T6_nail polish_Turn 5 and 7

D2A's question leads their discussion from negotiating meaning to negotiating the use of mode. In Turn 8, D2B first offered a positive answer to D2A's question 'yes, I can see you', and then quickly moved to her own focus on the audio mode again, insisting on pointing out that she could not hear clearly because of some noise. This time, in Turn 9, D2A attended to D2B's problem and replied to D2B that she could hear her clearly.

Since their communication was mediated through the researcher's mobile phone, when D2B stressed she could not hear clearly due to some noise, the researcher adjusted her mobile phone, in order for both interlocutors to hear and see each other clearly. Screenshots 27.1 and 27.2 show the process of the researcher adjusting her mobile phone position. Note that in both Turn 5 and Turn 7, which occurred before and after D2B pointed out her problem with the audio mode, D2A was always holding the speaker of her headset for D2A to hear clearly. Actually,

throughout the whole task interaction, D2A always held her speaker whenever she spoke to D2B except when she was showing her hand in front of the camera. This shows that she was highly aware of the potential hearing problem for D2B and was trying hard to avoid it. Possibly because the researcher's adjustment of her mobile phone position, D2B was finally able to hear her interlocutor clearly and she confirmed to D2A in Turn 10 'oh, it'sOKnow'.

Having solved the audio problem, D2A quickly moved back to negotiating meaning by giving another comprehension check of the original trigger in Turn 11 '*OK, so do you know what is nail polish?*'. D2B in Turn 12 repeated the words, still mispronouncing them as nail 'college' with a rising intonation. This was again a clear indicator of non-understanding from D2B. D2A replied by correcting D2B's pronunciation and stressing the word 'polish' twice in Turn 13. Soon after saying this, D2A showed her hand in front of the camera again, waving her fingers, trying to catch her peer's attention even before D2B's verbal reply (Screenshot 28). However, D2B missed D2A's hand gestures again. Her gaze direction suggested that she was first looking at her notes and then up to the screen but somehow, she still failed to notice her peer's gestures (Screenshot 29). This could be evidenced by her verbal reaction in Turn 14, which was a repetition of the trigger, as nail 'bolish'. She also looked confused, seemingly trying to get her partner to confirm her pronunciation (Screenshot 29.1). Both her verbal and visual reactions seemed to suggest that she still did not notice, or at least had not made sense of, D2A's hand movements in front of the camera.



Screenshot 28.1 Turn 13

D2A just finished saying polish and putting down the microphone.

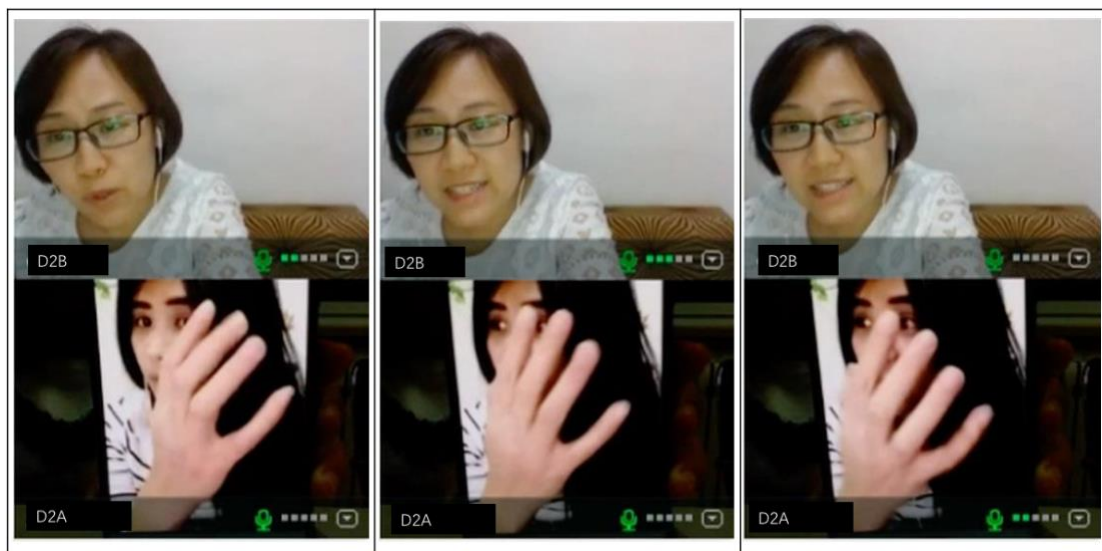
Screenshot 28.2 Turn 13

D2A put down the microphone and putting her hand towards the camera, but D2B did not see her hand movements.

Screenshot 28.3 Turn 13

D2A put her hand right in the centre of the camera and was looking at her own camera to adjust her hand position but D2B was trying to repeat the word polish and did not see D2A's hand movements.

Screenshot 28: D2_T6_nail polish_Turn 13



Screenshot 29.1 Turn 14 &15

D2B was repeating polish and looking at her peer's video image. D2A was showing her hand in front of the camera and looking at her own video image.

Screenshot 29: D2_T6_nail polish_Turn 14 &15

Since D2A still had not seen D2B's reaction to her hand gestures, D2A in Turn 15 asked again. *'can you see me? nails, do you know what is nails? nails'*. This turn includes both negotiation and a comprehension check as part of the meaning negotiation routine. D2A clearly showed her hand and nails right in the centre of her video image. D2A held this gesture for 12 seconds, which was very long in video SCMC, and demonstrated her insistence on showing her peer the meaning of the word through the video channel. And she also scaled down the trigger again by focusing only on 'nail' while showing her nails in front of her camera, trying to combine both the verbal and visual elements to explain the meaning of the word. It was important to point out that D2A asked the question 'can you see me?' before repeating the trigger, in other words, she negotiated the use of mode before negotiating meaning. Most probably, D2A knew that D2B could see her because D2B had confirmed this to D2A 30 seconds ago in Turn 8. Therefore, when D2A asked D2B whether she could see her for the second time, it appeared to be intended as a reminder to draw her interlocutor's attention to her video image, rather than a genuine question about visibility. This was the second time that D2A tried to negotiate the use of mode with her peer in the meaning negotiation episode.

Luckily, this time, after having been reminded by D2A, in Turn 16, D2B raised her head to look at D2A's video image while D2A was showing her hand and nails in front of the webcam. D2B saw her peer's hand movements and repeated the word 'nail' (Screenshots 30.1 and 30.2). That was the moment when she finally realised the meaning of the word nail. She quickly confirmed her understanding by saying 'oh, yes, yes' with a nod (Screenshot 30.3). After nodding, she moved her head slightly closer to the webcam and repeated the word nail again. During the stimulated recall interview, D2B confirmed that she achieved the understanding of the word 'nail' through seeing D2A's hand and nails from the video image (Video MNE VSRI Extract 4.4).



<p>Screenshot 30.1 Turn 16 D2B was repeating the word nail and looking at her peer's video image. D2A was showing her fingernails in front of the webcam.</p>	<p>Screenshot 30.2 Turn 16 D2B just finished repeating the word nail and was still looking at her peer's video image. D2A was showing her fingernails in front of the webcam.</p>	<p>Screenshot 30.3 Turn 16 D2B was saying 'oh' and nodding. D2A was showing her fingernails in front of the webcam.</p>	<p>Screenshot 30.4 Turn 16 D2B was moving closer to her webcam, looking at the webcam and was saying yes. D2A was still showing fingernails in front of the webcam.</p>
---	---	---	---

Screenshot 30: D2_T6_nail polish_Turn 16

Video MNE VSRI Extract 4.4: D2B understands the meaning of 'nail'

Researcher: oh, so when did you start to understand the word nail?

D2B: at this time, I know that the nail is 指甲 [nail]

Researcher: when?

D2B: when she hang [sic] up her hands and show nail to me

After D2B's clear confirmation of her understanding, in Turn 17, D2A quickly drew her hand away from the camera, held her headset microphone and reminded D2B of the full trigger, '*nail polish*', followed by a comprehension check '*do you know that?*'. Here, there was an overlap between D2A's question and D2B's answer. Right after D2A said '*nail polish*' before the comprehension check, D2B started smiling and confirmed '*yeah I know that*' in Turn 18. D2A was also smiling when D2B said she knew that (Screenshot 31).



Screenshot 31: D2_T6_nail polish_Turn 18

It seemed that in Turn 18, D2B already understood the meaning of 'nail polish', but it turned out to be a misunderstanding. According to the stimulated recall interview, at this point, D2B was thinking about a 'nail clipper' (see Video MNE VSRI Extract 4.5).

Video MNE VSRI Extract 4.5: D2B misunderstands 'polish' as 'clipper'

Researcher: yeah, so what were you thinking then? in Chinese?

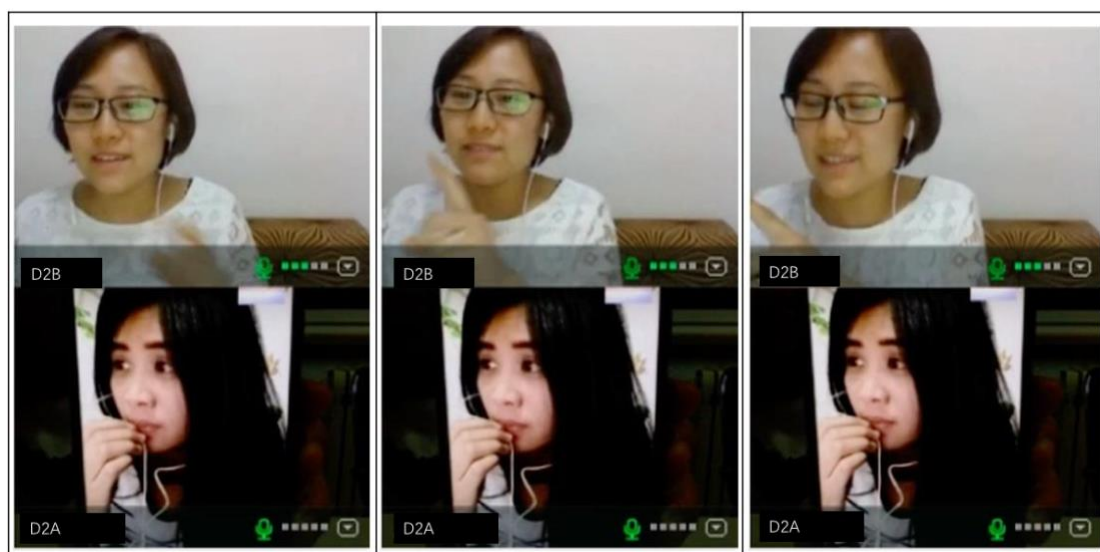
D2B: 指甲刀 [nail clipper]

Researcher: OK, OK

Despite D2B's smile and confirmation, D2A in Turn 19, persisted in offering her explanation of the 'nail polish', which was *'the thing that makes your nails colourful'*. There was another overlap, as D2B repeated the word 'nail polish' (Turn 20) while D2A was explaining the word to D2B. So D2B probably did not fully focus on listening to D2A's explanation. After the overlap, D2A performed another comprehension check *'do you know it now?'* in Turn 21. It seemed that D2B's smile and verbal confirmation were not sufficient for D2A to completely believe that D2B

understood, so she kept on asking comprehension check questions, trying to gain more details from D2B about the meaning of the words 'nail polish'.

In Turn 22, D2B was trying to explain her understanding of the word; accompanying this effort with gestures, as D2A had done. Note that, at this point, she misunderstood 'nail polish' as 'nail clipper'. While saying '*I know what it is*' as a verbal confirmation, she moved her fingers to the left side of the screen, trying to show the gesture of cutting her nails. Screenshot 32 shows the movements of D2B's hand movements from below the frame (Screenshot 32.1) to the left side of the frame (Screenshot 32.2) and then out of the left side of the frame (Screenshot 32.3). Although D2A was looking at her peer's video image all the time, she could not see D2B's hand gestures since they were out of frame.



Screenshot 32.1 Turn 22

D2B was saying 'I know what it is', while trying to use hand gestures to show what she meant in front of the webcam. D2A was listening to and looking at D2B.

Screenshot 32.2 Turn 22

D2B's finger moved to the left side of the frame, while looking at her peer's video image, and saying 'I know what it is'. D2A was listening to and looking at D2B.

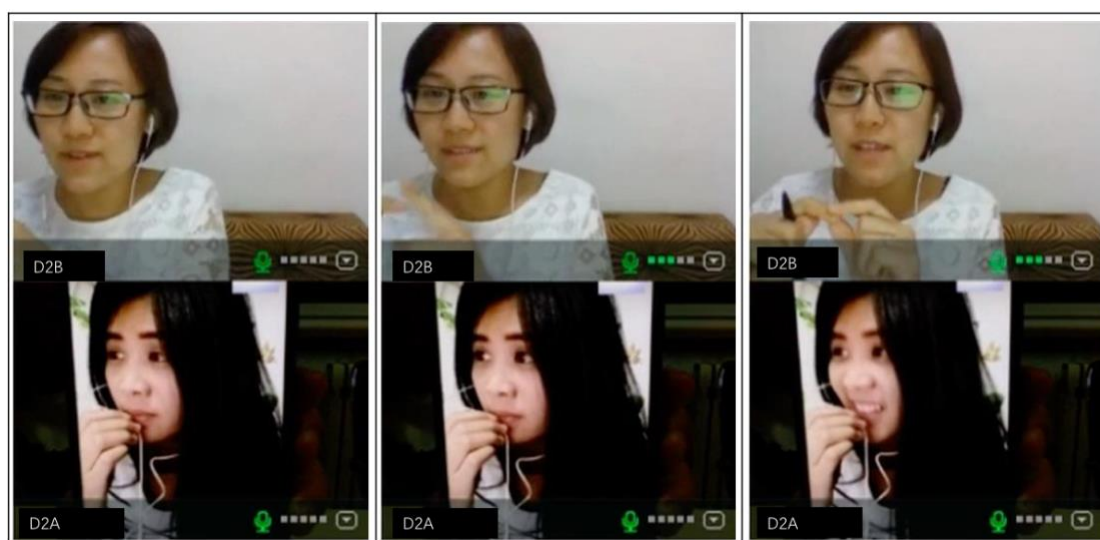
Screenshot 32.3 Turn 22

D2B's finger moved outside the left side of the frame and was looking at her own finger gestures. D2A was listening to and looking at D2B.

Screenshot 32: D2_T6_nail polish_Turn 22_1

However D2B had a quick look at her own video image (Screenshot 33.1) and realised that her hand gestures were out of the frame. So she quickly moved her hands to the right side (Screenshot 33.2) and then performed the gesture of cutting her finger nails within the camera frame (Screenshot 33.3) so D2A could see. Meanwhile she also gave a verbal explanation, '*if your nail longer, you can use it, make it short*', followed by a short clarification request '*right?*'.

Two points are important in D2B's performance in Turn 22. First, D2B had seen D2A using gestures to show the meaning of the words 'nail polish', and very quickly learned to use gestural means to convey her understanding to her peer. This showed her ability to quickly learn and use video to make their communication more effective. Another point is that she looked at her own video image while doing the hand gestures and noticed that her hands were out of the frame and so quickly brought her hands back into the camera frame so that her peer could see them. This response showed that she was aware of how her gestures were captured by the camera and what could be seen, or not. As a result of this awareness D2B was able to make quick adjustments to ensure the increasingly effective use of the visual mode in their communication. D2B's performance in this turn demonstrated a high level of multimodal communicative competence, which is crucial to the success in negotiating meaning between the members of the dyad.



Screenshot 33.1 Turn 22
D2B was saying 'make the nail', while trying to use

Screenshot 33.2 Turn 22
D2B was saying 'make the nail', while trying to use

Screenshot 33.3 Turn 22
D2B was saying 'if your nail longer, you can use it, make

gestures to show her explanation but her finger was out of frame. D2A was listening to and looking at D2B.	finger gestures to show her explanation. D2B moved her hands to the right side for her fingers to be seen in the frame. D2A was listening to and looking at D2B.	it short, right?' while performing the 'nail clipper' gesture with her hands in front of the camera. D2A was looking at D2B and smiling.
--	--	--

Screenshot 33: D2_T6_nail polish_Turn 22_2

Knowing that D2B was labouring under a misunderstanding, D2A quickly rejected D2B's guess by saying 'no' four times, while slightly shaking her head in Turn 23. After the negative response from D2A, D2B repeated the word and asked another confirmation check, '*nail polish, right?*', at the same time, D2B was also making the gesture of cutting her fingernails in front of the camera with a smile on her face (Screenshot 34).









Screenshot 34: D2_T6_nail polish_Turn 23_1

Before D2A started explaining again, D2B suddenly realised the meaning of 'nail polish'. According to the stimulated recall interview, D2B said she '*suddenly remember the information colourful*' and '*the girl, she (D2A) mentioned girls like it*', which enabled her to finally guess the correct meaning of 'nail polish' (Video MNE VSRI Extract 4.6). She expressed her 'sudden

248

realisation' both verbally by saying loudly '*oh, I know, I know that*' and visually with a big nod, touching her hair and raising up her right hand (Screenshot 35). D2B then quickly started explaining the meaning of nail polish, '*you can use it make your nails colourful*' and offering examples, such as '*maybe you can use red nail polish, black nail polish*', followed by a confirmation request, '*right?*'. Finally, after D2B offered her own explanations and examples of 'nail polish', D2A was convinced that D2B had arrived at the correct understanding, as was recalled by D2A during the interview (Video MNE VSRI Extract 4.7). Then, D2A quickly confirmed the accuracy of D2B's supposition with a 'yes' and moved to the next phase of their task discussion, which marked the end of this meaning negotiation routine.

		
		
<p>Screenshot 35.1 Turn 23 D2B was saying 'make the nail', while trying to use finger gestures to show her explanation but her finger gestures were out of frame to the left s. D2A was listening to and looking at D2B.</p>	<p>Screenshot 35.2 Turn 23 D2B was saying 'make the nail', while trying to use finger gestures to show her explanation. D2B moved her hands to the right side for her fingers to be seen in the frame. D2A was listening to and looking at D2B.</p>	<p>Screenshot 35.3 Turn 23 D2B was saying 'if your nail longer, you can use it, make it short, right?' while showing the 'nail clipper' gesture with her hands in front of the camera. D2A was looking at D2B and smiling.</p>

Screenshot 35: D2_T6_nail polish_Turn 23_2

Video MNE VSRI Extract 4.6: D2B's correct understanding of 'polish'

Researcher: so you, you understood it before she said it again?

D2B: yes, because I suddenly remember the information, colourful, colour, she just mentioned,

Researcher: yeah

D2B: because first of all, I realised the meaning of nail,

Researcher: yeah

D2B: I know nail is 指甲 [nail]

Researcher: yeah

D2B: and I guess, the word, the item is 指甲刀 [nail clipper]

Researcher: OK,

D2B: I confirmed it with her, and she want to explain more, and I suddenly remember the information colourful

Researcher: OK

D2B: and the girl, she mentioned girls like it

Researcher: yeah

D2B: so I suddenly realised this is 指甲油 [nail polish]

Researcher: yeah, OK, that's good

D2B: yes

Video MNE VSRI Extract 4.7: D2A's final confirmation

Researcher: now you know that she completely understood it?

D2A: yea

7.2.4.2 The use of multiple modes and semiotic resources in this episode

This meaning negotiation episode was exceptionally long, consisting of 25 turns in total. In general, the negotiation follows Smith's (2003) meaning negotiation routine, as it included all six stages of his model. However, this episode also incorporated stages that were not present in Smith's model. These were the explicit discussion of the use of modes in Turn 6, 7, 8, 9, 10 and 15. From Turn 6 to 10, D2A attempted to prioritise the visual mode so that she could express the meaning of the words through gestures. But D2B persisted in focusing on the audio

signal because there was interference and she could not hear her peer clearly. So the two interlocutors were negotiating mode and trying to solve the technical issue. In Turn 15, D2A managed to draw her peer's attention to her video image by asking again '*can you see me?*' although she most probably knew that D2B could see her video image. This was the second example of mode negotiation initiated by D2A. Her reminder to use the visual mode was very successful because D2B quickly understood the meaning of the word 'nail' after watching D2A's video image. Furthermore, D2B managed to quickly learn to use visual means properly to express herself to D2A. In this meaning negotiation episode, both interlocutors managed to use gestures to express the meaning of a previously unknown word and finally arrive at a correct understanding of it. Both interlocutors were highly aware of how they were framed by the webcam; in other words, they were aware of what their peer can see of their video images. For example, D2A first used gestures, right in the centre of her video image so that D2B could see her clearly. Soon after that, D2B adjusted her hand position from the left side of the frame to the middle so that D2A could see her gestures in the video image. Both interlocutors were good at using the multimodal resources effectively to facilitate the negotiation of meaning. This was a key reason for their final success in this unusually long meaning negotiation episode.

This episode also showed from a qualitative perspective how gaze, especially looking at the peer's video image, could be helpful for creating effective meaning negotiation. The previous gaze analysis offered statistical evidence of the positive relationship between looking at the peer's video image and the successful outcome of a meaning negotiation interaction. This multimodal analysis of a specific meaning negotiation episode illustrated this finding in more detail from a qualitative perspective. From the screenshots and the video clip, it could be seen that D2A was mostly looking at her peer's video image during their negotiated interaction and that only from time to time did she look at the task sheet or at her own video image to check how she appeared on screen. D2B sometimes took notes and looked down towards the task sheet, but when necessary D2A verbally directed D2B to concentrate on her video image when she (D2A) was about to use gestures to express her ideas. Except for taking notes and sometimes looking at the task sheet, D2B also mostly looked at her peer's video image. By

looking at each other's video images, they both obtained a wealth of important visual information during the negotiated interactions. This was another key reason for their success in negotiating meaning.

7.2.4.3 Attitudes towards multimodal SCMC: evidence from interviews

During the interview, when asked about her attitude towards audio and video conferencing, D2A showed a strong preference for video SCMC because it is more '*interactive*': (a) she can see her interlocutor's facial expressions, (b) use gestures to explain the words, and (c) tell whether her peer understood or not (Video MNE VSRI Extract 4.8). As a result of technical problems, one of her hands was holding her mobile phone, which prevented her using more manual gestures. In the interview D2A said that if their exchange had not been impacted by technological issues, she would have used '*more hand gestures to make it easier*' for her peer to understand. This shows her high level of awareness of the visual mode and an equally high level of competence in using it for more effective communication in the video SCMC environment. When it comes to the gaze issue, D2A confirmed that she cared how she looked in front of the webcam but most of the time, she focused on looking at her peer's video image, rather than the task sheet or her own video image.

Video MNE VSRI Extract 4.8: D2A outlines the advantages of multimodality

Researcher: when we compare video and audio, which one do you prefer
D2A: I prefer video conferencing because it is more interactive, I can see who I'm talking with, I can see their facial expressions so it is, make class more interesting and make me want to participate more, participate more
Researcher: OK, so, ah, compared to audio, what more information does it give to you in video?
D2A: ehm, what, just like you know, when we, I'll give you an example
Researcher: OK
D2A: just like we were talking about nail polish, there is no way I could have shown what is a finger

Researcher: ehm

D2A: it may take me more, a lot longer time for me to explain what it is, video is more helpful, it is more interactive

Researcher: and do you think video mode is easier for you to see if she understand [sic] or not

D2A: yes, sure, because if there's only audio, I can only tell from her voice

Researcher: OK

D2A: by plus video, I can see her facial expressions ... if we had a better condition, you know, we can see each other, and I can use my hands, maybe I can use more hand gestures, to make it more, to make it easier for her to understand

Researcher: yes, OK, OK, and then it was camera, when you were doing video conferencing task, there was a video camera and there was a task sheet, was it hard for you to take care of both

D2A: no, because after I had first rough understanding of task sheet, I don't need to focus all the time on the task sheet, I can see her video naturally, I don't need to focus on the task sheet

Similarly, in her interview, D2B also expressed a preference for video over audio SCMC because video communication is 'more straight forward, and she could see her peer's hand gestures, facial expressions and get more comprehensive information (see Video MNE VSRI Extract 4.9). But D2B also complained about the technical problems encountered in the audio channel during video communication, as was evident in Turns 6 to 10 of the interaction. She insisted, however, that '*if we ignore the technical issues, video is better*'. This comment shows her positive attitude towards video SCMC, and that she was aware of the wide variety of useful information afforded by the use of video, which was why she also looked at her peer's video image frequently during their task interaction. When asked about whether D2B cared how she looked in front of the webcam, she said when she was so fully engaged in the conversation that she forgot the presence of camera. This suggests that D2B was highly engaged in the task interaction. Moreover, it also reveals the intensive, absorbing and sometimes exacting nature of multimodal interaction.

Video MNE VSRI Extract 4.9: D2B also prefers multimodal SCMC

Researcher: just a few more questions about this course in general, ehm, so in the first class, you had two audio sessions, and the second day's class you had two video sessions, and compared audio and video, which one do you prefer? which one do you like better?

D2B: which one like better?

Researcher: yeah

D2B: 就是视频一是有一点声音之后，第二是我听起来 D2A 那边有一点杂音，有些时候我听不太清楚她在说什么，但是如果排除这些技术的问题的话，我觉得还是可以看见对方，比如那个 nail，就是她可以去给我描述，我会看到她的表情，好像更好一点 [in video communication, sometimes, there's some delay in the sound, and sometimes there are some noise from D2A's side, I couldn't hear her clearly. But if we exclude these technical issues, I think it's better if I can see my partner because she can describe things and I can see her facial expressions, for example, the nail]

Researcher: OK, so what do you think are the advantages of video or what additional information does video give to you compared to audio?

D2B: ehm, video, 就是更直观吧 [video, it's more straight forward], maybe I can get more information from the pictures,

Researcher: information like?

D2B: ehm, like 就是手势啊，表情啊，就是我可以看见对方的手势表情，得到的信息更全面吧 [like hand gestures, facial expressions, when I can see the partner's hand gestures, and facial expressions, I can get more comprehensive information]

Researcher: 就是你会不会关注这个 webcam

D2B: 就是我又要看屏幕又要看这个纸，会不会影响到我？ [you mean, I have to look at the camera and also look at the task sheet, is it hard for me]

Researcher: 嗯，对 [yes, right]

D2B: 不会 [no]

Researcher: 不会？ [no?]

D2B: 不会 [no]

Researcher: OK, that's good, do you care a lot about how you look in the camera?

D2B: ehm, to be honest, at the beginning of the talking, maybe I can I care about the way I look in the picture, but when we are very 投入怎么说 [how to say engaged?],就是非常投入到我们的 talking 里的时候, 就已经不记得就已经忽略了自己在里面的形象了, 就已经是沉浸在这个 talking 中了[when we were very immersed in our conversation, I already forget and ignore how I look in the camera, I completely focused on the talking] 但是在一开始我自己可能会注意这里头发有点翘, 会自己整理一下, 但是说到最后的话, 就不太注意, 不太 care 那个形象了[in the beginning I may look at myself from the camera and tidy my hair, but in the end, I don't care how I look for the camera]

This meaning negotiation episode is a good example in that it shows: (a) how multiple modes and semiotic resources can be used effectively to promote meaning negotiation and (b) that directing the use of mode in a multimodal SCMC environment could be helpful for negotiating meaning.

7.2.4.4 Comparison and findings for Dyads 2 and 3

While Dyad 2 members were very competent in using different modalities to negotiate, the participants in Dyad 3 offered a contrasting example of how the inability to use multimodality could lead to the failure of a negotiated interaction. Comparing D3B's hand gestures in negotiating the meaning of 'magnifying glass' and D2A's hand gestures in the 'nail polish' episode, we could see three clear differences. First, D2A explicitly asked her peer 'can you see me?' twice, to negotiate the use of modality with her peer and draw her peer's attention to the visual mode. But D3B did not draw D3A's attention to her gestures in any way. Second, when D2A showed her hand and nails in front of the camera, she placed her hand centrally in front of her webcam so that her gestures were clear and obvious enough to be seen by D2B. Similarly, when D2B wanted to mime the action of a nail clipper, she first erroneously held her hand outside the frame to the left, but quickly adjusted her hand position to the middle of the screen so that her peer could clearly see her. However, D3A's gestures, in attempting to explain 'magnifying glass' were in the lower left-hand corner of the screen, and so were partly obscured

by her name; unfortunately, she made no conscious adjustment of her hand positions. During this episode, D3B did not show any awareness of whether her gestures could be seen by her peer within the video frame. Third, when D2A was trying to show the meaning of 'nail', she was waving her hands and showing her nails in front of the camera constantly and repeatedly (for 12 seconds) until D2B finally guessed the correct meaning. However, D3A's hand gesture was very fleeting (less than a second) and was not repeated or emphasised in any way.

These three differences illustrate the different levels of ability in using the visual mode displayed by D2A and D3B. D3B seemed to have performed this gesture as an intuitive accompaniment to her own speech. But D2A demonstrated a good awareness of the computer mediation, and a deep understanding of how the video image mediates communication in a multimodal SCMC environment. In other words, it seems that D3B acted the same way as in a face-to-face communication, while D2A was able to understand and deal with the mediated nature of online multimodal communication. The comparison between the two participants' performances shows how the ability to use multimodal resources can affect the meaning negotiation process in video SCMC environments.

7.3 Findings

7.3.1 Proposing a multimodal communicative competence pyramid

Having analysed four individual episodes, this chapter moves on to summarising and comparing different dyads' multimodal performances to further examine how multiple modes were used to negotiate meaning in video SCMC. This study uses 'multimodal communicative competence' (MCC) to refer to students' ability to make use of multiple modes to negotiate meaning in video SCMC.

Participants demonstrate different levels of competence in making use of multimodal resources. For example, compared to D4A, who did not recognise the different affordances of audio and video SCMC, D4B showed some awareness of them. However, despite knowing that the visual mode could offer more information about the peer's (non-)understanding, D4B still did not look at her peer's video image frequently during their negotiated interactions. Compared to D4B, D1A and D1B not only had the awareness, but also looked at their peer's video image frequently to intentionally gain more clues about that peer's thoughts. In addition, D1B also intentionally used smiles, nods, and verbal communication to pretend she understood. Despite her negative attitude towards meaning negotiation, D1B was able to make flexible use of multiple modes to present herself clearly in front of the webcam. Similarly, in Dyad 3, both interlocutors were highly aware of the rich visual information available to them and often looked at each other's video image. As for presenting themselves in front of the webcam, D3A clearly outperformed D3B. While D3B made her 'enlarging' hand gesture quickly at the corner of the screen, D3B carefully adjusted her posture in relation to the webcam and put her hands around her eyes to make the iconic gesture of glasses in the middle of the screen for many seconds for her peer to see clearly. This gesture demonstrated her competence in making use of multiple modes and semiotic resources to negotiate meaning in video SCMC

Moreover, it seems that these four competences seem to build cumulatively on each other from (a) to (d). To illustrate, Table 25 summarises the individual competences demonstrated by each participant ('√' means yes; '/' means no). The table shows that those who possessed higher-level competences also demonstrated ownership of the basic ones. Therefore, the four competences seemed to be hierarchically ordered. Based on the above comparisons and analysis, a new model of multimodal communicative competence is then proposed in Figure 14.

Participant	Competency (a)	Competency (b)	Competency (c)	Competency (d)
D4A	/	/	/	/
D4B	√	/	/	/
D1A	√	√	/	/

D3B	√	√	/	/
D1B	√	√	√	/
D3A	√	√	√	/
D2A	√	√	√	√
D2B	√	√	√	√

Table 25: Participants' multimodal communicative competences

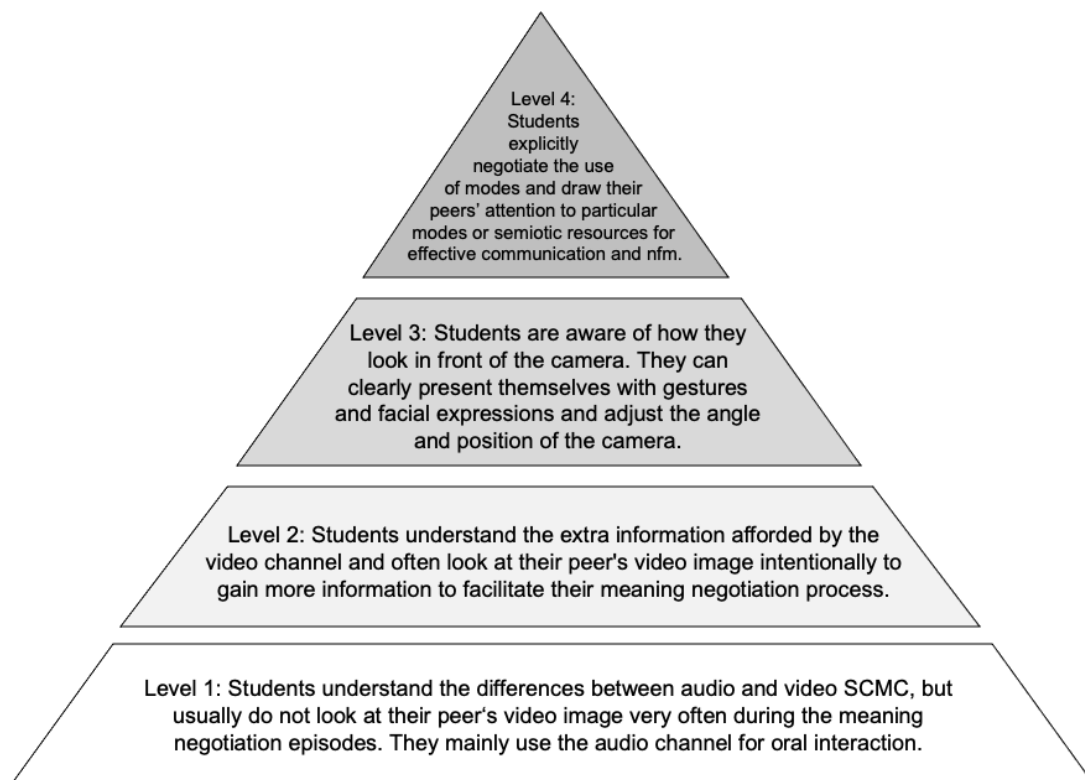


Figure 14: The multimodal communicative competence pyramid in video SCMC

At the most basic level (Level 1) students understand the differences between audio and video SCMC. Specifically, they are aware that they can gain more information from their peer's video image, through 'body language' such as: a) facial expressions, b) gestures, c) smiles, and d) nods. However, despite their awareness of the affordances of the video channel, students at this level do not look at their peer's video image very often during meaning negotiation episodes. They mainly use the audio channel for oral interaction. For example, D4A did not have this basic level of competence as she showed no realisation of the differences between audio and

video SCMC; however, her peer D4B was able to show this level of multimodal communicative competence (see Section 7.2.1).

The second level of MCC refers to those who not only understand the extra information afforded by the visual mode but also often look at their peer's video image intentionally to gain more information about their peer's gaze, facial expressions, and gestures. In other words, they both have the understanding and ability to take action during meaning negotiation. This level focuses on the viewer's perspective, when looking at the video image of their peers. For example, D1A and D2B both looked at each other's video image during the 'razor' episode (see Section 7.2.2).

While the second level focuses on the viewer's perspective in front of the screen, the third level deals with the presenter's perspective, in front of the camera. This level of MCC means students are aware of how they look in front of the camera. They are able to clearly present themselves in front of the camera, for example by repeating gestures many times, and carefully placing their gestures within the frame. This conduct is more effective than making natural gestures accompanying their speech, without any intentional adjustments. A good example of this aptitude was when D3A wanted to show a pair of glasses: she moved forward towards the camera and circled her eyes with her fingers. She also repeated this gesture many times for her peer to see it clearly (see Section 7.2.3 in this chapter). Another example is when D2B misunderstood the meaning of 'nail polish' as 'nail clipper' and wanted to show the gesture of cutting her nails (see Section 7.2.4 in this chapter). At first her hands were out of frame to the left, but she quickly noticed this and moved her hands slightly to the right so that her gestures are clearly in frame and therefore visible to her peer. The third level also includes the ability to make use of multiple modes and semiotic resources such as gestures and facial expressions to present themselves in front of the webcam. Students at this level are also able to adjust the webcam angle and position so that the camera can capture as much as possible of the presenter's face and body, to capture as much useful visual information as possible. This skill also ensures that the space captured in the video image is not wasted on unimportant information (e.g. background wall). For example, in Dyad 4 Task 3, the top third of the video

image of D4B was filled by the background, and her face occupies only the lower two-thirds of the screen, while none of her 'movements or gestures below neck level were captured by the camera (see Section 7.2.1). With so little useful visual information, it was hard for her peer to judge her levels of understanding or incomprehension during meaning negotiation episodes.

The highest level of multimodal communicative competence (Level 4) means students can clearly direct the peer on the choice of multiple modes and semiotic resources both as a viewer in front of the screen and as a presenter in front of the webcam. An appropriate example of high level MCC is when D2A asked D2B '*can you see me?*' twice in their negotiation for the meaning of 'nail polish' to draw her peer's attention on the video image; (see Section 7.2.4).

This multimodal communicative competence pyramid is based on the multimodal performances of four dyads (eight participants) during selected critical incidents (arising from 37 meaning negotiation episodes) in the current study. Conceivably, on top of the highest layer of the current pyramid, there might exist an even higher level of multimodal communicative competence where interlocutors are able to make flexible and appropriate use of different modes and semiotic resources without having to verbally negotiate the use of modes. In other words, if nothing else, this pyramid provides at least a basis for further development.

7.3.2 The relationships among different modes/channels in meaning negotiation in video SCMC

The multimodal analysis of the episodes has demonstrated different relationships among multiple modes and semiotic resources. The following analysis will summarise these relationships and offer specific examples. Findings of the multimodal analysis suggest that these modes and semiotic resources in video SCMC interactions are clearly hierarchically ordered and can therefore compliment/reinforce each other, or even compete against each other.

7.3.2.1 Hierarchy order

First of all, it is clear that the audio channel was still used as the default channel for communication in video SCMC negotiated interactions for all dyads, as most of the meaning negotiations were conducted orally using the audio channel. Even when the audio signal was disrupted, interlocutors (e.g. Dyad 2) attempted to report this problem to each other and to solve the technical problem using oral means, instead of resorting to an alternative mode to continue their negotiations (see Section 7.2.4). This behaviour demonstrated the hierarchy order, where orality was the prioritised mode for communication in video SCMC interactions.

7.3.2.2 The reinforcing/complementary relationship

Another type of relationship identified is that two modes/channels were used by participants in combination to express their ideas more explicitly in a mutually reinforcing or complementary way.

The reinforcing relationship occurred frequently in the reaction to response or reconfirmation stage (Smith, 2003), when the respondent confirmed her understanding or expresses agreement. For example, in the 'nail polish' episode (see Section 7.2.4), D2B first confirmed her understanding by saying '*I know that*' and smiling to the webcam at the same time. Another example is when D1A agreed to choose a razor for a paternal present. D1B nodded slightly twice while expressing her agreement by saying '*I think it is, maybe*' (see Section 7.2.2). While this sentence was a task-related response after the meaning negotiation routine, it still demonstrated how either speech or gesture was used to express the same meaning. In both examples, the respondents used both an oral confirmation and a facial expression (smile) or gesture (nod) to express the same meaning. In this case, the aural mode and the visual mode were used simultaneously to reinforce a statement denoting understanding or agreement.

The reinforcing relationship was also found in a reaction to response stage during negotiations in video SCMC interactions. For example, when guessing the meaning of the words 'magnifying

glass', the respondent D3A used an iconic hand gesture mimicking glasses around her eyes while testing a deduction by asking '*a pair of glasses?*' (see Section 7.2.3). The interlocutor used both words and gestures to express the same meaning, which means that the oral mode and the visual mode were used to reinforce each other in order to express meaning more explicitly.

While the above reinforcing relationship between the oral/aural and the gestural/visual mode occurred during the confirmation stage, the complementary relationship was identified in the resolution stage, including responses and reactions to response stages in video SCMC interactions. For example, when Dyad 2 was trying to understand the words 'nail polish', the initiator held up her fingernails in front of the webcam and asked her peer to look at her video image (see Section 7.2.4). In this case the oral mode was used by the initiator to direct the use of mode and remind her peer to focus on the visual mode. However, the initiator also used gesture to clarify the meaning of the word by showing the actual fingernail in front of the webcam. In this case, the oral mode and the visual mode were used to complement each other in negotiating the use of mode and meaning at the same time.

7.3.2.3 Competing relationships

Although the relationship between the oral and visual modes was normally harmonious in video SCMC interactions, there also existed competition between different modes or semiotic resources. For example, in the 'salt shaker' example by Dyad 4 (see Section 7.2.1), the oral mode served as the default mode of communication throughout the episode, visually interlocutors had two divergent options. They could either fix their gaze on the hard copy of the task sheet on the desk or look at each other's video image on the screen. The choice of gaze direction determined what visual information the two students could gain during the negotiation and subsequently what they could produce in their oral communication. In this case, the hard copies of the task sheet with the spot-the-difference pictures placed on the table was an important semiotic resource for students to refer to, if they wanted to complete the task

successfully. Therefore, both interlocutors emphasised 'in your picture' at the end of almost every turn of their speech to reinforce each other's focus on the task sheet, thereby reminding their peer to look at the task sheet instead of each other's video image. This case demonstrates a clear competing relationship between the task sheet as one type of visual semiotic resource and the video image on the screen as another type of visual semiotic resource.

While a wide range of modes and semiotic resources were available for students to choose from during video SCMC interactions, this variety could also lead to cognitive overload for interlocutors, as is noted by Guo and Möllering (2016). First, participants needed to follow the task instructions and try to complete the task. In order to achieve success, the dyad members needed to look at the task sheet and think about the order in which to describe the things in the picture (for the spot-the-difference task in Appendices 3 and 4), or how to explain an unknown word (in the problem-solving task in Appendices 5 and 6). Furthermore, as non-native speakers, participants also needed to organise the language which they should use to talk to their peer. Meanwhile, they had to listen to their peers carefully to be able to negotiate the meaning of an unknown word or to complete the set task. In addition, from time to time, they needed to look at the webcam to seek further information from their peers' video image and/or to present some visual information through the webcam to their peers. Last but not least, there might even be some technical issues, such as noise interference in the audio or frozen images in the video. Both problems were encountered and needed to be resolved in order to establish smooth communication and task completion. All these things required a high level of attention (and concentration) from the participants. However, learners only had limited attention, and as a result, they had to prioritise different modes and semiotic resources, possibly opting to sacrifice particular less important things in order to be able to focus on the more important modes/semiotic resources that could help them to complete the task effectively and successfully. Interlocutors in multimodal SCMC interactions chose to focus on the semiotic resources or modes that were most likely to help them complete the task, rather than trying to encompass all the modes and semiotic resources available to them. Some of them even

directed the use of modes by their peers strategically, focusing on the more effective mode for the current interaction, thus reducing their cognitive load.

The fine-grained turn-by-turn multimodal analysis of video SCMC interactions in the current study identified specific episodes illustrating three different types of relationships among various modes and semiotic resources involved in multimodal video SCMC. A hierarchy exists, with the oral/aural mode being the default mode of communication. Meanwhile, the oral/aural mode and the visual mode could be used to either reinforce one another in expressing the same meaning, especially during the reaction to response and confirmation stages, or to complement each other during the response stage. However, two different visual resources could have a competing relationship when learners had to choose either one or the other to focus on. Within the limited scale of the current study's multimodal analysis, no examples were found where gestures or facial expressions were used to completely replace oral communication.

7.4 Chapter conclusion

This chapter first briefly introduced the rationale for marrying the critical incident technique to a case study model. One meaning negotiation episode from each dyad was selected for multimodal analysis. Students' video stimulated recall interviews were used within the multimodal analysis to demonstrate participants' levels of comprehension, as well as illustrating their thought processes at certain points. Their attitudes towards meaning negotiation and audio/video SCMC were also analysed. The multimodal analysis identified four levels of multimodal communicative competence, upon which a proposed MCC pyramid were based. Finally, the relationships among multiple modes in video SCMC were analysed, the latter being informed by specific multimodal episodes.

The next chapter will present a discussion of the study and its findings, where the two research questions will be answered based on the findings of the three types of analysis reported in

Chapters 5, 6, and 7. Relevant literature will be discussed in relation to the findings of the current study.

8 Discussion

The previous three chapters presented three different types of data analysis, including an interaction analysis of meaning negotiation in audio SCMC (Chapter 5), a statistical analysis of the role of gaze in 37 meaning negotiation episodes carried out in video SCMC (Chapter 6), and a detailed qualitative multimodal analysis of four selected meaning negotiation episodes (Chapter 7).

This chapter will use the findings from these three types of data analysis to answer and discuss the two research questions of the thesis.

8.1 How do students negotiate meaning in audio SCMC and video SCMC?

Section 8.1 will answer research question 1 ‘How do students negotiate meaning in audio SCMC and in video SCMC?’ in three parts. Sections 1 and 2 will specifically discuss the findings in relation to meaning negotiation routines in audio and video SCMC. Section 3 will summarise and discuss meaning negotiation routines across the different modes of communication identified in the existing literature and present a comprehensive framework of meaning negotiation routines to demonstrate how the findings of the current study fill a gap in our knowledge and contribute to the research field.

8.1.1 Meaning negotiation routines in audio SCMC

In this study, an initial interaction analysis used the model developed by Smith (2003), a framework derived from the study of meaning negotiation routines in text-based SCMC

environments, as a basis for identifying different stages in meaning negotiation episodes in audio SCMC. Meanwhile, considering that different modes of communication have significant influences on the ways in which information is communicated and languages are learned online (Stockwell, 2010; Hampel and Stickler, 2012), the researcher took into consideration any potential differences or new stages which might emerge from the data.

The interaction analysis of ten meaning negotiation episodes led to two key findings. First, meaning negotiation routines in audio SCMC generally conformed to the model by Smith (2003), especially as compared to the framework by Varonis and Gass (1985). This indicates that meaning negotiation routines in audio SCMC include not only the basic stages (including trigger, indicator, response and reply to response) but also the confirmation and re-confirmation stages. The last two stages were identified by Smith (2003) as the ways in which learners achieved greater explicitness in text-based computer-mediated communication where they could neither hear nor see their peers. The fact that learners in the current study frequently used confirmation and reconfirmation stages in audio SCMC interaction seems to suggest that even in voice-based computer-mediated communication, learners still had a strong need for clarity, as has been highlighted by Smith (2003).

In addition to the routine identified by Smith (2003), some new stages also emerged repeatedly from the data. Based on interaction analysis of all meaning negotiation episodes in audio SCMC in the current study, an expanded model of negotiation for meaning routines in oral SCMC interactions was proposed (Figure 15). In this new model, which should be read from the top downwards, two new stages, CT and CI (confirmation of trigger and confirmation of indicator), were added to the routine. The two stages were used by learners to clarify non-understanding before moving on to resolving it. The data analysis showed that these stages occurred in seven out of ten audio SCMC meaning negotiation episodes, which indicated a strong need for learners to identify the source and nature of non-understanding in audio SCMC interactions.

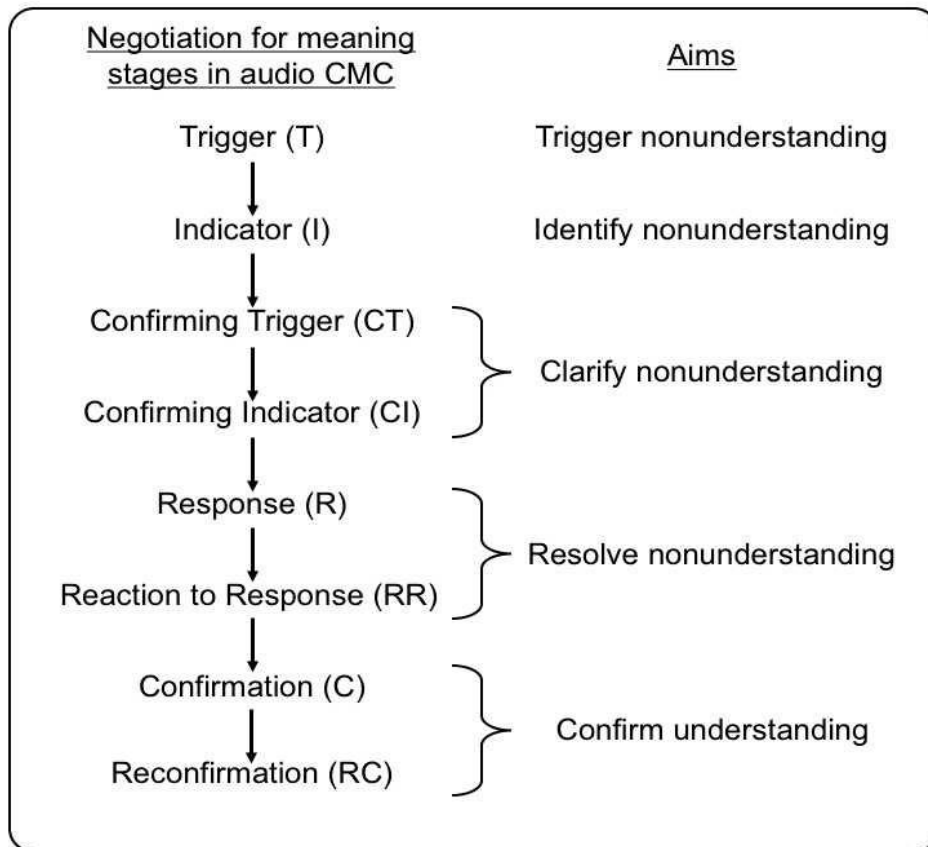


Figure 15: A negotiation for meaning routine for audio SCMC interactions: An expanded model

Therefore, it is important for researchers to explore why learners frequently experienced a failure of understanding during negotiated interactions in audio SCMC. A close look into the specific cases where these stages occur could help to reveal the fundamental reasons. Specifically, these two new stages were found at three possible points in meaning negotiation routines: (1) immediately after the indicator and before any resolution or explanation, (2) during the resolution stage (R or RR), and (3) after completing the resolution stage and following the confirmation request (Figure 16). These three possible routines were drawn from the meaning negotiation episodes identified in the data and are extensively discussed in Section 5.3.1.

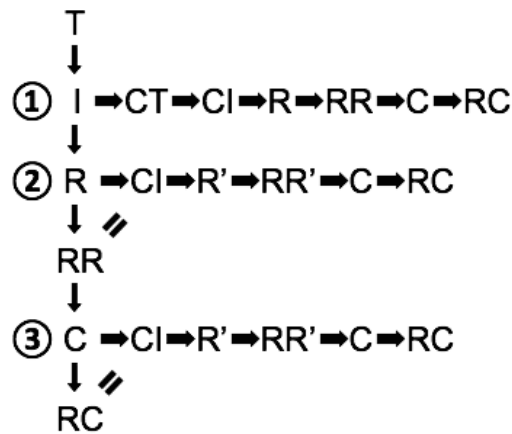


Figure 16: Three pathways of negotiation for meaning in audio SCMC interactions

The occurrence of three different pathways for meaning negotiation could be explained by the nature of the spoken medium, which makes possible different causes of non-understanding: (1) the meaning of the word, (2) its pronunciation, or (3) both. This differs from written interactions in text-based SCMC environments where pronunciation clearly cannot be a factor non-understanding.

The analysis of the oral interaction data reveals the causes of non-understanding at different linguistic levels were due to the specific technological affordances of the audio SCMC environment. First, as in the 'drawer/driver' episode (see Section 5.2.2.2), a speaker's failure to pronounce the word correctly or clearly could trigger non-understanding in the hearer. Second, oral communication differed from its written counterpart in that it required a higher level of attention and memory from the hearer. Once a phrase or a sentence was spoken, it was not preserved in any way, unlike in text-based SCMC where one could re-read past messages. So, if the hearer had poor listening skills or experienced a lapse in concentration, s/he could miss a word or an entire utterance. For example, in the 'stationery' episode (see Section 5.2.2.3), the hearer did not grasp the pronunciation of the word, although the initiator pronounced it many times. Moreover, their oral communication was mediated by synchronous audio-conferencing technology, so non-understanding could be caused by technical issues, which might also impede perception. Finally, how the respondent perceived the spoken sound of the trigger and

how she/he comprehended this information could both make a difference in meaning negotiation routines. For these reasons, when a respondent indicates non-understanding with a global indicator, the initiator may find it difficult to identify the level at which the cause of non-understanding is located. Therefore, it is often necessary for further confirmation of the trigger and the indicator to take place before resolving the non-understanding. In other words, in audio SCMC, it is hardly possible to move to resolving non-understanding unless both interlocutors have clearly identified what exactly has triggered it.

In summary, the researcher identified four layers of possible communication breakdown in audio SCMC interactions: (1) the initiator's expression or pronunciation, (2) the respondent's reception of and the attention to the spoken sound, (3) the respondent's perception of the spoken sound, and (4) respondent's comprehension of its meaning. All these factors, which were related to the audio channel of communication, could influence meaning negotiation routines differently from text-based written SCMC interactions. The proposed stages in meaning negotiation (CT and CI) could help interlocutors to understand the nature of the problem in this specific context. In this sense, synchronous audio communication might require an even higher level of explicitness and a more extended negotiation routine than text-based SCMC as claimed by Smith (2003).

Previous studies on audio SCMC interactions acknowledge the effects of pronunciation on meaning negotiation. For example, Jung and Jie (2012) identify pronunciation as a new type of trigger in video SCMC interactions between learners from different ethnic groups. Renner (2017) also report some 'clarification requests' where students tried to establish whether non-understanding had a phonological cause clarify the problem. Other researchers believe audio-based SCMC can promote pronunciation because learners can identify their own and their peers' pronunciation issues during negotiated interactions for communication breakdowns and thus, produce phonetically modified output (Jepson, 2005; Yamada, 2009; Bueno, 2011, 2013; Jung and Jie, 2012).

However, all these studies still use a framework developed for text-based SCMC to analyse meaning negotiation routines in audio interactions. This study filled in this research gap by proposing an extended meaning negotiation routine specifically for audio SCMC interactions. The CI and CT stages proposed in this study are devoted to identifying the sources of non-understanding, emanating either from their peers or from themselves. The three possible pathways specify the processes of how audio SCMC interactions might promote the development of pronunciation skills.

In conclusion, the discovery of two additional stages, CT and CI, reinforces previous findings on the effects of audio SCMC on pronunciation (Jung and Jie, 2012; Renner, 2017). Furthermore, the proposal of an expanded meaning negotiation routine and three possible pathways specifically for audio SCMC interactions could further contribute to the research field because it highlights the need for explicitness during such interactions and offers a more suitable framework for other researchers to analyse audio SCMC interactions.

8.1.2 Meaning negotiation routines in video SCMC-Directing the use of modes

8.1.2.1 Meaning negotiation stages in video SCMC interactions

Section 2 aims to answer research question 1 specifically in relation to meaning negotiation routines in video SCMC, but it also sheds light on research question 2, which asks 'What roles do multiple modes and semiotic resources play in meaning negotiation episodes in video SCMC?'. This section is based on the findings from the multimodal analysis in Chapter 7. The multimodal analysis followed a critical incident approach, in which four meaning negotiation episodes were selected for a fine-grained analysis of videoconferencing interactions, encompassing oral communication, gaze, facial expression and gesture. As has been widely acknowledged in the existing literature, all available modes and semiotic resources could contribute to meaning-making in video-based multimodal computer-mediated communication

(e.g. Kress and Van Leeuwen, 1996, 2001; Norris, 2004; Nelson, 2006; Royce, 2006; Dooly and Hauck, 2012). Therefore, the analysis of any multimodal interactions had to take a holistic approach, including scrutinising in combination all the semiotic resources and modes within the environment, rather than analysing any one particular mode independently. Owing to the extremely high level of detail involved in multimodal analysis and to the strict limits on registration periods for Ph.D. research in the UK, the researcher was able to analyse only a limited number of meaning negotiation episodes in video SCMC interactions. This dictated recourse to a critical incident approach. What the critical incidents have in common is that mode momentarily takes precedence over meaning. The episodes involved either participants' explicit mention of the use of modes or a conspicuous failure to use multiple modes in video SMC interactions. The multimodal analysis of these four episodes generated significant findings of meaning negotiation routines in video SCMC interactions.

The analysis reveals that the meaning negotiation routines in video SCMC interactions generally followed the model proposed by Varonis and Gass (1985). All the stages, including trigger, indicator, response and reply to response, were found in them. Two additional stages proposed by Smith (2003) and two new stages identified in the current study (CT and CI) were also found. Apart from the negotiation of meaning, a new type of negotiation was also identified in video SCMC interactions: directing the use of modes and semiotic resources. The following sections (Section 2.1.1-2.1.4) will demonstrate how these findings from the current study relate to the existing literature and contribute to the research field.

8.1.2.1.1 Trigger

Previous research on meaning negotiation in video SCMC seems to have reached a consensus that the lexical trigger is the most frequent source of non-understanding (Lee, 2006; Yanguas, 2010; Jung and Jie, 2012; Wang and Tian, 2013). But these studies do not, on the whole, specify whether the actual cause of non-understanding is semantic or phonological, or both.

Only Lee (2006) and Jung and Jie (2012) identify phonological triggers as a new type of trigger in video SCMC negotiated interactions.

In the current study, out of the four episodes subjected to multimodal analysis, three were initiated by lexical triggers, including 'salt shaker' by Dyad 4 (see Section 7.2.1), 'razor' by Dyad 1 (see Section 7.2.2) and 'magnifying glass' (see Section 7.2.3). However, the non-understanding of Dyad 4 in the 'nail polish' episode was caused by both pronunciation and meaning (see Section 7.2.4). Therefore, this study contributes in three ways to our understanding of trigger types in meaning negotiation episodes in video SCMC. First, it confirms that lexical triggers (here only referring to the meaning of the lexical item) are a frequent cause of non-understanding. Second, it confirms that the source of non-understanding can be caused by two aspects of a lexical item: either by its meaning or by its pronunciation, and occasionally by both. Finally, it highlights the distinction between meaning and pronunciation in video SCMC negotiation interactions, and thus encourages future researchers to be more specific in identifying not merely the type of trigger but the actual cause of misunderstanding.

According to Yanguas (2010), the key reason for the frequent occurrence of lexical triggers lies in the nature of the task design. In the current study, the spot-the-difference task had many small items in the pictures which needed to be described by interlocutors. The problem-solving tasks had fixed lexical items provided to students for meaning negotiation. In all 37 meaning negotiation episodes across all dyads and all tasks, only one involved a lexical trigger that was not intentionally seeded in the task design. This result largely confirms that task design plays an essential role in the frequent occurrence of lexical triggers. However, other types of tasks, particularly more open-ended ones such as opinion-gap tasks, might reveal different findings (see Section 2.3.3.4.4 in for a discussion of the effects of tasks on meaning negotiation). The relationship between task design and meaning negotiation clearly merits further research.

8.1.2.1.2 Confirmation of trigger (CT) and clarification of indicator (CI): new meaning negotiation stages identified in video SCMC

In this study, triggers were either followed by a local indicator, as in the 'magnifying glass' example (see Section 7.2.3), or by an extended conversation to confirm the trigger and clarify the indicator, as in the 'nail polish' episode (see Section 7.2.4). This means that the two new stages of meaning negotiation, confirmation of trigger (CT) and clarification of indicator (CI), identified in the current study as occurring in audio SCMC interactions, were also present in video SCMC interactions. For instance, in Dyad 2's 'nail polish' negotiation episode, the distortion in the audio channel disrupted the meaning negotiation routine, which was why D2A had to confirm with her peer the exact pronunciation of the trigger. And whereupon the initiator (D2B) attempted to clarify the indicator (CI) by stressing the word 'polish' and asking the respondent '*do you know that?*'. This example demonstrates that in negotiated interactions in video SCMC, pronunciation issues and technical distortion could both trigger non-understanding for the participants, just as in audio SCMC. To resolve non-understanding, interlocutors confirmed both trigger and indicator before negotiating the meaning of the unknown word.

Therefore, the meaning negotiation routine could be expanded to involve the two new stages, CT and CI identified from video SCMC interactions in the current study. These two new stages are essentially related to the technological affordances of the audio channel in video SCMC. The finding is similar to the expanded model in audio SCMC (see Section 8.1.1), indicating the shared voiced-based nature of communication in both audio and video SCMC. In other words, it highlights the similarity in meaning negotiation routines in both audio and video SCMC that are distinctive to text-based SCMC (Smith, 2003) or face-to-face non-mediated communication (Varonis and Gass, 1985). Moreover, this finding also fills in the research gap of the lack of a meaning negotiation framework specifically developed for video SCMC, which has been illustrated by those researchers who use the framework from text-based SCMC (Smith, 2003) or face-to-face communication (Varonis and Gass, 1985) to analyse meaning negotiation routines in video SCMC (e.g. Wang, 2006; Lee, 2006; Yanguas, 2010; Wang and Tian, 2013; Van der Zwaard and Bannink, 2014). More details about this research gap could be found in

Section 2.3 in the literature review. Apart from these two new stages in meaning negotiation episodes, another distinctive stage in video SCMC will be discussed later in Section 2.2.

8.1.2.1.3 Response and reaction to response

As for the response and reaction to response stage in meaning negotiation in video SCMC, researchers have identified different levels of use of multiple modes and semiotic resources by interlocutors (Yanguas, 2010; Wang and Tian, 2013) and pointed out the importance of visual cues in these negotiations (Lee, 2006; Wang, 2006; Yanguas, 2010; Wang and Tian, 2013). There is, however, some disagreement regarding the role of the video channel in negotiated interactions in video SCMC. Van der Zwaard and Bannink (2014, 2016), for example, argue that the video channel has a negative impact on meaning negotiation, pointing out that participants may pretend to understand the meaning during the reaction to response stage in order to save their face and avoid negotiating meaning. These disagreements are caused by the limitations in their research methods as they are only based on the interaction analysis of speech-based meaning negotiation routines using written descriptions of the visual cues. None use multimodal analysis, nor are multimodal transcriptions and screenshots provided to demonstrate the role of visual cues in detail. The present study used the multimodal analysis to demonstrate specifically how multiple modes were used in resolving non-understanding. In addition to analysing spoken interaction, the multimodal analysis also scrutinised further details, including frame-by-frame screenshots of students' gestures and facial expressions. These details add new empirical multimodal evidence of the important role of visual cues in negotiated interactions in video SCMC.

In particular, multimodal analysis reveals that iconic gestures, defined as 'representations of an action or object' by McNeill (1992), were frequently used during the resolution stage (including response and reaction to response). On the one hand, iconic gestures were used in the response stage by the initiator to explain the meaning of the trigger. For example, in the 'nail polish' episode, the initiator showed her hand right in front of the webcam to demonstrate the

meaning of the word 'nail'. On the other hand, iconic gestures were also used by the respondent in the reaction to response stage to test deduction and guess the meaning of the unknown word. For instance, when attempting to guess the meaning of the expression 'magnifying glass', the respondent placed her hands around her eyes in the shape of a pair of glasses.

Furthermore, comparisons across different dyads' multimodal performances during the response and reaction to response stages illustrate that multiple modes and semiotic resource were used to varying degrees by different interlocutors during their negotiation process, as proposed by Yanguas (2010) and Wang and Tian (2013). For example, Dyad 2 and Dyad 3 made extensive use of gestures and facial expressions when negotiating meaning for the word 'nail polish' and 'magnifying glass' respectively. However, Dyad 1 and Dyad 4 mostly focused on their task sheet and made little use of the visual channel to negotiate meaning in video SCMC interactions.

The multimodal analysis of the 'razor' episode by Dyad 1 in this study also identified a 'fake understanding' during the reaction to response stage in a video MNE (see Section 7.2.2). Specifically, D1B used nods, smiles together with oral confirmation to feign understanding of the trigger to her peer. The multimodal analysis also shows that she was trying to note down the word 'razor' in her own intuitive spelling. The stimulated recall interview confirms that she was unwilling to interrupt her peer and hoped to figure out the meaning of the word by herself in the later conversation or to look up the word in the online dictionary. Meanwhile, D1B also expressed her preference for video as opposed to audio SCMC. The multimodal analysis also shows her competence in using different visual cues to achieve her own communicative purpose, in this case, is to avoid meaning negotiation and look the word up in the online dictionary. In other words, D1B was not afraid of using video SCMC. Instead, she held negative attitudes towards negotiation for meaning, irrespective of the form of communication. This is different from Van der Zwaard and Bannink (2014, 2016) who report that NNS participants considered the videoconferencing part of the task 'scarier' than the text-chat part of the task (2016, p.636). The combination of multimodal analysis and the stimulated recall interview made

it possible for the researcher to distinguish the two potential causes and point out the exact reason why this student simulated understanding in video SCMC. In addition, this study identified much more cases of meaning negotiation episodes (37) in video SCMC than in audio SCMC (10), which also demonstrated that most participants did not consider the face issue as a big problem in their MNEs in video SCMC as has been reported by Van der Zwaard and Bannink (2014, 2016).

8.1.2.1.4 Confirmation and re-confirmation

The confirmation and reconfirmation stages received little attention from researchers. Only a few (Lee, 2006; Wang, 2006; Wang and Tian, 2013) mention that even in video SCMC interactions, interlocutors use the confirmation and reconfirmation stages to achieve explicitness during the negotiation.

In the current study, the confirmation and reconfirmation stages proposed by Smith (2003) were observed in some of the video SCMC interactions and proved to be highly important to the success of meaning negotiation in video SCMC. For example, Dyad 2 used the confirmation and reconfirmation stages to check whether the respondent had arrived at a correct understanding. At this stage, the initiator identified continued misunderstanding by the respondent, so issued another confirmation request and finally received a correct explanation of the expression 'nail polish' from the respondent. This final reconfirmation marked the successful conclusion of the meaning negotiation episode. This finding further confirms that when communication is mediated through technology, interlocutors seek to express themselves as explicitly as possible, whether in text-based SCMC or in audio and video SCMC.

Moreover, many visual cues, including facial expressions, head shakes, nodding, hand gestures, and smiles, were used together with speech to express denial, confirmation and reconfirmation in video SCMC. As was acknowledged in their stimulated recall interviews by both interlocutors in Dyad 2, visual cues gave them important indications of each other's

(non-)understanding. Therefore, this study argues that confirmation and reconfirmation stages play an important role in the success of meaning negotiation, and that the extensive and appropriate use of visual cues during these stages could be very helpful to the successful outcome. It is hoped that future researchers will choose to study the use of visual cues in these stages, particularly when a misunderstanding occurs.

Section 2.1 discussed specific meaning negotiated stages in video SCMC, including trigger, indicator, confirmation of trigger and clarification of indicator, response and reaction to response, confirmation and reconfirmation. These stages are the ones identified in audio SCMC negotiated interactions (see Figure 15 in Section 8.1.1). The next section will discuss the identification of a new phenomenon particularly in the video SCMC: directing the use of modes.

8.1.2.2 Directing the use of modes during negotiation of meaning

In early research about multimodal interactions in SCMC, the notion of 'mode switching' is proposed and highlighted by Sindoni (2013, 2014) who explores how learners switch between oral and written modes in multiparty video SCMC interactions (see Section 2.4.2 for more information). But 'mode switching' does not seem to be the most accurate term to describe the examples from the current study. The word 'switch' seems to indicate that students can only make use of one mode at a time. However, the 'nail polish' episode and the 'salt shaker' episode from the current study (see Section 7.2.1 and 7.2.4) show that interlocutors do not just switch modes. In both episodes, interlocutors used the audio channel to direct their peers to focus on the video images or the task sheet. As such, there was no 'switching' in this process, and the term 'directing the use of modes and semiotic resources' is preferred. This study proposes to define this notion as the process where interlocutors explicitly direct the use of proper multiple modes or semiotic resources in multimodal SCMC environments for better communication. This notion covers a wide range of possible combinations of modes and semiotic resources available in the SCMC environment rather than indicating a competing relationship between only two particular modes in the multimodal SCMC environment.

Previous studies have identified interlocutors' flexible or strategic use of multiple modes and semiotic resources in audio and video SCMC negotiated interactions (Wang and Tian, 2013; Guo and Möllering, 2016; Renner, 2017). For example, Wang and Tian (2013) report an episode where one interlocutor in video SCMC reminded her peer of the availability of the textual medium by saying 'you can type'. Similarly, Guo and Möllering (2016) report that in audio SCMC, one participant told her peer, 'hold on, I will pinyin it for you', which means typing the pronunciation code of a Chinese word through text-chat. Moreover, Renner (2017) finds that in 12% of all meaning negotiation episodes in audio SCMC interactions, students switched from the audio channel to text-chat. These examples relate to using text-chat to replace oral communication, but little is known so far about how participants make flexible use of the video channel for meaning negotiation in video SCMC. These examples are cases of 'directing the use of modes and semiotic resources' as defined in the current study, which – to the best of our knowledge - is the first to explicitly propose and discuss the notion.

Using the multimodal analysis, the current study was able to unveil how 'directing the use of modes' took place during meaning negotiation routines in video SCMC. For example, in Section 7.2.4, D2A twice asked D2B '*can you see me?*' to direct the use of the video channel during their meaning negotiation of the word 'nail polish'. By way of contrast, when Dyad 4 were engaged in the spot-the-difference task (see Section 7.2.1), both interlocutors repeatedly stressed the words 'in my picture' or 'in your picture', in an evident attempt to remind each other to focus on the task sheet on the table in front of them, instead of looking at the screen. This also is an example of directing the use of modes, although to a rather different end. In video SCMC interactions, multiple modes and semiotic resources are available for students to use. Explicitly directing the use of modes leads both interlocutors to focus on particular mode(s) so that they could negotiate meaning effectively. An in-depth discussion of the role of multimodal interactions in negotiation in video SCMC will be presented later in the chapter (see Section 8.2).

Introducing the concept of 'directing the use of modes and semiotic resources' contributes to the research field in a number of ways. First, the multimodal analysis combined with the stimulated recall interviews offers strong evidence of students' awareness of the affordances of different modes and semiotic resources within the multimodal SCMC environment and demonstrates some students' competence in making strategic use of multiple modes and semiotic resources. Second, this study offers new insights into how interlocutors negotiate the use of the video channel, which have not been reported before. Third, as compared to 'mode switching' (Sindoni, 2013, 2014), which implies a competing relationship between only the oral and the written modes, the notion 'directing the use of modes' extends the range of modes to include all modes and semiotic resources available in the SCMC environment and affords different potential relationships between multiple modes. For instance, the 'nail polish' example illustrates a complementary relationship between the audio and the video channel (see Section 7.3.2.2). Furthermore, this notion offers a new perspective for future researchers in this field to further examine meaning negotiation routines and the role of multimodal interactions in video SCMC.

8.1.3 Meaning negotiation routines in different modalities

Sections 1 and 2 identified meaning negotiation routines that occurred in audio and video SCMC interactions in the current study. These findings have remedied the lack of meaning negotiation routine models specifically suitable for audio and video SCMC interactions. So far, meaning negotiation routines for different communication contexts have been identified, including face-to-face communication, text-based SCMC interactions and multimodal audio and video SCMC interactions. Table 26 summarises meaning negotiation routines in all of these communication contexts.

Type of communication	Face-to-face SCMC	Text-based SCMC	Audio SCMC	Video SCMC
Author(s)	Varonis and Gass (1985)	Smith (2003)	Li (the current thesis)	Li (the current thesis)

Meaning negotiation routine	Trigger Indicator Response Reply to response	Trigger Indicator Response Reply to response Confirmation Reconfirmation	Trigger Indicator (Confirming trigger) (Clarifying indicator) Response Reply to response (Confirmation) (Reconfirmation) 'Directing the use of modes' may occur at any stage before resolution.	Trigger Indicator (Confirming trigger) (Clarifying indicator) Response Reply to response (Confirmation) (Reconfirmation) 'Directing the use of modes' may occur at any stage before resolution.
Feature	Non-mediated oral and visual communication	Mediated textual communication with mixed features of oral communication and written communication, meaning and form can trigger non-understanding	Mediated oral communication, pronunciation and meaning can trigger non-understanding	Mediated oral and visual communication with multiple modes and semiotic resources available. Both meaning and pronunciation can trigger non-understanding

Table 26: A comprehensive summary of meaning negotiation routines in different communication environments

Table 26 demonstrates how different modes of communication affect meaning negotiation routines. In face-to-face communication, both interlocutors share the same physical space where they can talk to each other and see each other clearly without mediation. Therefore, the meaning negotiation routines in face-to-face communication follow the most basic stages, including the trigger, the indicator, the response and the reaction to response (Varonis and Gass, 1985). In the other three contexts, communication is mediated by different online environments, the affordances of which play an important role in shaping meaning negotiation routines. For example, in text-based SCMC, Smith (2003) observes that learners often use confirmation and reconfirmation to make sure that the respondent understood a previously unknown word correctly. In written SCMC, the lack of visual cues obliges interlocutors to express themselves more explicitly. Meanwhile, since the chat history can be seen by

participants, the trigger and indicator stages are clear and easy to understand. However, in audio SCMC interactions, where interlocutors use the audio channel as the default means of communication, confusions can occur when the initiator fails to pronounce a word clearly, the audio signal is distorted or interrupted, or the respondent does not hear the word clearly. In these cases, interlocutors need to confirm the trigger and clarify the indicator before moving to resolution and confirmation. Therefore, the current study proposes to add two more additional stages - confirmation of trigger and clarification of indicator - to meaning negotiation routine model in audio SCMC interactions. Finally, in video SCMC interactions, where multiple modes and semiotic resources are available to interlocutors, the latter sometimes need to negotiate the use of mode with their peers during meaning negotiation episodes. This is why this study proposes to add 'directing the use of modes' to the meaning negotiation routines in multimodal video SCMC interactions.

Section 8.1 answers the first research question 'How do students negotiate meaning in audio SCMC and in video SCMC?'. In summary, this study finds that in audio SCMC, meaning negotiation routines follow that of Smith (2003) but with two additional stages, confirmation of trigger (CT) and clarification of indicator (CI), mainly due to potential problems caused by the audio channel. Therefore, an expanded meaning negotiation routine is proposed for audio SCMC interactions. In video SCMC, visual cues are used at different stages to negotiate meaning and a distinctive feature, directing the use of modes is found. With these new findings, meaning negotiation routines in all four communication contexts including face-to-face, text-based, audio and video SCMC, are fully elaborated.

However, some other factors relating to negotiation for meaning routines have not been explored. For example, in this study, all the participants, online teachers, as well as the researcher, are Chinese nationals and share a similar cultural background. Therefore, it is possible that such cultural background has certain influence on how participants negotiate meaning in audio and video SCMC that the researcher was not fully aware of. The central focus of this study is on negotiation for meaning routines in different communication contexts, but it

should be acknowledged that the cultural factor might have affected the findings of the study. After all, this study has identified 10 and 37 MNEs in audio and video SCMC respectively, and their meaning negotiation patterns are highly similar to the ones identified by Varonis and Gass (1985) and Smith (2003). This finding disagrees with Littlewood (2007) who claims that Chinese students appear to be goal-oriented and do not value spending time discussing the language and negotiating meaning. Within the current study, the eight participants have shown different attitudes towards negotiating meaning and demonstrated varying competences in their linguistic skills. Therefore, further research is needed to specifically examine whether the expanded meaning negotiation stages identified in this study are caused by cultural factors and to verify whether this comprehensive meaning negotiation model proposed above still holds with participants from a variety of social and cultural backgrounds.

8.2 What roles do multiple modes and semiotic resources play in meaning negotiation episodes in video SCMC?

The key difference identified between meaning negotiation routines in audio and video SCMC lies in the 'directing the use of multiple modes and semiotic resources' in multimodal video SCMC interactions. Therefore, Section 8.2 moves on to answering research question 2: 'What roles do multiple modes and semiotic resources play in meaning negotiation episodes in video SCMC?'. Section 8.1 offered some answers related to this question, particularly on the use of multiple modes and semiotic resources at different stages of meaning negotiation in Section 8.1.2.1.

Section 8.2 will answer research question 2 from the following perspectives: (a) the relationship between interlocutors' gaze directions and meaning negotiation outcomes in Section 1; (b) findings from the multimodal analysis, including the relationships between multiple modes and

channels in meaning negotiation episodes in Section 2.1 and the multimodal communicative competence pyramid in video SCMC in Section 2.2; and (c) the differences between video SCMC and face-to-face communication in Section 2.3.

8.2.1 The role of the video channel in meaning negotiation in video SCMC

Exploring gaze in video SCMC interactions could offer insights into what role the video channel plays in the technology-mediated communication environment. The key difference between the affordances of audio and video SCMC is the availability of the visual mode in video SCMC. Due to the lack of a shared physical communication environment and the limited visibility of bodily gestures and posture, gaze has become one of the most important sources of information in video systems interactions (Sindoni, 2014). Only a small number of researchers have explored the role of gaze in video SCMC interactions (e.g. Develotte, Guichon and Vincent, 2010; Wang and Tian, 2013; Lamy and Flewitt, 2011; Satar, 2013, see more details in Section 2.4.2.3). It seems that researchers interested in multimodal interactions in SCMC tend to agree that interlocutors, no matter whether they are learners or teachers, have different levels of competence in making use of the webcam, the video channel and the multimodal resources in video SCMC. However, most of these studies do not focus on meaning negotiation episodes, with the exception of Wang and Tian (2013). In addition, although some of these studies have identified different types of gaze or levels of webcam use through multimodal analysis, none of them has used statistical analysis to quantify the role of gaze in video SCMC.

This study contributes to the research field by quantifying the time spent by participants on different gaze directions and correlating this factor with the outcomes of meaning negotiation in video SCMC. The physical setting of the task interactions and rigorous coding procedures using a multimodal annotation tool are two key elements that make the statistical analysis possible.

Previous literature has emphasised the embodied nature of task interactions during video SCMC, which informs the design of the current study. As Chanier and Lamy (2017) argue, in video SCMC, meanings are constructed 'through learners' physical relationship to tools ..., through learners' engagement with still and moving images...' (p.431). In the current study, the hard copies of the task sheets were placed on learners' desks during all task interactions. This setting made it possible to clearly and easily distinguish whether a participant was looking at the task sheet or at the screen. Guichon and Wigham (2016) and Develotte et al. (2010) also highlight the physical elements of the communication context beyond 'the screen's edge' (Jones, 2004, p. 24). In this study, students' gaze at the task sheet demonstrates how the physical set-up of the wider communication environment beyond the screen could have a substantial influence on the way interlocutors negotiate meaning in video SCMC.

For statistical analysis to be effective, it was essential to ensure the accuracy of the time spent in each direction. Four types of gaze direction were identified including (1) gaze directed at peer's video image; (2) gaze directed at the task sheet; (3) gaze directed at other directions; and (4) unidentifiable gaze directions. Three methods were used to identify and confirm these gaze directions, including a pre-task gaze direction test, constant gaze at one direction and a video stimulated recall interview (see more details in Section 6.1.3). All 37 meaning negotiation episodes, adding up to 61 minute and 3.6 seconds, were coded for statistical analysis. A multimodal annotation tool 'ELAN' was used to code gaze directions frame by frame and calculate the specific amount of time in each direction. The length of each frame is 0.013 seconds, ensuring that the coding was highly fine-grained and the statistics are as accurate as possible.

The correlation and regression analysis generate a statistically significant result (Figure 17): the more time participants spent looking at their peers' video image during negotiated interactions, the more likely they were to be successful in negotiating the meaning of unfamiliar lexis. Students' answers from the stimulated recall interview also confirm this finding as most of them (all except for D4A) reveal that they gained more information about their peers' thoughts

and (non-)understanding by looking at her video image during meaning negotiation (see Section 7.2).

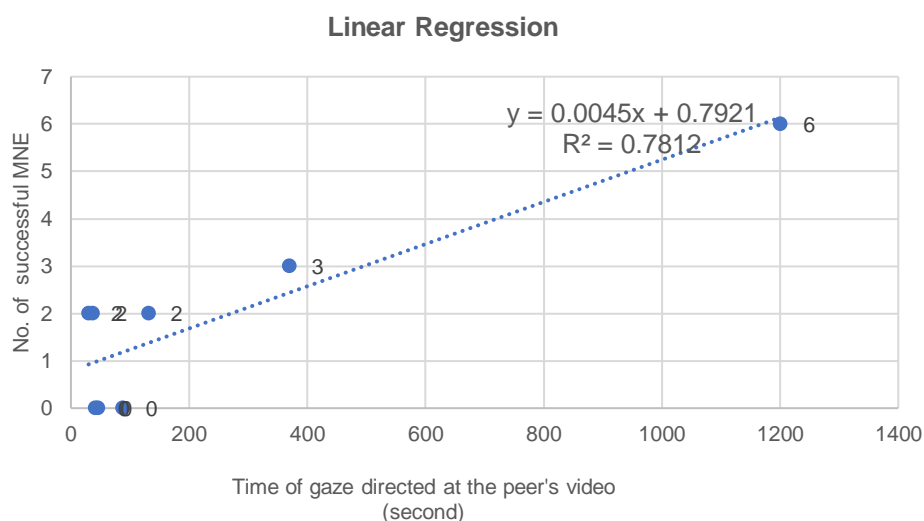


Figure 17: Linear regression between the time of gaze directed at peer's video and number of successful MNEs

This is possibly the first statistical analysis of students' gaze directions to have been carried out in relation to the success of meaning negotiation in video SCMC. On the one hand, this quantitative finding contributes to the research field by offering statistical evidence to prove that gaze direction plays an important role in meaning negotiation in video SCMC interactions. On the other hand, it seems to contradict some previous findings of how the use of video can distract students' attention or cause cognitive overload (Lee, 2006; Yanguas, 2010). A possible explanation of the different findings may be caused by students' different level of multimodal communicative competence, which will be further discussed in Section 8.2.2.1. The implication of the statistical result for online peer interactions is that interlocutors should look at each other's video image more frequently during meaning negotiation in video SCMC. But since this is a first attempt to establish the role of gaze direction in video SCMC, further investigation would be welcome to test the generalisability of this result with different tasks, participants and in different contexts.

8.2.2 The role of multimodal interactions in meaning negotiation in video SCMC

Section 1 established the importance of gaze direction using a statistical analysis, which partly answered the question of the role of multimodal interactions in meaning negotiation episodes in video SCMC. Section 2 will complete the answer by discussing the key findings from the multimodal analysis in Chapter 7.

The multimodal analysis focuses on critical incidents, which led to the identification of one meaning negotiation episode from each dyad. These four episodes were selected because the dyad had a clear discussion about the use of modes in their video SCMC interactions, or demonstrated a lack of awareness, or use of multiple modes and semiotic resources. These episodes were annotated on multiple levels to record speakers' use of modes such as speech, facial expressions, hand gestures, head movements and gaze, and to record their framing choices in front of the webcam. All these multimodal elements were analysed in a comprehensive way to demonstrate how learners used multiple modes to negotiate meaning in video SCMC. Two key findings emerged from the multimodal analysis: interlocutors possessed significantly different levels of competence in the use of modality, as is represented in the pyramid model, and the relationships among different modes and semiotic resources.

8.2.2.1 The multimodal communicative competence pyramid for meaning negotiation in video SCMC

It is widely agreed that interlocutors tend to have varying degrees of ability in using the webcam and multiple modes in video SCMC (e.g. Develotte et al., 2010; Wang and Tian, 2013). Mayer (2005) originates the notion of multimodal competence, which entails developing the metacognitive strategies necessary for 'allocating, monitoring, coordinating, and adjusting [...] limited cognitive resources' (p.36) when dealing with mediated learning situations. Hampel and Hauck (2006) point out that interlocutors need to develop the ability to use different modes

critically so as to 'familiarise themselves with the 'grammar' of other modes, such as the visual'. Similarly, as Kramersch (2006) argues, learners these days not only need to know how to communicate meanings, but also have to understand the process of meaning making itself. However, the notion from these studies is often theoretical or observational, lacking empirical evidence. Only Develotte et al. (2010) use a multimodal analysis and propose five degrees of use of webcam by online language teachers. So far, no specific multimodal communicative competences by students in video SCMC have been identified and demonstrated by a multimodal analysis.

The current study identified four levels of multimodal communicative competence by analysing four meaning negotiation episodes in video SCMC and comparing the use of multiple modes and semiotic resources in these episodes. The different levels appear to build on one another (see Section 7.3.1). Therefore, the study proposes a multimodal communicative competence pyramid for meaning negotiation in video SCMC (Figure 18).

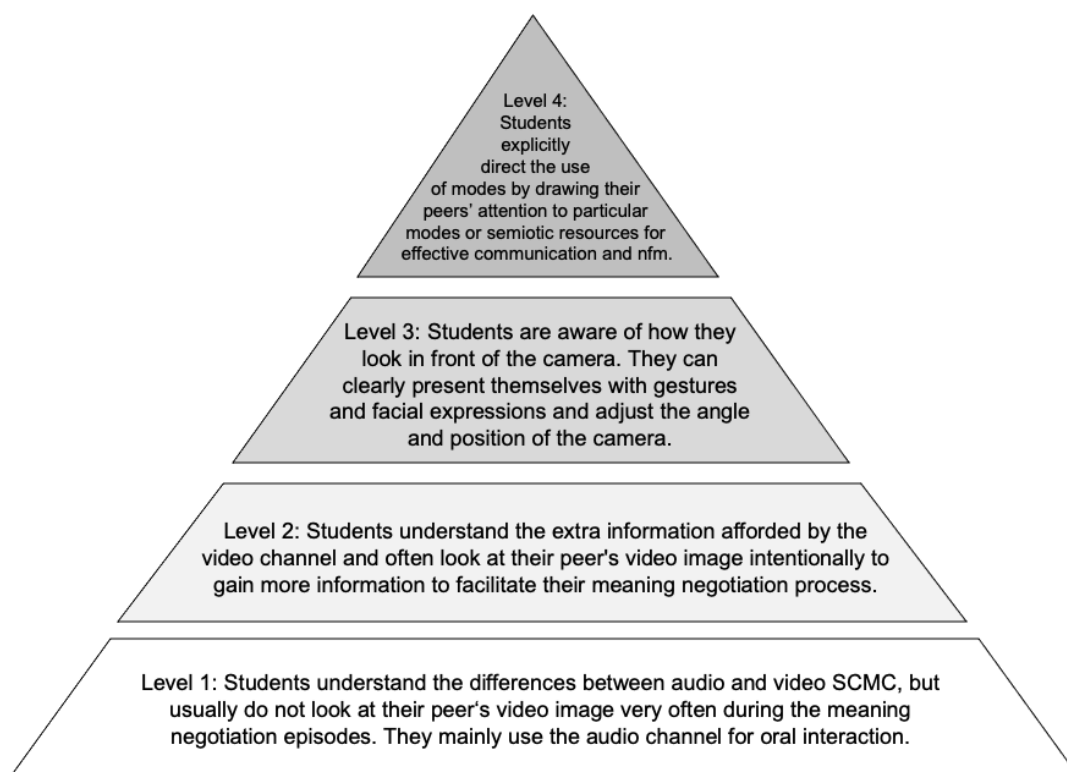


Figure 18: The multimodal communicative competence pyramid for meaning negotiation in video SCMC

This finding contributes to the research field in the following ways. First, the levels are based on empirical evidence of participants' multimodal performances during meaning negotiation in video SCMC. Therefore, the model was derived inductively from the data. Taking the form of a pyramid, the model specifies four different levels of multimodal communicative competence, which have not been captured before. These four levels have enriched the notion of multimodal communicative competence. Furthermore, although developed from meaning negotiation episodes, these competences can also be applied to a wider range of multimodal interactions, such as online teachers' instructions and students' collaborative activities in video SCMC. It is hoped that future studies can use this pyramid as a framework to examine students' multimodal communicative competence. It also has the potential to be further developed, expanded and revised in different research contexts.

The direct implication of this model is that online learners need sufficient training in multimodal communicative competence to be able to make good use of multiple modes and semiotic resources for communicating with their peers in video SCMC and to maximise their learning opportunities. As Hubbard (2013) argues, the assumption that students in modern days are naturally competent in using a wide range of technologies for online learning is not accurate. In fact, many learners need sufficient targeted training to develop 'the level of readiness needed for effective use of technology in language learning tasks and activities' (p.166). Through the comparison of the different levels of multimodal communicative competence of the four dyads, the study reinforces the necessity of training in the use of multiple modes for communicating in technology-mediated environments. The pyramid also potentially offers a practical guide for such training for both online teachers and students as it is inductively derived from students' multimodal performances in video SCMC interactions.

8.2.2.2. The relationships among different modes in video SCMC

As Jewitt (2015) points out, the central aim of multimodal studies is to explore the full range of communicative modes and the relationships that are created between them. The relationships between different modes are hierarchically ordered. They could reinforce or complement, or even compete against each other (Kress and Van Leeuwen 2010; Hampel and Stickler, 2012; Sindoni, 2013, 2014), rather than different modes simply being 'the alternative means of representing the same thing' (Kress and Van Leeuwen, 1996, p. 76). However, little empirical evidence in the form of multimodal analysis has been reported to confirm these hypotheses and to demonstrate the relationships between different modes in multimodal video SCMC. Section 8.1.2.2 in this chapter reported that although previous studies have mentioned the strategic use of different modes in online multimodal interactions, there is a particular lack of research exploring the relationships between visual cues and other modes in video SCMC.

Existing meaning negotiation studies in video SCMC highlight the importance of the visual semiotic resources in video SCMC (Lee, 2006; Wang, 2006; Yanguas, 2010; Jung and Jie,

2012; Wang and Tian, 2013; Guo and Möllering, 2016). For example, Wang (2006) highlights the importance of visual cues in negotiated interactions in video SCMC, arguing that facial expressions or head movements, such as nodding, laughing, or a puzzled look are examples of indicating or confirming (non-)understanding. Wang and Tian (2013) report examples of minimal reaction to response accompanied by visual cues. But these studies only emphasise the importance of the video channel itself and lack insightful analysis and discussion on the relationships among various types of visual semiotic resources and between visual and oral modes.

This study filled in the gap by using multimodal analysis to exemplify the relationships between multiple modes and semiotic resources (see Section 7.3.2). Three types of relationships were identified among different modes and semiotic resources in video SCMC. First, since the audio channel was used as the default channel for conversations, it is clear that a hierarchy exists between the audio channel and the visual channel. This hierarchy relationship is very similar to the case in face-to-face communication. The essential reason for this hierarchy is that interlocutors could choose where to look, but they had to listen and speak to each other both in face-to-face communication and in video SCMC.

Second, the video channel and the audio channel could also be used to complement or reinforce each other. The use of facial expressions, hand gestures and head movements in the video channel could strengthen the oral expression in the audio channel, as was theorised by Kress and Van Leeuwen (2010) and Hampel and Stickler (2012). However, no examples were found where gestures or facial expressions were used to completely replace oral communication. Unlike face-to-face communication, where interlocutors could see each other clearly and sometimes a gesture or a facial expression was sufficient to express some ideas, video SCMC is mediated and interlocutors tend to feel the need to communicate with each other explicitly rather than succinctly.

Third, a competitive relationship was found between a textual resource, the task sheet and the screen in the 'salt shaker' episode engaged in by Dyad 4 (see Section 7.2.1). The gaze analysis in Chapter 6 also clearly illustrates that students' gaze can only focus on one direction at a certain point of time, indicating the competitive relationship between the task sheet and the peer's video image on the screen and other directions. This episode offers another possibility of a competitive relationship, one that is largely different from the competitive relationship proposed by Sindoni (2013, 2014) between oral and written mode.

Interestingly, unlike Guo and Möllering (2016) or Renner (2017), who report the use of text-chat messages during meaning negotiation, in the four meaning negotiation episodes for the multimodal analysis, text-chat was never used by any interlocutor. Therefore, this study was unable to identify the relationship between text-chat and other channels. The lack of use of the written mode could be caused by the task instruction which stresses practising participants' oral English proficiency.

These relationships show students' ability to make use of audio and video channels and other multimodal semiotic resources (such as the task sheet) to negotiate meaning in video SCMC. In other words, the multiple modes and semiotic resources play an important role in enhancing students' mutual understanding during meaning negotiation episodes in multimodal computer-mediated environments.

8.2.3 The differences between video SCMC and face-to-face communication

As was made clear in the multimodal communicative competence pyramid, the foundation of all levels of multimodal competence is the understanding of the fundamental differences between face-to-face communication and video computer-mediated communication. Kern (2014) explains these differences by highlighting the mediating effect of SCMC, particularly in light of the limited size of the screen (see also Guichon and Wigham 2016 in Section 2.4.2.1.2).

Therefore, this section discusses the particular limitations of video SCMC as they relate to this study.

The mediation process distinguishes video SCMC from face-to-face communication. One important feature of video SCMC is the size of the screen image. Whereas communication in a shared physical environment provides interlocutors with visual images at actual size, the size of images in video SCMC is often highly limited. From the viewer's perspective, this severely restricts the amount of information that can be derived from the interlocutor's visual image and from the contextual physical environment.

The size of the video images on the computer screen usually depends on the interface of individual video conferencing systems, as well as the size of the screen of the device used for communication. For example, in this study, most participants used their personal laptop and the university's video conferencing system for their task interactions. In this interface, typically, students' video images were located at the top right-hand corner of the screen. For example, on a 13.3-inch laptop, with the video SCMC interface maximised on the screen, the size of the participants' onscreen image is a rectangle of 30mmx44mm, and the size of a participant's face only takes up around an area of 10mm*12mm (Figure 19). The average size of an average woman's face is about between 144mm*187mm ('human head', 2020). To make it easier to compare the size differences, this study uses the term 'video actual ratio' to refer to the size of the person or object that can be seen in a video interface. This is calculated as a percentage, arrived at by dividing the size of the on-screen image by the actual size of the person or object. In this study, the ratio is usually between 6.6% and 10%. Therefore, it is much harder to notice any movements, or minor facial expressions in video SCMC than in face-to-face communication, due to the limited size of the screen and the interface design.

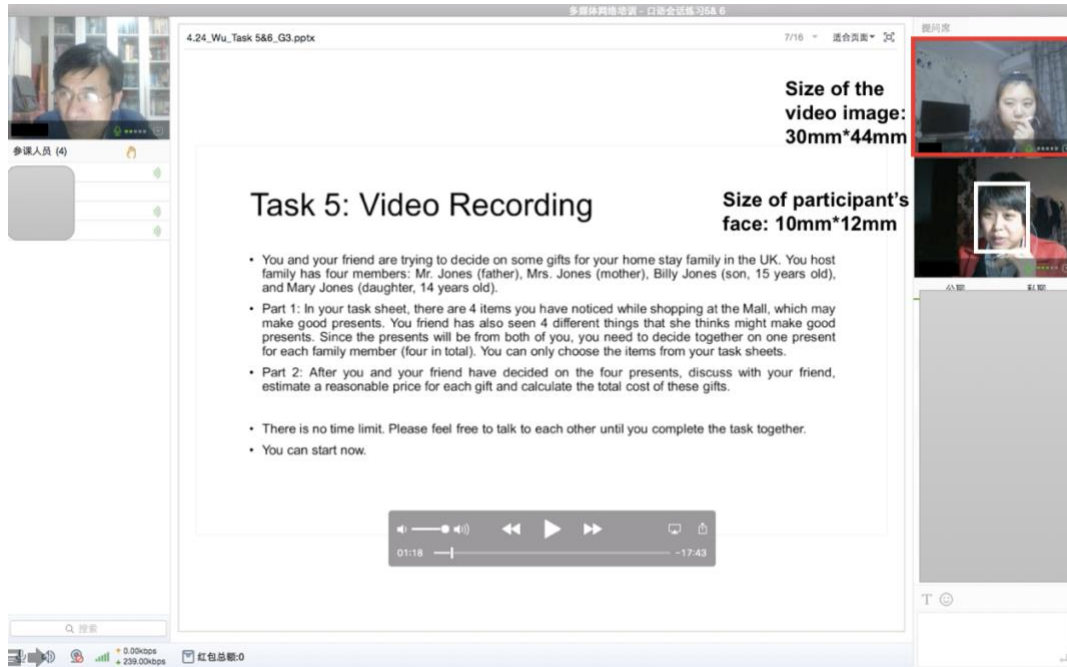


Figure 19: The size of interlocutors' face in the video SCMC interface

Another limitation of video SCMC, as compared to face-to-face communication, is that the webcam usually cannot capture all the gestures of participants (Kern, 2014; Guichon and Wigham, 2016), which could be important for meaning negotiation. As seen in Chapter 7, when D3B was explaining the meaning of a 'magnifying glass', she seemed to have made an 'enlarging' gesture at the bottom of the frame (see Section 7.2.3). However, it was not clear and long enough to be noticed by her peer, which is one of the reasons why Dyad 3 did not manage to negotiate the meaning of the word successfully. This technological limitation pushes interlocutors to develop their multimodal communicative competence and be aware of which of their gestures can be clearly projected through the webcam and which cannot, so that they can make proper adjustments.

Moreover, there is always a trade-off to be made between the visibility of the face and the visibility of gestures. Since the camera can capture only a limited amount of space, the closer one sits to the camera, the larger the image of the face, but the rest of the body will be out of the frame, and vice versa. Usually, students tend to sit close to the webcam, which allows them to see their peer's video image clearly and to show their face clearly, at the cost of missing all

the gestures below the shoulder. In the case of less competent communicators, the position and the angle of the webcam may not be properly set, and most of the video image is not the students' face but the background. For example, in Dyad 4 Task 3, the top one-third of D4B's video image captures the background, and her face occupies the lower two-thirds of the screen, while no movements or gestures below her neck are captured by the camera. Given that so little relevant visual information was on screen, it would have been harder for her partner to use visual cues to judge just how much D4B actually understood during meaning negotiation episodes.

In summary, Section 8.2 answered research question two, 'What roles do multiple modes and semiotic resources play in meaning negotiation episodes in video SCMC?' from three main perspectives. First, quantitative gaze analysis demonstrated a statistically significant correlation between the amount of time interlocutors spend looking at their peers' video image and the number of successful meaning negotiation episodes. This result showed the importance of multiple modes and semiotic resources and multimodal interaction, and especially of the skilful use of the video channel, in meaning negotiation episodes in video SCMC. However, not all interlocutors were able to make full use of different multimodal resources. The extent to which multiple modes and semiotic resources could enhance learners' understanding during negotiated interactions depends on their multimodal communicative competence. Four levels of multimodal communicative competences were identified based on the multimodal analysis. In addition, three types of relationships were identified between different modes in video SCMC, including a hierarchy order, a reinforcing or a competing relationship. Finally, the limitations of video SCMC and its effects on the role of multimodal interactions for meaning negotiation were discussed.

This chapter used the findings from the data analysis chapters to answer the two research questions identified following the literature review. The next chapter will contain a conclusion to the whole thesis.

9 Conclusion

9.1 Summary of the research findings

The study sets out to explore meaning negotiation episodes in computer-mediated communication environments, as was introduced in Chapter 1. Chapter 2 reviewed the relevant literature, covering (1) the historical development of this field, such as the interaction hypothesis and a model of meaning negotiation developed from face-to-face interaction studies; (2) recent meaning negotiation studies in synchronous audio and video SCMC interactions; and (3) studies focusing on interlocutors' multimodal interactions in synchronous video SCMC. The literature review identifies the following two major research gaps which the current study aims to fill:

- (1) How do students negotiate meaning in audio SCMC and video SCMC?
- (2) What roles do multiple modes and semiotic resources play in meaning negotiation episodes in video SCMC?

To answer the above questions, the following research methods (Chapter 3 and 4) were used to collect relevant data. The context in which the research took place was the online education department at a major Chinese university. The SCMC software was developed by the university and has all the basic features of video conferencing software, including presentation slides, an audio channel, a video channel and a text-chat block. Eight adult part-time undergraduate degree students were grouped into four dyads and asked to complete four tasks in two SCMC sessions. In each session, each dyad performed one task in audio and one task using video. The tasks were lexically seeded information gap tasks to elicit meaning negotiation by the participants. Students' task interactions during audio and video SCMC were screen video recorded for multimodal annotation, transcription and data analysis. The screen video recordings were also used as the basis for the video stimulated recall interviews, in which further data were gathered about students' thoughts and (non-)understandings at particular points of meaning negotiation episodes in audio and video SCMC.

Three types of data analysis were carried out, including: (1) an interaction analysis of all audio SCMC negotiated interactions; (2) a statistical analysis of students' gaze directions during meaning negotiation episodes in video SCMC; and (3) a multimodal analysis of students' verbal interactions, gaze directions, facial expressions, and gestures. The findings from these three types of analysis provided answers to the two research questions.

First, the interaction analysis (Chapter 5) of all the audio meaning negotiation episodes identified two new meaning negotiation stages: confirmation of trigger (CT) and clarification of indicator (CI). These two new stages were used by participants mainly because of the additional possibilities for non-understanding associated with speech-based online communication. An expanded meaning negotiation routine was proposed for audio SCMC interactions.

Next, a quantitative analysis of students' gaze directions (Chapter 6) established a statistically significant correlation between the amount of time students spend looking at their peer's video image and the number of successful negotiation episodes. The more students made use of the video channel in video SCMC interactions, the more likely they were to be successful. This is the first attempt to use quantitative methods to examine the role of the visual mode in negotiating the meaning of unknown lexical items, in video SCMC interactions. It establishes the importance of gaze direction in contributing to the success of such negotiations.

Finally, a number of findings could be derived from the multimodal analysis (Chapter 7). Firstly, in terms of meaning negotiation routines, interlocutors tended to direct the use of modes and semiotic resources during meaning negotiation episodes in video SCMC. Directing the use of modes and semiotic resources allows participants to focus intensively on the use of particular mode(s) in order to negotiate meaning effectively in an environment where multiple modes and semiotic resources are available. Secondly, based on students' different levels of multimodal performance in video SCMC interactions, a multimodal competence pyramid was proposed, which specifies four different degrees of making use of the multiple modes and semiotic resources available in the video SCMC environment. It is suggested that online learners would

benefit from training to become effective multimodal communicators in computer-mediated environments. Last but not least, the multimodal analysis identified specific examples of different relationships among different modes and semiotic resources in meaning negotiation episodes in multimodal SCMC interactions, including a hierarchy order, reinforcing/complementary and competing relationships. These examples contribute to answering the question of the role of multimodality in meaning negotiation episode in video SCMC.

9.2 Contribution to knowledge

Building on the meaning negotiation frameworks developed by Varonis and Gass (1985) as well as that of Smith (2003), this study contributes to the research field by identifying an expanded routine for both synchronous audio and video SCMC interactions. Therefore, meaning negotiation routines in four different contexts (including face-to-face, text-based SCMC, audio SCMC and video SCMC) are all specifically identified based on the affordances and limitations of each type of communication. In the field of multimodality research, the study adds to the existing literature by using quantitative evidence to illustrate the important role of gaze and the video channel in video SCMC interactions. The study also offers specific qualitative evidence to reveal a variety of relationships between different modes and channels in multimodal video SCMC interactions. Moreover, the multimodal communicative competence pyramid emerging from the study can also be helpful for developing such competence for video SCMC interactions, the need for which has been stressed by many researchers. The discussion chapter (Chapter 8) presents in more detail how this study fills in research gaps and contributes to knowledge.

9.3 Limitations of the study

Section 9.3 will acknowledge the limitations of the current study.

9.3.1 Limited episodes in the multimodal analysis

Owing to time constraints, only one episode from each participating dyad could be subjected to multimodal analysis. This is one of the limitations of this study. But the high level of detail in the multimodal analysis contributed to the validity of the eventual result. And gaze analysis, quantitatively measured, offers an overall understanding of students' performances in all MNEs in relation to where they looked. Moreover, video stimulated recall interview data was also helpful in corroborating the results of gaze analysis and multimodal analysis. However, it would of course have been satisfying to be able to analyse a larger selection of such episodes.

9.3.2 Task design

The task design has an important influence on meaning negotiation routines across different communication contexts (e.g. Pica et al., 1993; Smith, 2003; Yanguas, 2010). This study used two types of tasks, including a spot-the-difference task and a problem-solving task, both of which are closed information gap tasks and have been frequently used by researchers to elicit meaning negotiation (e.g. Smith, 2003; Yanguas, 2010; Guo and Möllering, 2016). It is possible that if other types of information gap tasks had been used, the research findings might have been different. In addition, the choice of lexical items in the task could also influence meaning negotiation routines and results. The lexical items selected denote concrete objects, and many of them could be easily explained in simple language or using gestures. However, if abstract lexical items had been chosen, it was not clear whether the meaning negotiation routines and results would be similar.

9.3.3 Contextual factors

The technical affordances and limitations of the particular video conferencing software used also play an important role in the meaning negotiation routines and students' multimodal performances studied here. For example, the size of the interlocutors' video image in the video SCMC interface was limited to a small corner of the screen which might limit the amount of visual information interlocutors can see. Other contextual factors, including students' level of proficiency in English, their attitudes towards meaning negotiation, the pairing of interlocutors, their shared Chinese cultural background as well as technical issues such as the internet speed, the clarity of the webcam, headset quality, might all have some influence on the findings of the current study.

9.4 Implications

Section 9.4 will offer some potential implications for research and teaching practice based on the findings of the current study.

9.4.1 Implications for future research

Therefore, it is hoped that more studies should be carried out to examine the findings of the current study. For researchers interested in meaning negotiation, it would be useful to test whether the expanded meaning negotiation routines framework for audio and video SCMC interactions proposed in the current study adequately describes meaning negotiation episodes in different audio and video SCMC contexts, with different task designs and target lexical items. In terms of multimodality, it would be helpful for future researchers to focus on the following aspects: (1) further exploring the role of the video channel in multimodal SCMC interactions using quantitative research methods as in the gaze analysis in the current study as it reveals the statistical relationships between the use of multimodality and language acquisition; (2)

identifying more examples of different relationships among multiple modes and semiotic resources to gain deep insights into the role of multimodality in the computer-mediated communication; and (3) verifying and further expanding the multimodal competence pyramid based on participants' multimodal performances in video SCMC interactions, which can be very helpful for the training of developing multimodal communicative competence.

9.4.2 Implications for online teaching practice

The findings of this study suggest that synchronous audio and video SCMC environment have the potential for students to develop their oral skills. It is recommended that language teachers could make use of audio and video SCMC as an alternative way for students to practice their oral English with their peers through well-designed tasks. Particularly due to the COVID-19 pandemic all over the world, video SCMC has become a very useful way to connect students and teachers who could not physically gather together in face-to-face classrooms. Therefore, it is more important than ever for both online learners and teachers to develop their multimodal communicative competence to make their online learning/teaching more effective. According to the gaze analysis (Chapter 6) and the multimodal analysis (Chapter 7), more training in multimodal communicative competence is needed. It is hoped that the pyramid proposed in Chapter 7 could prove useful as a practical guide for developing this. Particularly, interlocutors are advised to pay attention to their peers' video images and make full use of the video channel during video SCMC interactions. When feeling overwhelmed by multiple modes and semiotic resources, interlocutors should learn to direct the use of modes and semiotic resources with their peers in order to communicate smoothly in the multimodal SCMC environments.

References

Adams, R. (2007) 'Do second language learners benefit from interacting with each other?', in Mackey, R. (ed.) *Conversational interaction in second language acquisition*. Oxford: Oxford University Press, pp. 29–51.

Altorfer, A., Jossen, S., Würmle, O., Käsermann, M., Foppa, K. and Zimmermann, H. (2000) 'Measurement and meaning of head movements in everyday face-to-face communicative interaction', *Behavior Research Methods, Instruments and Computers*, 32(1), pp. 17-32.

Andersen, P. A. (1998) *Nonverbal communication: Forms and functions*. Mountain View, CA: Mayfield Publishing.

Andersen, P. A. (2008) *Nonverbal Communication: Forms and Functions*. Long Grove, IL: Waveland Press Inc.

Andresen, J. (2014) *Linguistics and Evolution: A Developmental Approach*. Cambridge: Cambridge University Press.

Argyle, M., Ingham, R., Alkema, F. and McCallin, M. (1973) 'The different functions of gaze', *Semiotica*, 7(1), pp. 19-32.

Aston, G. (1986) 'Trouble-shooting in interaction with learners: The more, the merrier?', *Applied Linguistics*, 7(2), pp. 128–143.

Bax, S. (2003) 'The end of CLT: A context approach to language teaching', *ELT Journal*, 57(3), pp. 278-287.

Bax, S. (2011) 'Normalisation revisited: The effective use of technology in language education', *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, 1(2), pp. 1-15.

Beattie, G. and Shovelton, H. (1999) 'Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation', *Semiotica*, 123(1-2), pp. 1-30.

Berger, P. and Luckmann, T. (1967) *The Social Construction of Reality*. London: Allen Lane.

Biesta, G. (2010) 'Pragmatism and the philosophical foundations of mixed methods research', *Sage handbook of mixed methods in social and behavioural research*, 2, pp. 95-118.

Birdwhistell, R. L. (1952) *Introduction to Kinesics: An Annotation System for Analysis of Body Motion and Gesture*. Washington, DC: Department of State, Foreign Service Institute.

Blake, R. J. (2005) 'Bimodal CMC: The glue of language learning at a distance', *CALICO Journal*, 22(3), pp. 497-511.

Bloom, B. S. (1953) 'Thought-processes in lectures and discussions', *The Journal of General Education*, 7(3), pp. 160-169.

Bloom, B. S. (1954) 'The thought processes of students in discussion classes', in French, S. J. (ed.) *Accent on teaching: Experiments in general education*. New York: Harper Brothers, pp. 23-46.

Boomer, D. S. and Allen T. D. (1964) 'Speech rate, filled pause and body movement in interviews', *Journal of Nervous and Mental Disease*, 139(4), pp. 324-327.

Boomer, D. S. and Dittman, A. P. (1964) 'Speech rate, filled pause, and body movement in interviews', *Journal of Nervous and Mental Disease*, 139, pp. 324-327.

Bourne, J., Franks, A., Hardcastle, J., Jewitt, C., Jones, K., Kress, G. and Reid, E. (2004). *English in urban classrooms: A multimodal perspective on teaching and learning*. New York: Routledge.

Braun, V. and Clark, V. (2006) 'Using thematic analysis in psychology', *Qualitative Research in Psychology*, 3(2), pp. 77-101.

Bremer, K., Broeder, P., Roberts, C., Simonot, M. and Vasseur, M. (1988) *Ways of achieving understanding: Communicating to learn in an L2*. Strasbourg, France: European Science Foundation.

Brock, C., Crookes, G., Day, R. and Long, M. (1986) 'The differential effects of corrective feedback in native speaker conversation', in Day, R. (ed.) *Talking to learn: Conversations in second language acquisition*. Rowley, MA: Newbury House, pp. 229-236.

Bryman, A. (2004) *Social Research Methods (2nd ed.)*. New York: Oxford University Press.

Bueno-Alastuey, M. C. (2011) 'Perceived benefits and drawbacks of synchronous voice-based computer-mediated communication in the foreign language classroom', *Computer Assisted Language Learning*, 24(5), pp. 419–432.

Bueno-Alastuey, M. C. (2013) 'Interactional feedback in synchronous voice-based computer-mediated communication: Effect of dyad', *System*, 41(3), pp. 543–559.

Chanier, T. and Lamy, M. N. (2017) 'Researching Technology-mediated Multimodal Interaction', in Chapelle, C. A. and Sauro S. (eds.) *The Handbook of Technology and Second Language Teaching and Learning*. Oxford: Wiley-Blackwell, pp. 428-443.

Chapelle, C. A. (1997) 'CALL in the year 2000: Still in search of research paradigms?', *Language Learning and Technology*, 1(1), pp. 19–43.

Chapelle, C. A. (2009) 'The Relationship between Second Language Acquisition theory and Computer-Assisted Language Learning', *The Modern Language Journal*, 93 (s1), pp. 741–753.

Chen, H. (2014). Foreign lexical acquisition in SCMC. *Foreign Language Education*, 35(3), 65-68.

Cherryholmes, C. H. (1992). Notes on pragmatism and scientific realism. *Educational Researcher*, 14, pp.13–17.

Chun, D. M. (1998) 'Using computer-assisted class discussion to facilitate the acquisition of interactive competence', in Swaffar, J., Romano, S., Markley, P. and Arens, K. (eds.), *Language learning online: Theory and practice in the ESL and L2 computer classroom*. Austin, TX: Labyrinth Publications, pp. 57-80.

Chun, D. M. (2003) 'Computer Applications in Second Language Acquisition: Foundations for Teaching, Testing and Research', *The Modern Language Journal*, 87(2), pp. 337-339.

Ciekanski, M. and Chanier, T. (2008) 'Developing online multimodal verbal communication to enhance the writing process in an audio-graphic conferencing environment', *ReCALL*, 20(2), pp. 162-182.

Codreanu, T. and Celik, C. C. (2013) 'Effects of webcams on multimodal interactive learning', *ReCALL*, 25(1), pp. 30-47.

Cohen, C. and Guichon, N. (2014) 'Researching nonverbal dimensions in synchronous videoconference-based interactions', *CALICO (Computer Assisted Language Instruction Consortium)*. University of Athens, Ohio, United States, May. Available at: <https://hal.archives-ouvertes.fr/hal-01079071> (Accessed: 20 January 2016).

Cohen, C. and Guichon, N. (2016) 'Analysing multimodal resources in pedagogical online exchanges', *Language-Learner Computer Interactions: Theory, methodology and CALL applications*, 2, pp. 187-213.

Cohen, C. and Wigham, C. R. (2019) 'A comparative study of lexical word search in an audioconferencing and a videoconferencing condition', *Computer Assisted Language Learning*, 32(4), pp. 448-481.

Cook, J. (2015) 'Negotiation for meaning and feedback among language learners', *Journal of Language Teaching and Research*, 6(2), pp. 250-257.

Cosnier, J. (2008) 'Les gestes du dialogue', in Cabin, P. and Dortier, J. F. (eds.) *La communication, e´tat des savoirs*. Auxerre: Editions Sciences Humaines, pp. 119–128.

Crawford, J., Butler-Henderson, K., Rudolph, J. and Glowatz, M. (2020) 'COVID-19: 20 Countries' Higher Education Intra-Period Digital Pedagogy Responses', *Journal of Applied Teaching and Learning (JALT)*, 3(1), DOI: 10.37074/jalt.2020.3.1.7.

Creswell, J. W. (2014) *Research Design: Qualitative, Quantitative and Mixed Methods Approaches*(4th ed.). Thousand Oaks, CA: Sage.

Crotty, M. (1998) *The foundations of social research: Meaning and perspective in the research process*. London: Sage.

Cunningham, J. D. and Akiyama, Y. (2018) 'Synthesizing the practice of SCMC-based telecollaboration: A scoping review', *Calico Journal*, 35(1), pp. 49-76.

Damon, W. and Phelps, E. (1989) 'Critical distinctions among three approaches to peer education', *International journal of educational research*, 13(1), pp. 9-19.

Davies, A. (2004). 17 The Native Speaker in Applied Linguistics. *The handbook of applied linguistics*, 431.

Davies, A. (2012). Who owns English? Questioning the native speaker. *Journal for Language Teaching*, 46(2), 191-201.

Deleuze, G. (1983) *L'image-mouvement*. Paris: Editions de Minuit.

Develotte, C., Guichon, N., and Kern, R. (2008) 'Allo Berkeley? Ici Lyon... Vous nous voyez bien?' Étude d'un dispositif de formation en ligne synchrone franco-américain à travers les discours de ses usagers. *Alsic. Apprentissage des Langues et Systèmes d'Information et de Communication*, 11(2).

Develotte, C., Guichon, N. and Vincent, C. (2010) 'The use of the webcam for teaching a foreign language in a desktop videoconferencing environment', *ReCALL*, 22(3), pp. 293-312.

Doerr, N. M. (Ed.). (2009). *The native speaker concept: Ethnographic investigations of native speaker effects* (Vol. 26). Walter de Gruyter.

Dooly, M. and Hauck, M. (2012) 'Researching multimodal communicative competence in video and audio telecollaborative encounters', in Dooly, M. and O'Dowd, R. (eds.) *Researching Online Interaction and Exchange in Foreign Language Education*. Telecollaboration in Education (3). Bern: Peter Lang Publishing Group, pp. 135-162.

Doughty, C. and Pica, T. (1986) 'Information gap tasks: Do they facilitate second language acquisition?', *TESOL quarterly*, 20(2), pp. 305-325.

Doughty, C. and Williams, J. (1998) *Focus on form in classroom second language acquisition*. New York: Cambridge University Press.

Ekman, P. and Friesen, W. V. (1969) 'The repertoire of nonverbal behaviour: Categories, origins, usage, and coding', *Semiotica*, 1, pp. 49-98.

ELAN Software. (2020) In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/ELAN_software.

Ellis, R. (2000) 'Task-based research and language pedagogy', *Language teaching research*, 4(3), pp. 193-220.

Ellis, R. (2003) *Task-based language learning and teaching*. Oxford: Oxford University Press.

Ellis, R. (2008) *The study of second language acquisition* (2nd ed.). Oxford: Oxford University Press.

Ellis, R. and Barkhuizen, G. (2005) *Analyzing learner language*. Oxford, England: Oxford University Press.

Ericsson, K. A. and Simon, H. A. (1996) *Protocol analysis: using verbal reports as data*. Cambridge, MA: Oxford University Press.

Faraco, M. and Kida, T. (2008) 'Gesture and the negotiation of meaning in a second language classroom', in McCafferty, S. and Stam, G. (eds.) *Gesture: Second language acquisition and classroom research*. New York: Routledge, pp. 280-297.

Feng, J., Chen, Y., & Shen, R. (2015). Cross-cultural SCMC negotiation of meaning and its role in lexical acquisition among tertiary Chinese students. *Technology Enhanced Foreign Language Education*, 2015(2), 10-16.

Ferguson, C. (1971) 'Absence of copula and the notion of simplicity: A study of normal speech, baby talk, foreigner talk and pidgins', in Hymes, D. (ed.), *Pidginization and creolization of languages*. Cambridge: Cambridge University Press, pp. 117–140.

Foster, P. (1998) 'A classroom perspective on the negotiation of meaning', *Applied linguistics*, 19(1), pp. 1-23.

Foster, P. and Ohta, A. S. (2005) 'Negotiation for meaning and peer assistance in second language classrooms', *Applied Linguistics*, 26(3), pp. 402-430.

Gass, S. M. and Mackey, A. (2017) *Stimulated Recall Methodology in Applied Linguistics and L2 Research (2nd edition)*. New York: Routledge.

Gerwing, J. and Allison, M. (2011) 'The flexible semantic integration of gestures and words: Comparing face-to-face and telephone dialogues', *Gesture*, 11(3), pp. 308-329.

Grayson, D. M. and Monk, A. F. (2003) 'Are you looking at me? Eye contact and desktop video conferencing', *ACM Transactions on Computer-Human Interaction*, 10(3), pp. 221–243.

Guba, E. (1990) *The Paradigm Dialog*. Newbury Park, CA: Sage.

Guichon, N., Blin., F., Wigham, C.R. and Thouësny, S. (2014) *ISMAEL LEarning and TEaching Corpus*. Dublin, Ireland: Centre for Translation and Textual Studies and Lyon, France: Laboratoire ICAR.

Guichon, N. and Cohen, C. (2014) 'The impact of the webcam on an online L2 interaction', *Canadian Modern Language Review*, 70(3), pp. 331-354.

Guichon, N. and Cohen, C. (2016) 'Multimodality and CALL', in Farr, F. and Murray, L. (eds.) *The Routledge Handbook of Language Learning and Technology*. Routledge, pp. 509-521. Available at: <https://www.routledgehandbooks.com/doi/10.4324/9781315657899.ch36> (Accessed: 25 February 2018).

Guichon, N. and McLornan, S. (2008) 'The effects of multimodality on L2 learners: Implications for CALL resource design', *System*, 36(1), pp. 85-93.

Guichon, N. and Wigham, C. R. (2016) 'A semiotic perspective on webconferencing-supported language teaching', *ReCALL*, 28(1), pp. 62-82.

Gumperz, J. J. (1982) *Discourse Strategies*. Cambridge: Cambridge University Press.

Gumperz, J. J. (2003) 'Interactional sociolinguistics: A personal perspective', in Schiffrin, D., Tannen, D., and Hamilton, H. E. (Eds), *The Handbook of Discourse Analysis*. Oxford: Blackwell, pp. 215–228.

Guo, S. and Möllering, M. (2016) 'The implementation of task-based teaching in an online Chinese class through web conferencing', *System*, 62, pp. 26-38.

Hall, Edward T. (1966) *The hidden dimension*. New York: Doubleday.

Hampel, R. (2003) 'Theoretical perspectives and new practices in audio-graphic conferencing for language learning', *ReCALL*, 15(1), pp. 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), pp. 105-121.

Hampel, R. and Hauck, M. (2006) 'Computer-mediated language learning: Making meaning in multimodal virtual learning spaces', *JALT CALL Journal*, 2(2), pp. 3-18.

Hampel, R and Stickler, U. (2005) 'New skills for new classrooms: Training tutors to teach languages online', *Computer Assisted Language Learning*, 18(4), pp. 311-326.

Hampel, R. and Stickler, U. (2012) 'The use of videoconferencing to support multimodal interaction in an online language classroom', *ReCALL*, 24(2), pp. 116-137.

Hatch, E. (1978a) 'Discourse analysis and second language acquisition', in Hatch, E. (ed.) *Second language acquisition: A book of readings*. Rowley, MA: Newbury House, pp. 401-435.

Hatch, E. (1987b) 'Acquisition of syntax in a second language', in Richards, J. (ed.), *Understanding second and foreign language learning: Issues and approaches*. Rowley, MA: Newbury House, pp. 34-70.

Heift, T., & Vyatkina, N. (2017). Technologies for teaching and learning L2 grammar. *The handbook of technology and second language teaching and learning*, 26-44.

Holliday, A. (2006). Native-speakerism. *ELT Journal*, 60(4), 385-387.

Huang, L. (2014) 'Video-stimulated verbal recall: A method for researching cognitive processes and strategic behaviours', *SAGE Research Methods Cases*. DOI: <https://dx.doi.org/10.4135/978144627305013508883> (Accessed: 10 July 2017).

Huang, R. H., Liu, D.J., Tlili, A., Yang, J. F. and Wang, H.H. (2020) *Handbook on Facilitating Flexible Learning During Educational Disruption: The Chinese Experience in Maintaining Undisrupted Learning in COVID-19 Outbreak*. Beijing: Smart Learning Institute of Beijing Normal University.

Hubbard, P. (2013) 'Making a case for learner training in technology-enhanced language learning environments', *Calico Journal*, 30(2), pp. 163-178.

Hubbard, P. and Levy, M. (2016) 'Theory in computer-assisted language learning research and practice' in Farr, F. and Murray, L. (eds.) *The Routledge handbook of language learning and technology*. Routledge, pp. 50-64. Available at: <https://www.routledgehandbooks.com/doi/10.4324/9781315657899.ch2> (Accessed: 25 February 2018).

Human head. (2020) In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Human_head.

iiMedia Research (2020) *Business Research Survey on the Online Education for Chinese Primary and Secondary School Students in the First Quarter of 2020*. Available at: https://www.sohu.com/a/384295245_533924 (Accessed: 20 May 2020).

Jefferson, G. (1984) 'Notes on some orderlinesses of overlap onset', in D'Urso, V. and Leonardi, P. (eds.), *Discourse analysis and natural rhetoric*. Padua, Italy: Cleup Editore, pp. 11-38.

Jepson, K. (2005) 'Conversations and negotiated interaction in text and voice chat rooms', *Language Learning and Technology*, 9(3), pp. 79-98.

Jewitt, C. (2009) *The Routledge handbook of multimodal analysis*. London: Routledge.

Jewitt, C. (2011) 'Different approaches to multimodality', in Jewitt, C. (ed.) *The Routledge handbook of multimodal analysis*. London: Routledge, pp. 28-39.

Jewitt, C. (2015) 'Multimodal analysis', in Georgakopoulou, A. and Splilioti, T. (eds.) *The Routledge handbook of language and digital communication*. London: Routledge, pp. 69-84.

Jewitt, C., Bezemer, J. J. and O'Halloran, K. L. (2016) *Introducing multimodality*. London: Routledge.

Jonassen, D., Davidson, M., Collins, M., Campbell, J. and Haag, B. B. (1995) 'Constructivism and computer-mediated communication in distance education', *American journal of distance education*, 9(2), pp. 7-26.

Jones, R. H. (2004) 'The problem of context in computer-mediated communication', in Levine, P. and Scollon, R. (eds.) *Discourse and technology – Multimodal discourse analysis*. Washington DC: Georgetown University Press, pp. 20-33.

Jones, R. H. (2012) *Discourse analysis: A resource book for students*. London: Routledge.

Jung, S. K. and Jie, J. (2012) 'Analysis of meaning negotiation patterns between non-native speakers in video conferencing', *English Teaching*, 67(2), pp. 247-269.

Kaneko, A. (2009) *Comparing computer-mediated communication (CMC) and face to face (FTF) communication for the development of Japanese as a foreign language*. PhD thesis. The University of Western Australia. Available at: https://research-repository.uwa.edu.au/files/3215844/Kaneko_Aki_2009.pdf (Accessed: 15 April 2016).

Kasper, G. and Kellerman, E. (1997) 'Introduction: Approaches to communication strategies', in Kasper, G. and Kellerman, E. (eds.) *Communication strategies: Psycholinguistic perspective*. New York: Longman, pp. 1-13.

Kelm, O. (1992) 'The use of synchronous computer networks in second language instruction: A preliminary report', *Foreign Language Annals*, 25, pp. 441-454.

Kendon, A. (1982) 'The study of gesture: some observations on its history', *Recherches Semiotique/Semiotic Inquiry*, 2(1), pp. 25-62.

Kendon, A. (2004) *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Kendon, A. (2014) 'Semiotic diversity in utterance production and the concept of 'language'', *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369(1651). DOI: 10.1098/rstb.2013.0293.

Kenning, M. M. (2010) 'Differences that make the difference: a study of functionalities in synchronous CMC', *ReCALL*, 22(1), pp. 3-19.

Kern, R. G. (1995) 'Restructuring classroom interaction with networked computers: Effects on quantity and characteristics of language production', *The Modern Language Journal*, 79(4), pp. 457-476.

Kern, R. G. (2014) 'Technology as Pharmakon: The promise and perils of the internet for foreign language education', *The Modern Language Journal*, 98(1), pp. 340-357.

Kim, N. Y. (2017) 'Effects of Types of Voice-Based Chat on EFL Students' Negotiation of Meaning According to Proficiency Levels', *English Teaching*, 72(1), pp. 159-181.

Kim, S. Y. (2006) 'Korean sixth-grade elementary school students' meaning negotiation in task-based CMC', *Multimedia-Assisted Language Learning*, 9(2), pp. 99-127.

Kramsch, C., (2006) 'From communicative competence to symbolic competence', *The modern language journal*, 90(2), pp. 249-252.

Krashen, S. (1977) 'Some issues relating to the monitor model', in Brown, H., Yorio, C. and Crymes, R. (eds.), *On TESOL '77*. Washington: TESOL, pp. 144-158.

Krashen, S. (1980) 'The input hypothesis', in Alatis, J. (ed.) *Current issues in bilingual education*. Washington: Georgetown University Press, pp. 168-180.

Kress, G. (2009) 'Social Semiotic Approach to Multimedia', *Advances in Language and Education*, p.15.

Kress, G. and Van Leeuwen, T. (1990) *Reading Images: Sociocultural aspects of language and education*. Geelong, Vic.: Deakin University.

Kress, G. and Van Leeuwen, T. (1996) *Reading images: The grammar of visual design*. London: Routledge.

Kress, G. and Van Leeuwen, T. (2001) *Multimodal discourse: The modes and media of contemporary communication*. London: Arnold.

Kress, G. and Van Leeuwen, T. (2010) *Multimodal discourse: The modes and media of contemporary communication* (Repr. d. publ., Hodder Education, 2001.). Bloomsbury Acad.

Lai, C., Fei, F. and Roots, R. (2008) 'The contingency of recasts and noticing', *CALICO Journal*, 26, pp. 70-90.

Lai, C. and Zhao, Y. (2006) 'Noticing and text-based chat', *Language Learning and Technology*, 10(3), pp. 102-120.

Lamy, M. and Flewitt, R. (2011) 'Describing online conversations: Insights from a multimodal approach' in Develotte, C., Kern, R. and Lamy, M. (eds.) *Décrire la conversation en ligne: Le face à face distancié*. Lyon: Presses Universitaires de Lyon.

Lee, H. G. (2006) 'Meaning negotiation in task-based video conferencing', *The Journal of the Korea Association of Multimedia-Assisted Language Learning*, 9(2), pp. 128-153.

Levine, P. and Scollon, R. (2004) *Discourse and Technology: Multimodal Discourse Analysis*. Washington: Georgetown University Press.

Lee, H., Hampel, R. and 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, pp. 26-38.

Li, C. C. and Lewis, T. (2018) 'Negotiation for Meaning Routines in Audio SCMC Interactions: An Expanded Framework', *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, 8(3), pp. 50-72.

Lincoln, Y. S. and Guba, E. G. (1985) *Naturalistic inquiry*. Thousand Oaks, CA: Sage.

Littlewood, W. (2007) 'Communicative and task-based language teaching in East Asian classrooms', *Language teaching*, 40(3), pp. 243-249.

Liu, S. H. J. (2017) 'Text-based negotiated interaction of NNS-NNS and NNS-NS dyads on Facebook', *ReCALL: The Journal of EUROCALL*, 29(3), pp. 294-312.

Long, M. (1980) *Input and interaction in second language acquisition*. PhD thesis. The University of California at Los Angeles.

Long, M. (1981) 'Input, interaction, and second language acquisition', in Winitz, H. (ed.) *Native language and Foreign language acquisition*. New York: Annals of the New York Academy of Science, Vol. 379, pp. 259-278.

Long, M. (1983) 'Native speaker/non-native speaker conversation in the second language classroom' in Clarke, M. and Handscombe, J. (eds.) *Pacific perspective on language learning and teaching*. Washington: TESOL, pp. 207-225.

Long, M. (1988) 'Understanding second language acquisition Rod Ellis. Oxford: Oxford University Press, 1985. Pp. 327', *Studies in Second Language Acquisition*, Cambridge University Press, 10(1), pp. 79–82.

Long, M. (1991) 'Focus on form: A design feature in language teaching methodology', *Foreign language research in cross-cultural perspective*, 2(1), pp. 39-52.

Long, M. (1996) 'The role of the linguistic environment in second language acquisition', in Ritchie, W. C. and Bhatia, T. K. (eds.) *Handbook of second language acquisition*. San Diego: Academic Press, pp. 413-468.

Long, M. H. and Crookes, G. (1992) 'Three approaches to task-based syllabus design', *TESOL quarterly*, 26(1), pp. 27-56.

Long, M. , and Robinson, P. (1998) 'Focus on form: Theory, research, and practice', in Doughty, C. and Williams, J. (eds.) *Focus on form in classroom second language acquisition*. New York: Cambridge University Press, pp. 15-41.

Lyle, J. (2003) 'Stimulated recall: A report on its use in naturalistic research', *British educational research journal*, 29(6), pp. 861-878.

Lyster, R. (2002) 'The importance of differentiating negotiation of form and meaning in classroom interaction', in Burmeister, P., Piske, T. and Rohde, A. (eds.) *An integrated view of language development: Papers in Honor of Henning Wode*. Trier, Germany: Wissenschaftlicher Verlag, pp. 381-397.

Lyster, R. (2001) 'Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms', *Language Learning*, 51(s1), pp. 265-301.

Mackey, A. (2007) 'Introduction: The role of conversational interaction in second language acquisition', in Mackey, A. (ed.) *Conversational interaction in second language acquisition: A collection of empirical studies*. Oxford: Oxford University Press, pp. 1–26.

Mackey, A., Abbuhl, R. and Gass, S. M. (2011) 'Interactionist approach', in Gass, S. M. and Mackey, A. (eds.) *The Routledge Handbook of Second Language Acquisition*. New York: Routledge, pp. 7-23.

Mackey, A. and Goo, J. (2007) 'Interaction research in SLA: A meta-analysis and research synthesis' in Mackey, A. (ed.) *Conversational interaction in second language acquisition: a series of empirical studies*. Oxford: Oxford University Press, pp. 407-453.

Mackey, A., Oliver, R. and Leeman, J. (2003) 'Interactional input and the incorporation of feedback: An exploration of NS-NNS and NNS-NNS adult and child dyads', *Language Learning*, 53, pp. 35-66.

Martinec, R. (2004) 'Gestures that co-occur with speech as a systematic resource: The realization of experiential meanings in indexes', *Social Semiotics*, 14(2), pp. 193-213.

Maxwell, J. A. (2005) *Qualitative research design: an interactive approach (2nd ed.)*. Thousand Oaks, London, New Delhi: SAGE.

Mayer, R. E. (2005) *The Cambridge Handbook of Multimedia Learning*. New York: Cambridge University Press.

McCafferty, S. G. and 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.

Mertens, D. M. (2010) *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods (3rd ed.)*. Thousand Oaks, CA: Sage.

Meyer, S. B. and Lunay, B. (2013) 'The Application of Abductive and Retroductive Inference for the Design and Analysis of Theory-Driven Sociological Research', *Sociological Research Online*, 18(1), pp. 86–96.

Morgan, D.L. (2007) 'Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods', *Journal of mixed methods research*, 1(1), pp. 48-76.

Nelson, M. (2006) 'Mode, meaning, and synaesthesia in multimedia L2 writing', *Language Learning and Technology*, 10(2), pp. 56-76.

Neuge, T. (2004) *Conversational analysis of chatroom talk*. PhD thesis. The University of South Australia, Adelaide, Australia.

Nik, A., Adams, R. and Newton, J. (2012) 'Writing to learn via text-chat: Task implementation and focus on form', *Journal of Second Language Writing*, 21(1), pp. 23-39.

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

Oliver, R. (2000) 'Age differences in negotiation and feedback in classroom and pair work', *Language learning*, 50(1), pp. 119-151.

O'Rourke, B. (2005) 'Form-focused interaction in online tandem learning', *CALICO Journal*, 22(3), pp. 433-466.

O'Rourke, B., Prendergast, C., Shi, L., Smith, B. and Stickler, U. (2015) 'Eyetracking in CALL - present and future', in Gimeno Sanz, A., Levy, M., Blin, F. and Barr, D. (eds.) *WorldCALL: Sustainability and computer-assisted language learning*. London: Bloomsbury, pp. 285-298.

Patton, M. Q. (1990) *Qualitative Evaluation and Research Methods*. Newbury Park: Sage.

Pellettieri, J. (1999) *Why-Talk?: Investigating the role of task-based interaction through synchronous network-based communication among classroom learners of Spanish*. PhD thesis. The University of California at Davis. Available at: <https://learntechlib.org/p/119206> (Accessed: 15 August 2017).

Pellettieri, J. (2000) 'Negotiation in cyberspace: The role of chatting in the development of grammatical competence', in Warschauer, M. and Kern, R. (eds.), *Network-based language teaching: Concepts and practice*. Cambridge: Cambridge University Press, pp. 59–86.

Perniss, P. (2018) 'Why we should study multimodal language', *Frontiers in psychology*, 9:1109. Available at: <https://doi.org/10.3389/fpsyg.2018.01109> (Accessed: 10 July 2019).

Phillips, D. C. and Burbules, N. C. (2000) *Postpositivism and educational research*. Lanham, MD: Rowman and Littlefield.

Pica, T. (1991) 'Classroom interaction, negotiation, and comprehension: Redefining relationships', *System*, 19(4), pp. 437-452.

Pica, T. (1994) 'Research on negotiation: What does it reveal about second-language learning conditions, processes, and outcomes?', *Language learning*, 44(3), pp. 493-527.

Pica, T., Kanagy, R. and Falodun, J. (1993) 'Choosing and using communication tasks for second language research and instruction', in Crookes, G. and Gass, S. (eds.) *Tasks and second language learning*. Clevedon, UK: Multilingual Matters, pp. 9-34.

Pinnow, R. J. (2011) 'I've got an idea: A social semiotic perspective on agency in the second language classroom', *Linguistics and Education*, 22(4), pp. 383-392.

Pressley, M., Graham, S. and Harris, K. (2006) 'The state of educational intervention research as viewed through the lens of literacy intervention', *British Journal of Educational Psychology*, 76(1), pp.1-19.

Renner, J. (2017) 'Negotiation of meaning and language-related episodes in synchronous, audio-based Chinese-German eTandem', *Language Learning in Higher Education*, 7(1), pp. 137-159.

Riazi, A. (2016) *The Routledge Encyclopedia of Research Methods in Applied Linguistics*. Oxon: Routledge.

Richards, K. (2003). *Qualitative inquiry in TESOL*. Hampshire: Palgrave Macmillan.

Rosell-Aguilar, F. (2005) 'Task design for audiographic conferencing: Promoting beginner oral interaction in distance language learning', *Computer Assisted Language Learning*, 18(5), pp. 417-442.

Rost, M. and Ross, S. (1991) 'Learner use of strategies in interaction: Typology and teachability', *Language Learning*, 41(2), pp. 231–249.

Rouhshad, A., Wigglesworth, G. and Storch, N. (2016) 'The nature of negotiations in face-to-face versus computer-mediated communication in pair interactions', *Language Teaching Research*, 20(4), pp. 514-534.

Royce, T. (2002) 'Multimodality in the TESOL classroom: Exploring visual-verbal synergy', *Tesol Quarterly*, 32(2), pp. 191-205.

Royce, T. (2006) 'Multimodal communicative competence in second language contexts', in T.D. Royce and W.L. Bowcher (eds.) *New directions in the analysis of multimodal discourse*. New Jersey: Lawrence Erlbaum Associates, pp. 361-390.

Ryan, J. and Gass, S. (2012) 'Stimulated recall', in Barnard, R. and Burns, A. (eds.) *Teacher cognition and practice: International case studies to illuminate methodological issues*. Bristol, UK: Multilingual Matters, pp. 144-161.

Sacks, H. (1992) *Lectures on conversation*. Cambridge, MA: Blackwell.

Satar, H. M. (2013) 'Multimodal language learner interactions via desktop videoconferencing within a framework of social presence: Gaze', *ReCALL*, 25(1), pp. 122-142.

Satar, H.M. (2015) 'Sustaining multimodal language learner interactions online', *Calico journal*, 32(3), pp. 480-507.

Satar, H. M. (2016) 'Meaning-making in online language learner interactions via desktop videoconferencing', *ReCALL*, 28(3), pp. 305-325.

Satar, H. M. and Ozdener, N. (2008) 'The effects of synchronous CMC on speaking proficiency and anxiety: Text versus voice chat', *Modern Language Journal*, 92(4), pp. 595-613.

Satar, H. M. and Wigham, C. R. (2017) 'Multimodal instruction-giving practices in webconferencing-supported language teaching', *System*, 70(2017), pp. 63-80.

Saunders, M., Lewis, P. and Thornhill, A. (2019) *Research Methods for Business Students (8th edition)*. England: Pearson Education Limited.

Sauro, S. (2011) 'SCMC for SLA: A research synthesis', *Calico Journal*, 28(2), pp. 369-391.

Schegloff, E. (2000) 'Overlapping talk and the organization of turn-taking for conversation', *Language in Society*, 29(1), pp. 1-63.

Schmidt, R. (1990) 'The role of consciousness in second language learning', *Applied Linguistics*, 11(2), pp. 129-158.

Schmidt, R. (1993) 'Awareness and second language acquisition', *Annual Review of Applied Linguistics*, 13, pp. 206–226.

Scollon, R. and Scollon, S. (2003) *Discourses in Place: Language in the Material World*. London: Routledge.

Seedhouse, P. (2008) 'Learning to talk the talk: Conversation analysis as a tool for induction of trainee teachers', in Garton, S. and Richards, K. (eds.) *Professional encounters in TESOL: Discourses of teachers in training*. Basingstoke: Palgrave Macmillan, pp. 42-57.

Shekary, M. and Tahirin, M. H. (2006) 'Negotiation of meaning and noticing in text-based online chat', *The Modern Language Journal*, 90(4), pp. 557-572.

Sime, D. (2008) 'Because of her gesture, it's very easy to understand - learners' perceptions of teachers' gestures in the foreign language class', in McCafferty, S. G. and Stam, G. (eds.)

Gesture: Second language acquisition and classroom research. New York: Routledge, pp. 259-279.

Sindoni, M. G. (2011) 'Online conversations: A sociolinguistic investigation into young adults' use of videochats', *Classroom Discourse*, 2(2), pp. 219-235.

Sindoni, M. G. (2013) *Spoken and Written Discourse in Online Interactions: A Multimodal Approach*. London: Routledge.

Sindoni, M. G. (2014) 'Through the looking glass: a social semiotic and linguistic perspective on the study of video chats', *Text and Talk*, 34(3), pp. 325-347.

Smith, B. (2001) *Taking students to tasks: Task-based computer-mediated communication and negotiated interaction in the ESL classroom*. PhD thesis. University of Arizona, Tucson.

Smith, B. (2003) 'Computer-mediated negotiated interaction: An expanded model', *The Modern Language Journal*, 87(1), pp. 38-57.

Smith, B. (2004) 'Computer-mediated negotiated interaction and lexical acquisition', *Studies in second language acquisition*, 26(3), pp. 365-398.

Smith, B. (2005) 'The relationship between negotiated interaction, learner uptake, and lexical acquisition in task-based computer-mediated communication', *Tesol Quarterly*, 39(1), pp. 33-58.

Smith, B. (2009) 'The relationship between scrolling, negotiation, and self-initiated self-repair in an SCMC environment', *CALICO Journal*, 26(2), pp. 231-245.

Sotillo, S. M. (2000) 'Discourse functions and syntactic complexity in synchronous and

asynchronous communication', *Language Learning and Technology*, 4(1), pp. 82-119.

Stickler, U. and Hampel, R. (2015) 'Qualitative research in CALL', *Calico journal*, 32(3), pp. 380-395.

Stickler, U. and Shi, L. (2015) 'Eye movements of online Chinese learners', *Calico Journal*, 32(1), pp. 52-81.

Stockwell, G. (2010) 'Effects of multimodality in computer-mediated communication tasks', in Thomas, M. and Reinders, H. (eds.) *Task-based language learning and teaching with technology*. London: Continuum International Publishers Group, pp. 83-104.

Swain, M. (1985) 'Communicative competence: Some roles of comprehensible input and comprehensible output in its development', in Gass, S. and Madden, C. (eds.) *Input in second language acquisition*. Rowley, MA: Newbury House, pp. 235–253.

Swain, M. (1995) 'Three functions of output in second language learning', in Cook, G. and Seidlhofer, B. (eds.) *Principle and practice in applied linguistics: Studies in honor of Henry G. Widdowson*. Oxford: Oxford University Press, pp. 125–144.

Swain, M. (2000) 'The output hypothesis and beyond: Mediating acquisition through collaborative dialogue', in Lantolf, J. P. (ed.) *Sociocultural theory and second language learning*. Oxford: Oxford University Press, pp. 97-114.

Swain, M., & Sharon L. (1998). Interaction and second language learning: two adolescent French immersion students working together. *Modern Language Journal*, 82(3), 320–338.

Tannen, D. (2005) *Conversational style: Analyzing talk among friends*. New York: Oxford University Press.

Tannen, D. (2012) 'Turn-taking and intercultural discourse and communication', in Paulston, C., Kiesling, S. and Rangel, E. (eds.) *The handbook of intercultural discourse and communication*. Chichester: Wiley-Blackwell, pp. 135–157.

Tashakkori, A. and Teddlie, C. (1998) *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.

Topping, K. J. and Ehly, S. (1998) *Peer-assisted learning*. Mahwah, NJ: Lawrence Erlbaum Associates.

Van der Zwaard, R. and Bannink, A. (2014) 'Video call or chat? Negotiation of meaning and issues of face in telecollaboration', *System*, 44 (2014), pp. 137-148.

Van der Zwaard, R., and Bannink, A. (2016) 'Nonoccurrence of negotiation of meaning in task-based synchronous computer-mediated communication', *The Modern Language Journal*, 100(3), pp. 625-640.

Van der Zwaard, R., and Bannink, A. (2018). 'Reversal of participation roles in NS-NNS synchronous telecollaboration'. *Calico journal*, 35(2), pp. 162-181.

Van der Zwaard, R., and Bannink, A. (2019). 'Towards a comprehensive model of negotiated interaction in computer-mediated communication'. *Language Learning and Technology*, 23(3), pp. 116-135.

Van Lier, L. (1996) *Interaction in the language curriculum: Awareness, autonomy and authenticity*. London: Longman.

Van Lier, L. (2004) *The Ecology and Semiotics of Language Learning: A Sociocultural Perspective*. Norwell, MA: Kluwer Academic Publishers.

Varonis, E. and Gass, S. (1985) 'Non-native/Non-native conversations: A model for negotiation of meaning', *Applied Linguistics*, 6(1), pp. 71-90.

Wagner-Gough, J. and Hatch, E. (1975) 'The importance of input data in second language acquisition studies', *Language Learning*, 25(2), pp. 297-308.

Wang, Y. (2004) 'Supporting synchronous distance language learning with desktop videoconferencing', *Language Learning and Technology*, 8(3), pp. 90-121.

Wang, Y. (2006) 'Negotiation of meaning in desktop videoconferencing-supported distance language learning', *ReCALL*, 18(1), pp. 122-146.

Wang, Y. (2007) 'Task design in videoconferencing-supported distance language learning', *Calico Journal*, 24(3), pp. 591-630.

Wang, Y. and Tian, J. (2013) 'Negotiation of meaning in multimodal tandem learning via desktop videoconferencing', *International Journal of Computer-Assisted Language Learning and Teaching*, 3(2), pp. 41-55.

Warschauer, M. (1996) 'Comparing face-to-face and electronic discussion in the second language classroom', *CALICO Journal*, 13(2-3), pp. 7-26.

Watanabe, Y. and Swain, M. (2007) 'Effects of proficiency differences and patterns of pair interaction on second language learning: Collaborative dialogue between adult ESL learners', *Language teaching research*, 11(2), pp. 121-142.

Wen, Q. (2016) 'The production-oriented approach to teaching university students English in China', *Language Teaching*, pp.1-15. DOI:10.1017/S026144481600001X.

Wen, Q. (2018) 'The production-oriented approach to teaching university students English in China', *Language Teaching*, 51(4), pp. 526-540.

White, L. (1991) 'Adverb placement in second language acquisition: Some effects of positive and negative evidence in the classroom', *Second Language Research*, 7, pp. 133–161.

White, L. (2003) *Second language acquisition and universal grammar*. Cambridge: Cambridge University Press.

Wigham, C. R. (2017) 'A multimodal analysis of lexical explanation sequences in web conferencing-supported language teaching', *Language Learning in Higher Education*, 7(1), pp. 81-108.

Wildfeuer, J. (2014) *Film discourse interpretation: Towards a new paradigm for multimodal film analysis*. New York: Routledge.

Worajittiphon, K. (2012) 'Negotiation for meaning in synchronous EFL chat', *Procedia Engineering*, 32, pp. 1157-1162.

Yamada, M. (2009) 'The role of social presence in learner-centred communicative language learning using synchronous computer-mediated communication: Experimental study', *Computers and Education*, 52(4), pp. 820-833.

Yamada, M. and Akahori, K. (2007) 'Social presence in synchronous CMC-based language learning: How does it affect the productive performance and consciousness of learning objectives?', *Computer Assisted Language Learning*, 20(1), pp. 37-65.

Yanguas, I. (2010) 'Oral computer-mediated interaction between L2 learners: It's about time!', *Language Learning and Technology*, 14(3), pp. 72-93.

Yilmaz, Y. (2011) 'Task effects on focus on form in synchronous computer-mediated communication', *The Modern Language Journal*, 95(1), pp. 115-132.

Yinger, R. J. (1986) 'Examining thought in action: A theoretical and methodological critique of research on interactive teaching', *Teaching and teacher education*, 2(3), pp. 263-282.

Yuksel, D. and Inan, B. (2014) 'The effects of communication mode on negotiation of meaning and its noticing', *ReCALL*, 26(3), pp. 333-354.

Zeng, G. and Takatsuka, S. (2009) 'Text-based peer-peer collaborative dialogue in a computer-mediated learning environment in the EFL context', *System*, 37(3), pp. 434-446.

Zhang, W., Wang, Y., Yang, L. and Wang, C. (2020) 'Suspending Classes Without Stopping Learning: China's Education Emergency Management Policy in the COVID-19 Outbreak', *J. Risk Financial Management*, 13(3), pp. 55-60. Available at <https://doi.org/10.3390/jrfm13030055> (Accessed: 5 May 2020).

Zheng, P. (2011), The features of language output and interaction patterns of dyads with different English language proficiency in text-based synchronous computer-mediated communication. *Foreign Language World*, 2011(3), 60-67.

Appendices

Appendix 1: Mock IELTS Speaking Test

Part 1:

The examiner asks the candidate about him/herself, his/her home, work or studies and other familiar topics. This part will last 4-5 minutes.

Part 2:

Describe a day when you thought the weather was perfect.

You should say:

where you were on this day

what the weather was like on this day

what did you do during the day

and explain why you thought the weather was perfect on this day.

You will have to talk about the topic for 1-2 minutes. You have 1 minute to think about what you are going to say. You can make some notes to help you if you wish.

Part 3:

Discussion topic: Types of weather

Example questions:

What types of weather do people in your country dislike most? Why is that?

What jobs can be affected by different weather conditions? Why?

Are there any important festivals in your country that celebrate a season or a type of weather?

Appendix 2: Opinion Gap Tasks

Task 1: Opinion Gap Task (Video Conferencing)

Topic: Food

What's your favourite food? List 3 of them and discuss why you like the food.

Do you cook? If so, what dishes are you good at cooking? If not, who does the cooking in your family?

Do you order takeaway food very often? Do you eat in restaurants very often?

What are the benefits and problems of ordering take-away food? What was your most enjoyable/terrible experience with takeaway food, if any?

What was your most enjoyable/terrible experience at a restaurant? What is the restaurant?

Where is it? What did you eat? Why do you like it?

Task 2: Opinion Gap Task (Audio Conferencing)

Topic: Shopping

Do you like shopping? Why or why not?

How do you usually do shopping? Go to shops or order online?

What things would you buy online and what things would you buy from shops?

What are the advantages and problems of shopping online?

What is your favourite shop? Why?

What is your most enjoyable/terrible shopping experience? Describe when, where, who, what, why?

Appendix 3: Task 3

Task 3: Spot-the-differences Sheet A (video conferencing)

Task Instruction:

You and your partner each have a picture. There are seven differences in the two pictures.

Please describe your picture to each other and find as many differences as possible.



Task 3: Spot-the-differences Sheet B (video conferencing)

Task Instruction:

You and your partner each have a picture. There are seven differences in the two pictures.

Please describe your picture to each other and find as many differences as possible.



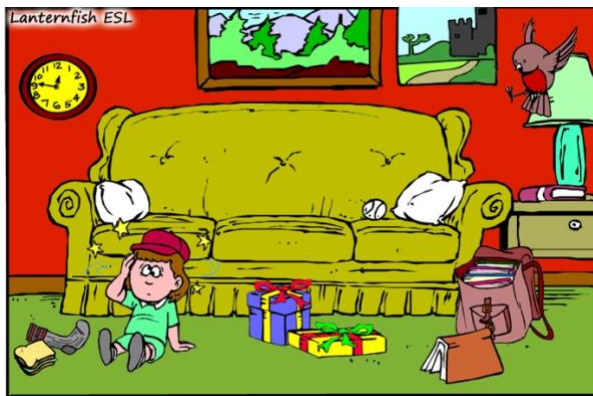
Appendix 4: Task 4

Task 4: Spot-the-differences task for Student A (audio conferencing)

Task Instruction:

You and your partner each have a picture. There are seven differences in the two pictures.

Please describe your picture to each other and find as many differences as possible.



Task 4: Spot-the--differences task for Student B (audio conferencing)

Task Instruction:

You and your partner each have a picture. There are seven differences in the two pictures.

Please describe your picture to each other and find as many differences as possible.



Appendix 5: Task 5

Task 5: Problem-solving task for Student A (video conferencing)

Part 1:

You and your friend are trying to decide on some gifts for your homestay family in the UK. Your host family has four members: Mr Jones (father), Mrs Jones (mother), Billy Jones (son, 15 years old), and Mary Jones (daughter, 14 years old). Below are some items you and your friend have noticed while shopping at the Mall, which may make good presents. Your friend has been shopping at the Mall and has also seen some (different) things that he/she thinks might make good presents. Since the presents will be from both of you, you must decide together on one present for each family member (four total).



Razor



Perfume



Stationery



Skateboard

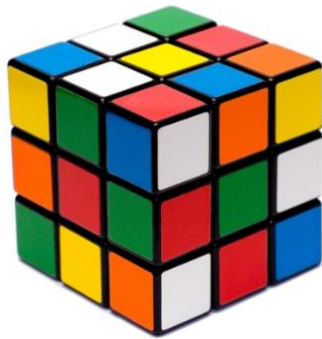
Part 2:

After you and your friend have decided on the four presents, discuss with your friend, estimate a reasonable price for each gift and calculate the total cost of these gifts.

Task 5: Problem-solving task for Student B (video conferencing)

Part 1:

You and your friend are trying to decide on some gifts for your homestay family in the UK. Your host family has four members: Mr Jones (father), Mrs Jones (mother), Billy Jones (son, 15 years old), and Mary Jones (daughter, 14 years old). Below are some items you and your friend have noticed while shopping at the Mall, which may make good presents. Your friend has been shopping at the Mall and has also seen some (different) things that he/she thinks might make good presents. Since the presents will be from both of you, you must decide together on one present for each family member (four total).



Rubik's cube



Bouquet



Cosmetics



Magnifying Glass

Part 2:

After you and your friend have decided on the four presents, discuss with your friend, estimate a reasonable price for each gift and calculate the total cost of these gifts.

Appendix 6: Task 6

Task 6: Problem-solving task for Student A (audio conferencing)

Part 1:

Students at BFSU are having a giant flea market sale to raise money for a trip to the Great Wall. In addition to working as 'sales assistants' next Saturday, all students have been asked to donate (give) some items that they no longer needed for sale. These items will be re-sold at the BFSU flea market sale next Saturday. Imagine that you and your partner are students living/staying in a BFSU dormitory on campus. Below is a list of used (old) items you have found in your room. Your chat partner also found some items! Together, decide on four items in total that you can donate to the flea market sale.



Roller Skating Shoes



Portable clothes rack



Blender



Nail polish

Part 2:

Sometimes, it is not easy to sell things at a flea market because the items may be old, broken, out of fashion, etc. After you and your partner have decided on the four items you will donate, discuss how you will convince (persuade) people to buy these items, and discuss at what price you plan to sell these items. You may wish, for example, to talk about the usefulness of the items, their value, their condition, etc.

Task 6: Problem-solving task for Student B (audio conferencing)

Part 1:

Students at BFSU are having a giant flea market sale to raise money for a trip to the great wall. In addition to working as 'sales assistants' next Saturday, students have all be asked to donate (give) some items that they no longer need for sale. These items will be re-sold at the BFSU flea market sale next Saturday. Imagine that you and your chat partner are students in a BFSU dormitory on campus. Below is a list of used (old) items you have found in your room. Your chat partner also found some items! Together, decide on four items in total that you can donate to the flea market sale.



Toaster



Piggy bank



Go (Board game)



Phone accessories

Part 2:

Sometimes, it is not easy to sell things at a flea market because the items may be old, broken, out of fashion, etc. After you and your partner have decided on the four items you will donate, discuss how you will convince (persuade) people to buy these items, and discuss at what price you plan to sell these items. You may wish, for example, to talk about the usefulness of the items, their value, their condition, etc.

Appendix 7: Information sheet for participants

Dear participants,

Thank you very much for participating in these advanced oral English classes. The aim of this project is to explore how meaning is negotiated through synchronous video/audio conferencing classes and how multiple modes are used in video SCMC meaning negotiation interactions. This study is designed by Chenxi Li, a doctoral researcher at the Centre for Research in Education and Educational Technology from the Open University in the UK.

As research participants, you will be asked to take a pre-task vocabulary test, four online sessions, and a post-task video stimulated recall interview. Your performance in video/audio conferencing classes will be video recorded as research data for analysis and presentation.

The data generated will be stored in the Open University UK and will be destroyed after five years. Your participation will be anonymised. The confidentiality of the information you provide will be safeguarded subject to any legal requirements. As a research participant, you have the right to withdraw your participation at any time of the study.

I hope you can enjoy participating in this research project. A summary copy of the research findings will be forwarded to you on request. If you have any problems with this research project, please feel free to contact the research at Cecilia.Li@open.ac.uk.

Kind Regards


Chenxi (Cecilia) Li

Faculty of wellbeing, education and language studies

Centre for Research in Education and Educational Technology

The Open University, Milton Keynes, UK, MK7 6AA

Appendix 8: Sample signed consent form

 The Open University

Centre for Research in Education and Educational Technology

Consent form for Belwal online language learners participating in a research project
Meaning negotiation patterns in synchronous audio and video conferencing ELT classes


Name of participant: _____

Name of principal investigator(s): Chenxi (Cecilia) Li

1. I consent to participate in this project, the details of which have been explained to me, and I have been provided with a written statement in plain language to keep.
2. I understand that my participation will involve a stimulated recall interview and I agree that the researcher may use the results as described in the plain language statement.
3. I acknowledge that:
 - (a) the possible effects of participating in this research have been explained to my satisfaction;
 - (b) I have been informed that I am free to withdraw from the project at any time without explanation or prejudice and to withdraw any unprocessed data I have provided;
 - (c) the project is for the purpose of research;
 - (d) I have been informed that the confidentiality of the information I provide will be safeguarded subject to any legal requirements;
 - (e) I have been informed that with my consent the data generated will be stored in the Open University UK and will be destroyed after five years;
 - (f) if necessary any data from me will be referred to by a pseudonym in any publications arising from the research;
 - (g) I have been informed that a summary copy of the research findings will be forwarded to me, should I request this.

I consent to this stimulated recall interview being audio-taped/video-recorded yes no

I wish to receive a copy of the summary project report on research findings yes no

Participant signature:  Date: 2017. 4. 19

Chenxi (Cecilia) Li
Email: Cecilia.Li@open.ac.uk
Centre for Research in Education and Educational Technology
The Open University, UK

Appendix 9: Interview questions

Part 1:

Video stimulated recall interview questions about gaze directions:

E.g.:

1. Do you care how you look in the camera?
2. Did you look at your own video image and/or your peer's video image very often during task interactions?
3. Where were you looking at this point of the interaction? Why?
4. What information (if any) can you obtain by looking at your peer's video?

Part 2:

Video stimulated recall interview questions about thoughts and (non-)understandings during meaning negotiation episodes

E.g.:

1. Did you know this word before?
2. Here when she said razor, did you get it?
3. but were you taking notes?
4. Did you think interrupting people in the middle and asking 'what do you mean by this' is not polite?
5. What were you thinking when she said magnifying glass, you asked a pair of glasses?
6. Why didn't you try to confirm with her whether what you guessed is correct or not?

Part 3:

Additional interview questions (not stimulated, not about particular meaning negotiation episodes)

1. What are your opinions and preferences for peer interaction through audio and video SCMC?
2. What are your opinions on the student-centred task based language teaching method?
3. What are your opinions on technology-related issues in audio and video SCMC?

4. Could you please tell me about your previous language learning experience? How confident are you about your oral English?
5. How do you usually deal with new words in daily conversation or during your daily reading, listening exercises and conversations with native speakers (if any)?
6. Could you please tell me your basic personal information, particularly about your job, and what are the chances of using English in your job or daily life?