Towards the Multimodal Unit of Meaning:
A Multimodal Corpus-based Approach

Yaoyao Chen, MA, BA.

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

January 2019

CONTENTS

ACKNOWLEDGMENT I

ABSTRACT II

LIST OF FIGURES III

LIST OF TABLES VII

CHAPTER 1. INTRODUCTION 1

1.1. Introduction 1

1.2. Research questions and research design 5

1.2.1. Framework development 6

1.2.2. The first study 7

1.2.3. The second study 10

1.3. Thesis overview 12

CHAPTER 2. LITERATURE REVIEW 14

2.1. Introduction 14

2.2. Multimodality 15

2.2.1. Definition, origin and strands of research 16

2.2.2. Corpora: monomodal vs. multimodal 20

2.2.2.1. Representativeness 20

2.2.2.2. Size 26

2.2.2.3. Machine-readable 29

2.2.2.4. A standard reference 34

2.3. Formulaic sequences 36

2.3.1. Defining Formulaic Sequences 38

2.3.2. Frequency and formulaic sequences 39

2.3.3. The processing advantage of formulaic sequences 41

2.4. Gesture: definition, frameworks and analysis 42

2.4.1. Defining gestures 43

2.4.2. Categorising gestures 45

2.4.3. Frameworks for segmenting gesture phases 49

2.4.4. Temporal and semantic relationship between speech and gesture 53

2.4.5. Analytical software 56

2.5. Towards the meaning of gestures of high occurrence 59

2.5.1. Emblems 60

2.5.2. Gesture families 63

2.5.3. Recurrent gestures 66

2.6. The functions of gestures 69

2.7. Interpreting gesture-meaning association 71

2.7.1. Gesture and metaphor 72

2.7.2. Analogic links in gesture elements 74

2.8. Towards the multimodal units of meaning 76

2.9. Summary 80

CHAPTER 3. DATA AND FRAMEWORK DEVELOPMENT FOR SPEECH-GESTURE ALIGNMENT 83

3.1. Introduction 83

3.2. An overview of NMMC 83

3.2.1. The sub-corpora of NMMC 84

3.2.2. Setup of NMMC 88

3.2.3. Speech transcripts and representation in NMMC 91

3.2.4. Ethics 94

3.3. Gesture phases and coding procedures 96

3.4. Gesture coding schemes 101

3.4.1. Coding preparation phase 102

3.4.1.1. Fleeting preparation coded as no preparation 102

3.4.1.2. Lengthy preparation 103

3.4.1.3. No preparation between successive strokes 104

3.4.1.4. Preparation following post-stroke hold 104

3.4.2. Coding the stroke phase 106

3.4.2.1. Palm Up Open Hand Gesture 106

3.4.2.2. Bounded-space gesture 108

3.4.2.3. Beat gesture 110

3.4.2.4. Bi-directional symmetric movements 111

3.4.2.5. Bi-directional asymmetric movements 114

3.4.2.6. Continuous movements 114

3.4.3. Coding the pre-stroke and post-stroke hold phases 117

3.5. Transcribing and representing gestures 118

3.8. Summary 122

CHAPTER 4. TOWARDS A SPEECH-GESTURE PROFILE OF FORMULAIC SEQUENCES: A CORPUS-BASED APPROACH 124

4.1. Introduction 124

4.2. Approach 126

4.2.1. The video for the case study and the TFSs 126

4.2.2. Coding speech functions 129

4.2.3. Segmenting gestures in ELAN 131

4.3. Results of the case study 133

4.3.1. Temporal relationship between TFS and gesture 133

4.3.2. Identifying gesture patterns based on functional variations 140

4.3.2.1. TFSs as checking understanding with beat/beat-like strokes, beat/beat-like strokes with long post-stroke holds and post-stroke hold phases. 141

4.3.2.2. TFSs as a combination of checking understanding and discourse marker and embedded in stroke phases 153

4.3.2.3. TFSs as discourse markers with no gesture 159

4.4. Results of the extended data analysis 165

4.4.1. Temporal relationship between TFS and gesture 166

4.4.2. Identifying gesture patterns based on functional variations 167

4.4.2.1. TFSs checking understanding and their co-occurring gestures 168

4.4.2.2. TFSs as discourse markers and the co-occurring gestures 176

4.4.2.3. TFS as claiming mutual understanding and the co-occurring gestures 180

4.5. Discussion 184

4.5.1. Temporal relationship between TFS and gesture 184

4.5.2. Functional coordination between TFS and gesture 187

4.5.2.1. TFSs checking understanding with beat/beat-like strokes, beat/beat-like strokes with long post-stroke holds and post-stroke hold phases 187

4.5.2.2. TFS as checking understanding and as discourse markers with the Palm Up Open Hand gesture 190

4.5.2.3. TFS claiming mutual understanding with a fling gesture 191
4.5.3. Meaning on a continuum 192
4.5.4. The discourse functions of gestures 195

4.6. Summary 198

CHATPER 5. EXPLORING THE UNIT OF MEANING OF THE CIRCULAR GESTURE 201

5.1. Introduction 201

5.2. Approach 204

5.2.1. Defining and selecting the circular gesture 204
5.2.2. The form of the circular gesture and discourse context 213
5.2.2.1. Number of circles 214
5.2.2.2. Direction of rotation 215
5.2.2.3. Shape of hand 216
5.2.2.4. Size of circles 217
5.2.2.5. Space of gesture 218
5.2.3. The unit of meaning model 219
5.2.4. The use of corpus techniques in multimodal analysis 222
5.2.4.1. Concordance 223
5.2.4.2. Frequency analysis 225
5.2.4.3. Clusters/N-Grams 230
5.2.5. Coding lexical and grammatical patterning 231
5.2.6. The multimodal unit of meaning of the circular gesture 236

5.3. Data analysis and results 237

5.3.1. Temporal relationship between speech and the circular gesture 238
5.3.1.1. The length of the words 238

5.3.1.2. The position of circular gesture and linguistic patterning 243
5.3.2. The multimodal unit of meaning of the circular gesture 244
5.3.2.1. Colligation 248
5.3.2.2. Semantic preference 252
5.3.2.3. Semantic prosody 256

5.4. Discussion 259
5.4.1. Corpus approach and gesture-speech relationship 260
5.4.2. The multimodal unit of meaning of the circular gesture 262
5.4.3. The functions of the circular gesture 264
5.4.3.1. Referential function 265
5.4.3.2. Segmenting utterances 266
5.4.3.3. Underlining noteworthy information 267
5.4.3.4. Maintaining discourse coherence 269
5.4.3.5. Enhancing inter-subjectivity 272

5.5. Summary 273

CHAPTER 6. TOWARDS A NEW UNDERSTANDING OF SPEECH-GESTURE RELATIONSHIP: TEMPORAL, MEANING AND DISCURSIVE FUNCTIONAL PERSPECTIVES 276

6.1. Introduction 276

6.2. Temporal relationship between speech and gesture 279
6.2.1. TFSs and non-iconic gestures 279
6.2.2. The position of recurrent gestures 285

6.3. Meaning coordination between speech and gesture and metaphor 287
6.3.1. Conceptual Metaphor Theory as an explanatory tool 288
6.3.2. Gesture parameters and their metaphoric links 292

6.4. Functions of gestures 294

6.4.1. For speech production 295

6.4.2. For communication 297

6.4.3. Interactive gestures 301

6.4.4. Discourse functions 303

6.5. Summary 306

CHAPTER 7. CONCLUSION 308

7.1. Thesis overview 308

7.2. Limitations and avenues for future research 310

7.3. Implications for applied linguistics 312

7.4. Applications of the approaches 318

7.5. Summary 326

APPENDIX I SPEECH TRANSCRIPTION CONVENTIONS FOR THE NOTTINGHAM MULTIMODAL CORPORA 328

APPENDIX II A SAMPLE OF PARTICIPANT CONSENT FORM IN NMMC330

APPENDIX III CONSENT FORM FOR RESEARCHERS (SIGNED) 331

GUIDE TO CD-ROM DATA DISK 332

REFERENCES 333

Acknowledgment

I would like to thank many people who have helped me during my PhD. First and foremost, I am greatly indebted to my supervisor and mentor, the most inspiring, intelligent and kind Svenja Adolphs, who has provided me with invaluable guidance and support. I am also very grateful to my other supervisor, Louise Mullany, who has always been generous in offering encouragement and help whenever I needed.

Further thanks go to Dawn Knight, Simon Harrison, and Keiko Tsuchiya, for their useful feedback and suggestions for my works. Also, thanks to all the administrative staff in the School of English at the University of Nottingham; to my PhD fellows and friends at Nottingham, Abdulmalik Ofemile, Leigh Clark, Beatriz González Fernández, Yaxiao Cui, Jin Zhang, Xiaofan Xu, Guilia Grisot, Marianna Kyriacou, Aya Van Renterghem, Christine Muir, Linqiao Wu, Ruijun Zhang for their support and accompany.

Many thanks and love to my dearest parents, without whose support this PhD would not be possible. Finally, special thanks to my dear Jiaqi.

Abstract

While there has been a wealth of research that uses textually rendered spoken corpora (i.e. written transcripts of spoken language) and corpus methods to investigate utterance meaning in various contexts, multimodal corpus-based research beyond the text is still rare. Monomodal corpus-based research often limits our description and understanding of the meaning of words and phrases, mainly due to the fact that meaning is constructed by multiple modes (e.g. speech, gesture, prosody, etc.).

Hence, focusing on speech and gesture, the thesis explores multimodal corpus-based approaches for investigating multimodal units of meaning, using recurrent phrases, e.g. "(do) you know/see what I mean", and gesture as two different, yet complementary points of entry. The primary goal is to identify the patterned uses of gesture and speech that can assist in the description of multimodal units of meaning. The Nottingham Multimodal Corpus (250,000 running words) is used as the data base for the research.

The main original contributions of the thesis include a new coding scheme for segmenting gestures, two multimodal profiles for a target recurrent speech and gesture pattern, and a new framework for classifying and describing the role of gestures in discourse. Moreover, the thesis makes important implications for our understanding of the temporal, cognitive and functional relationship between speech and gesture; it also discusses potential applications, particularly in English language teaching and Human-Computer Interaction. These findings are of value to the methodological and theoretical development of multimodal corpus-based research on units of meaning.

List of figures

Figure 2.1: Framework for segmenting gesture space (McNeill, 2005, p. 274). 53
Figure 2.2: The purse hand. 61
Figure 2.3: The pray hand. 61
Figure 2.4 The Drinking Gesture (Brookes, 2005) 62
Figure 2.5: Two examples of the ring hand shape in the R-gesture family (Kendon, 2004). 64
Figure 2.6: Open Hand Supine gesture (Kendon, 2004). 64
Figure 2.7: Examples of Open Hand Prone gestures (Kendon, 2004). 65
Figure 2.8: The continuum of relationship to conventions in gesture (McNeill, 2005, p. 10). 68
Figure 2.9: Gestures segment idea units (Calbris, 2011). 69
Figure 2.10: The analogic links of the transverse movement of the level hand (Calbris 1990: 140, see in Calbris, 2011). 75
Figure 3.1: The set-up of camera in NMMC (Knight, 2011). 90
Figure 3.2: A screen shot of image in NMMC. 90
Figure 3.3: An episode of the transcript in NMMC. 93
Figure 3.4: Kendon's (2004) gesture transcription convention. 119
Figure 3.5: An example of multimodal transcription (Kendon, 2004). .. 120
Figure 3.6: An example of gesture transcription co-occurring with a TFS. 121
Figure 3.7: An example of gesture transcription for circular gestures. .. 122
Figure 4.1: Distribution of speech functions in TFS. 131
Figure 4.2: A screen shot of ELAN 132
Figure 4.3: A TFS co-occurring with a stroke phase with a long retraction phase. 134

Figure 4.4: A TFS co-occurring with an independent stroke/gesture..... 135
Figure 4.5: A TFS co-occurring with a preparation phase. 136
Figure 4.6: A TFS embedded in an extensive stroke phase..... 137
Figure 4.7: A TFS co-occurring with a post-stroke hold phase..... 137
Figure 4.8: A TFS co-occurring with a retraction phase. 138
Figure 4.9: A TFS co-occurring with a post-stroke hold phase and retraction. 138
Figure 4.10: A TFS co-occurring with a post-stroke hold phase and preparation. 139
Figure 4.11: Distribution of the temporal relationship between TFSs and gestures in Supervisor 6. 139
Figure 4.12: A TFS checking understanding co-occurring with a beat gesture. 143
Figure 4.13: A TFS as checking understanding co-occurring with a beat-like gesture. 145
Figure 4.14: A TFS checking understanding co-occurring with a beat-like gesture with a long post-stroke hold phase. 148
Figure 4.15: A TFS as checking understanding co-occurring with a post-stroke hold phase. 150
Figure 4.16: A TFS as checking understanding co-occurring with a circular gesture. 152
Figure 4.17: A TFS as a combination of checking understanding and discourse marker co-occurring with an embedded in stroke phase. 155
Figure 4.18: A TFS as a combination of checking understanding and discourse marker co-occurring with a preparation phase. 158
Figure 4.19: A TFS as a discourse marker co-occurring with no gesture. 161
Figure 4.20: A TFS as discourse marker co-occurring with a beat-like stroke/gesture. 164

Figure 4.21: Distribution of temporal relationship between TFSs and gestures in other speakers. 167
Figure 4.22: A TFS checking understanding co-occurring with a beat-like gesture. 169
Figure 4.23: A TFS checking understanding co-occurring with a beat-like gesture. 170
Figure 4.24: A TFS checking understanding co-occurring with an PUOH gesture. 172
Figure 4.25: A TFS checking understanding co-occurring with an PUOH gesture. 173
Figure 4.26: A TFS checking understanding co-occurring with a post-stroke hold phase. 174
Figure 4.27: A TFS as a discourse marker co-occurring with a bounded-space gesture. 178
Figure 4.28: A TFS claiming mutual understanding co-occurring with a fling gesture. 181
Figure 4.29: A TFS as claiming mutual understanding co-occurring with a fling gesture. 183
Figure 4.30: Speech-gesture functional profile for multimodal TFSs. ... 193
Figure 5.1: Examples of Cyclic Gestures from Ladewig (2014b). 206
Figure 5.2: An example of a circular gesture changing gesture space as the hand rotates. 209
Figure 5.3: An example of a circular gesture performed by a pen. 210
Figure 5.4: A collection of images captured by the camera. 212
Figure 5.5: An example of a circular gesture with 3 circles. 214
Figure 5.6: An example of a circular gesture with 1 circle. 215
Figure 5.7: An example of a circular gesture with a finger-pointing hand shape. 217

Figure 5.8: An example of a circular gesture with a particularly large size. 218
Figure 5.9: An example of a circular gesture performed in the upper peripheral space. 218
Figure 5.10: A screen shot of Elan. 224
Figure 5.11: A screen shot of a concordance list obtained by searching ‘CG’. 225
Figure 5.12: The concordance plot of circular gestures. 226
Figure 5.13: The correlation between the length of the video and the occurrence of the circular gesture. 227
Figure 5.14: The distribution of circular gestures by supervisors and students. 228
Figure 5.15: Frequency of number of words co-occurring with the circular gestures. 240
Figure 5.16: The average duration of circular gestures performed by each supervisor and student in each video. 242
Figure 5.17: The unit of meaning for circular gestures in relation to co-occurring speech. 246
Figure 5.18: Distribution of frequency of occurrence of speech patterns. 247
Figure 5.19: A deictic gesture with a referential function. 266
Figure 5.20: A circular gesture with discourse marking function. 267
Figure 5.21: A circular gesture co-occurring with a noun phrase. 269
Figure 5.22: A circular gesture co-occurring with a noun phrase. 269
Figure 5.23: A circular gesture co-occurring with an it-clause. 271
Figure 5.24: A circular gesture co-occurring with conjunction. 272
Figure 5.25: A circular gesture co-occurring with checking understanding. 273

List of tables

Table 2.1 An overview of existing multimodal corpora (Knight, 2011) .. 23

Table 3.1: An overview of the lectures. 85

Table 3.2: An overview of the supervision meetings in NMMC. 86

Table 4.1: Frequency of TFS variations. 128

Table 4.2: Gesture phases co-occurring with the TFSs as checking understanding. 141

Table 4.3: Temporal relationship between the TFSs checking understanding and beat gestures. 144

Table 4.4: Temporal relationship between the TFSs checking understanding and beat-like gestures. 146

Table 4.5: Temporal relationship between the TFSs checking understanding and beat/beat-like gestures with long post-stroke hold phases. 148

Table 4.6: Temporal relationship between TFSs checking understanding and post-stroke hold phases. 151

Table 4.7: Temporal relationship between the TFSs checking understanding and the circular gestures. 153

Table 4.8: Gesture phases co-occurring with the TFSs as a combination of checking understanding and discourse marker. 154

Table 4.9: Temporal relationship between the TFSs as a combination of checking understanding and discourse marker and embedded in stroke phases. 156

Table 4.10: Temporal relationship between the TFSs as a combination of checking understanding and discourse marker and preparation phases. 158

Table 4.11: Gesture types co-occurring with TFSs as discourse markers.159
Table 4.12: Temporal relationship between TFSs as discourse markers and gesture phases. 162
Table 4.13: Instances of TFSs from other speakers..... 166
Table 4.14: Gesture types co-occur with TFSs as checking understanding in other speakers 168
Table 4.15: Temporal relationship between the TFSs checking understanding and gestures in other speakers..... 175
Table 4.16: Gesture types co-occur with TFSs as discourse markers in other speakers. 176
Table 4.17: Temporal relationship between TFSs as discourse markers and gestures in other speakers. 179
Table 4.18: Gestures co-occurring with TFSs claiming mutual understanding with other speakers..... 183
Table 5.1: An overview of supervision meetings. 205
Table 5.2: The number of circular gestures per minute. 227
Table 5.3: The first 10 most frequently clusters/N-Grams to the right of 'cg'. 230
Table 5.4: Descriptive coding scheme for speech co-occurring with circular gestures. 233
Table 5.5: Frequency of the number of words co-occurring with the circular gestures. 239
Table 5.6: Examples of speech co-occurring with circular gestures containing more than 10 words..... 241
Table 5.7: The average duration of circular gestures performed by each supervisor and student. 242
Table 5.8: Top 10 most frequently occurring clusters to the right of the node 'CG'. 243

Table 5.9: Distribution of frequency of occurrence of speech patterns.. 247

Table 5.10: Number of verb phrases that occur twice or more. 254

Chapter 1. Introduction

1.1. Introduction

This thesis aims to explore multimodal corpus-based approaches for investigating speech and gesture as two integral components of spoken language. Since the advent of corpus linguistics and the rapid advance in computer technology, language, or texts, has become a type of data that can be systematically marked-up and analysed by computational software on a large scale (Adolphs, 2006; Biber, Conrad, & Reppen, 1998; Kennedy, 1998; McEnery & Wilson, 1996; Sinclair, 1991). Due to this, linguistic knowledge that was proposed based on researcher instinct and invented data can be re-examined and refined drawing on authentic language collected from daily life (Adolphs, 2008; Sinclair, 1991). One of the major insights is the shift of the analytical unit itself from single word to multiword. As stated by Sinclair (2004), corpus evidence indicates that words are not completely disconnected or isolated from one another, but tend to be associated and co-occur with certain other words; hence, the units of meaning not only lie in single words, but also, more often, in multiword expressions.

In spoken language, with the aid of spoken corpora and corpus tools (refer to Chapter 2, 2.2.2. for details), the extended research for the units of meaning at the phrasal level has been well executed in numerous studies on lexical bundles (Biber & Barbieri, 2007; Biber, Conrad, & Cortes, 2004; Hyland, 2008), formulaic sequences (Schmitt, 2004; Wray, 2002, 2008), discourse markers (Fung & Carter, 2007) and pragmatic markers (Erman, 2001), etc. However, such investigations mostly remain at the textual level, and this is problematic as spoken language is inherently multimodal (Adolphs & Carter, 2013). There is a distinct lack of knowledge regarding how other modes (e.g. gesture, facial expression, prosody, etc.) interact with

speech, whether the uses of other modes are as formulaic as speech, and what roles the other modes play in the units of meaning. We also lack theoretical and methodological frameworks that can help us explore and describe these queries.

In view of the aforementioned issues, the current thesis intends to take on the challenges and explore corpus-based approaches in order to examine the multimodal units of meaning in speech and gesture and the role of gestures by drawing on a large multimodal corpus, i.e. the Nottingham Multimodal Corpus. The major contributions of the thesis entail the development of a gesture coding scheme and transcription convention, together with two main studies, both of which aim to allow new speech-gesture meaning patterns to emerge. The two studies are designed to use different points of entry, with the first starting with a group of recurrent multiword expressions, i.e. a formulaic sequence (Wray, 2008), and the second with a recurrent gesture. The two studies are not only complementary as two distinctive target aspects of the multimodal units of meaning, they are also incremental as the second study is much boarder in scope and much more complex in design than the first.

Before the two studies can proceed, the initial issue that requires to be addressed is the ways in which gestures can be segmented and represented along with speech in accordance with the research agenda of the thesis. To examine how speech and gesture co-occur, previous researchers have developed frameworks to segment gestures into different phases with reference to the changes of the trajectory, velocity, movement patterns, etc. of gestures (refer to Chapter 2, 2.4.3. and Chapter 3, 3.3. for information) (Kendon, 2004; Kita, van Gijn, & van der Hulst, 1998; McNeill, 2005). With minor variations, the gesture phases included in a framework can include the preparation phase, pre-stroke hold phase, stroke phase, post-stroke hold phase and retraction phase. The key is to identify the stroke phase, which is

the most laborious part of a gesture where the meaning lies in (e.g. the movement of extending the hands horizontally to the opposite directions as someone says “separate”). For a gesture to exist, it must have a stroke phase. In other words, the existence of one stroke equals one independent gesture. It is acknowledged that the strokes are usually in line with the speech components they are associated with in meaning, and the associated speech is called ‘the lexical affiliate’ (Schegloff, 1984), or ‘the idea unit’ (Kendon, 2004) (refer to Chapter 2, 2.4.4. for more information of the speech-gesture temporal and semantic relationship).

Whereas all previous gesture studies have provided similar definitions for the gesture phases in their frameworks, the actual coding processes in each phase involves numerous important decisions and few publications have offered detailed descriptions of these. This leads to the development of a data-based coding scheme for segmenting gestures. The gesture coding scheme constructed here contains the most fine-grained guidelines for demarcating the frequently occurring movement patterns and gesture types, which have not been systematically examined, synthesised and reported in the existing literature.

The thesis also designs a new format for speech-gesture transcription, which effectively represents the alignment of the audio/speech and visual/gestural data streams and the manual construction of multimodal concordances (refer to Chapter 3, 3.5. for details). This new format of data representation is very much consistent with the core of the corpus methods, offering considerable assistance in the analysis of emerging speech-gesture patterns. These initial methodological explorations are not only crucial for laying the ground for the investigations of the two main studies of the thesis, but also valuable in shedding new light on the systematic segmentation, representation and analysis of gestures in future research.

After the gesture coding schemes and transcription conventions have

been developed, the thesis proceeds with the first study, aiming for the development of a multimodal corpus-based approach for examining speech-gesture relationship in a group of formulaic sequences, "(do) you know/see what I mean". A speech-gesture profile is established based on the emerging patterns of speech-gesture associations, representing and describing the roles of speech and gesture in this formulaic sequence. As the major outcomes of the study, the speech-gesture profile built up can be used in further research on the same formulaic sequence, and the multimodal corpus-based approach developed in this study can be adapted to future investigations of other multiword strings. Moreover, the first study also develops a framework for describing the functions of gestures in relation to speech in discourse based on the dataset. This functional framework of gestures is then verified and applied in the second study.

The second study focuses on investigating the multimodal units of meaning based on a recurrent gesture, i.e. the circular gesture (refer to Chapter 2, 2.5. for previous research on recurrent gestures). Unlike corpus-based research on texts, large-scale corpus-based investigations of the meaning of frequently occurring gestures (i.e. head nods as a sub-category of gesture) only started around ten years ago (Adolphs & Carter, 2013; Carter & Adolphs, 2008; Knight, 2011). Whereas the technologies for building, annotating and analysing textual/monomodal corpora has become more sophisticated, such technological support for conducting multimodal research on frequently occurring gestures is rather insufficient. As a significant field of research, most existing gesture studies tend to adopt anthropological and qualitative approaches to investigate gesture meanings, providing thick descriptions of a few examples (refer to Chapter 2, 2.2. for definitions and strands of research in multimodality) (Hazel, Mortensen, & Rasmussen, 2014; Kendon, 1995, 2004; Mondada, 2007, 2014; Streeck, 2009a). As manual annotation and analysis of gestures takes considerable time, even the

currently largest corpus-based research on the functions of recurrent hand-and-arm gestures, i.e. the cyclic gesture, does not exceed 500 instances (Ladewig, 2011). In view of this, we lack empirical evidence in terms of what those recurrent gestures in the existing literature actually mean in real-life communication. Moreover, as previous research on recurrent gestures only describe the general notion or function of a recurrent gesture (e.g. the cyclic gesture is often related to continuity), there is little research that has attempted to establish a fine-grained lexical and grammatical framework for the emerging speech patterns co-occurring with a certain recurrent gesture. In addition, we also lack a comprehensive framework for describing the discourse functions of recurrent gestures in accordance with the emerging speech patterns.

Therefore, the second study of the thesis expands on the research on cyclic gestures (i.e. outward rotational gestures that remain in the same position), selecting all the circular movements of various orientations and trajectories (i.e. circular gestures) in the corpus and building up a lexical and grammatical framework for the speech co-occurring with them. For this purpose, Sinclair's (2004) unit of meaning framework, i.e. *collocation*, *colligation*, *semantic preference* and *semantic prosody*, is adopted to describe the relationship between the circular gestures and the emerging speech patterns. The multimodal unit of meaning built up for the circular gestures can be used as a framework for future research on the same gesture, and the approach developed in the current study can be used to examine the units of meaning of other recurrent gestures.

1.2. Research questions and research design

The following three sub-sections will focus on demonstrating the research questions during the framework development, the first study and the second

study, along with the methods for answering these questions.

1.2.1. Framework development

The framework development in the thesis mainly attempts to provide answers to the following questions:

1) Can gestures, to a great extent, be segmented in mute in order to avoid the influence of speech?
2) How can the emerging movement patterns and phenomena be coded in different gesture phases?

The first question is a response to the criticism of McNeill (1992, 2005), who points out the issue of the influence of speech on the objectivity of gesture segmentation and analysis in most gesture studies. That is, there is a circulation existing during the process of segmenting gestures with the assist of speech and then examine the temporal and semantic relationship between speech and gesture. To generate the most objective results of speech-gesture relationship, the gestures are best segmented independent of speech, at least initially. Speech can be involved at the check-up stage when necessary. To test the possibility of this, the current research ran a pilot study and selected 25 gestures from 5 videos and segmented them in mute. The gesture segmentation and all the annotations were carried out with the assistance of the ELAN[1] software (Sloetjes & Wittenburg, 2008). The results show that all the gestures can be correctly demarcated into gesture phases in mute once the target gesture has been identified. In view of the findings of the pilot study, all the target gestures in the thesis are segmented isolated from speech.

For the second question, the development of a gesture coding scheme

[1] https://tla.mpi.nl/tools/tla-tools/elan/

has the dual function of recording and reporting the coding processes, and providing a framework for future researchers. The thesis adopts a five-phase gesture coding scheme, including preparation (the preparation movement before the stroke), pre-stroke hold (the hold period immediately before the stroke after the preparation phase), stroke (the apex of a gesture), post-stroke hold (the hold period directly after the stroke) and the retraction phases (the relaxation period) (refer to Chapter 3, 3.3.3. for more detailed definitions and illustrations). The key to developing this gesture coding scheme is through repetitive observations of the various movement patterns and phenomena in each gesture phases, and synthesising them into inclusive guidelines. The gestures for the first study (49 instances) are used as the basis for this. The coding scheme (refer to Chapter 3, 3.4.1. for details) reveals 4 types of phenomenon in the preparation phase, and 6 in the stroke phases including 3 frequently occurring gestures and 3 movement patterns. The scheme even determines the minimum length of the pre-stroke and post-stroke phases. This coding scheme is used as the guidelines for coding gestures in both studies.

1.2.2. The first study

The first study explores the multimodal units of meaning in a group of target formulaic sequences (TFSs), the three questions are:

1) What is the temporal relationship between the TFSs and the co-occurring gestures?
2) What is the unit of meaning of the TFSs in accordance with their co-occurring gestures?
3) What are the discourse functions served by the gestures co-occurring with the TFSs?

The first study comprises a case study followed by extended analysis, drawing on the supervision-meeting sub-corpus of the Nottingham Multimodal Corpus. The video used for the case study lasts for 47:52 minutes with 11,705 running words, and 64 instances of "(do) you know/see what I mean" are identified, all of which are used by the supervisor. In the remaining ten videos, 12 instances of the TFSs are found in another 3 supervisions. These 12 instances are then used to verify the findings of the case study. It is acknowledged that the dataset is narrow due to the limited size of the multimodal corpus; however, even through preliminary observation, patterns of gesture uses already emerge in this small dataset. Moreover, as the goal of the study is to explore a multimodal corpus-based approach for investigating multimodal units of meaning rather than providing generalisations, the relatively small size of the corpus does not limit the research findings. A limited number of instances has been selected to allow for a full exploration of a new approach to multimodal analysis of spoken units of meaning.

The exploration of the temporal coordination between speech and gesture is a common practice for gesture research as it reveals the extent to which the meaningful parts of gestures align with the speech segments of which they are iconic (Kendon, 2004; McNeill, 1992, 2005). For the first study, the temporal analysis is designed to investigate which gesture phases co-occur with the TFSs, what types of movements co-occur with the TFSs when they co-occur with the strokes, and whether the gesture phases and the TFSs are closely aligned. To accomplish this, all the gestures co-occurring with the TFSs are segmented into five phases and transcribed based on the gesture coding scheme and transcription format that are customised for the current thesis. Based on the frequency of occurrence of the emerging gesture phases in line with the TFSs, a general picture of the temporal relationship between the speech TFSs and the gestures can be generated. Such analysis contributes

to the debate concerning where the start and end points of speech and gesture align. It can also imply whether gesture can be a criterion for segmenting formulaic sequences.

For the second research question, exploring the multimodal units of meaning means identifying the emerging speech-gesture functional coordination and then establishing a speech-gesture profile based on the emerging patterns. To do this, a framework for categorising the speech functions of all the TFSs is built up based on the discourse context, and then the number of different gesture phases co-occurring with each type of function is counted. In this way, the researcher can identify what gesture phases predominantly co-occur with which function, i.e. the emerging patterns of speech-gesture functional association. Such findings can suggest whether the gestures synchronised with the formulaic language are as formulaic as the speech counterparts. It can also shed new light on understanding the cognitive link underpinning the meaning association between speech and gesture (refer to Chapter 2.7 for details).

In terms of the third research question concerning the discourse functions of gestures, the thesis intends to build up a comprehensive framework based on the discourse contexts. Previous researchers have discussed some functions of the gestures, and the accepted ones include demarcative, referential, pragmatic, and discourse marking functions (refer to Chapter 2, 2.6. for more information). Demarcative function means that gestures segment the utterances into words, phrases, and utterances as they are aligned with different segments of speech (Calbris, 2011). Referential function refers to those gestures that contribute to the meaning of utterances (Calbris, 2011; Kendon, 2004; McNeill, 1985, 2005). Pragmatic function means that gesture can perform functions such as offering, requesting, etc. (Kendon, 1995, 2004). The discourse marking function refers to gestures that can foreground the noteworthy and/or new information in the utterances (McNeill, 2005). Until

now, there is a lack of research that describes the contributions of gestures in discourse systematically and at scale.

Hence, for the first study, the focus of the functional analysis of gestures is on their role in discourse with reference to the TFSs. Relevant aspects include whether they contribute to the utterance meaning, whether they segment utterances into units, whether they mark discourse relationships, whether they contribute to the coherence of interactions, and whether they help enhance inter-subjectivity between participants. The findings as such can be a valuable contribution to the debate on whether gestures are primarily for communication (Alibali, Heath, & Myers, 2001; Bavelas, Gerwing, Sutton, & Prevost, 2008) or speech production (Beattie & Aboudan, 1994; Butterworth & Beattie, 1978; Hadar & Butterworth, 1997). They can also contribute to the knowledge of the interactive nature of gestures (Bavelas, Chovil, Lawrie, & Wade, 1992).

1.2.3. The second study

The second study intends to explore the multimodal units of meaning of the circular gestures, and, in accordance with the first study, the research questions are as follows:

1) What is the temporal relationship between the circular gestures and speech?
2) What is the unit of meaning of the circular gestures in accordance with their co-occurring speech?
3) What are the discourse functions served by the circular gestures?

The second study also draws on the supervision meetings in the Nottingham Multimodal Corpus and 570 instances of circular gestures emerge from 8 videos with approximately 100,000 running words. Although

the primary goal is to explore a multimodal corpus-based approach for examining the unit of meaning of a recurrent gesture, this is by far the largest dataset used to investigate the meaning variations of a recurrent gesture. To explore the questions, all the circular gestures are segmented in ELAN and all the speech components co-occurring with them are also annotated. The study is also exploratory in terms of using monomodal corpus tools in multimodal research. Hence, in order to allow automatic word processing, all the gestures are annotated in plain texts using 'CG# #' with the co-occurring speech lying in the middle of the two hashtags. The abbreviation of CG can then be used as the node, or the searched item, during automatic textual search and analysis. For textual analysis, AntConc is chosen as a valuable and free tool that is sufficient for the analytical needs of the current study.

For the first research question, i.e. the temporal analysis between speech and gesture, it intends to discover the frequency of the different lengths of speech co-occurring with the circular gestures, i.e. whether they tend to be single words, or multiword expressions, or even clauses. To do this, all the speech parts co-occurring with the circular gestures are manually counted and analysed in Excel. Furthermore, the temporal analysis also examines the tendency of the location of the circular gestures in utterances, i.e. whether they tend to occur at the beginning, middle or end of utterances. For this, all the annotated text files are processed in AntConc for cluster analysis, i.e. sequencing the frequency of the words to the right of CG. By observing the most frequently co-occurring lexical items to the right of CG, the tendency of the positions of the circular gestures can be efficiently examined.

The key for building up the multimodal unit meaning of the circular gestures is identifying the emerging lexical and grammatical patterns in the co-occurring speech. To do this, five instances are chosen in sequence each time from each video for pattern analysis (40 instances for each round) and a lexical and grammatical framework is then established when no new patterns

can be identified. This sampling process is repeated three times (120 instances involved in this pilot). The concordance list of CG is also useful for spotting patterns as all the instances are listed in alphabetic order, and, with its assistance, minor revisions are made to the framework established during the sampling processes. The final framework is then used to categorise all the instances in the concordance list. To describe the relationship between the circular gestures and the emerging speech patterns, three of Sinclair's (2004) concepts in the multimodal unit of meaning framework, i.e. *colligation*, *semantic preference* and *semantic prosody*, are used here. The colligation category includes all the grammatical patterns, semantic preference refers to all the semantic patterns and semantic prosody entails all the functional patterns. After completing this, the multimodal unit of meaning of the circular gestures can be built up.

For the final research question, the discourse functions of the circular gestures are analysed in accordance with the speech co-occurring with them. The functional framework developed in the first study based on the gestures co-occurring with the TFSs is also used for describing the circular gestures.

1.3. Thesis overview

In terms of the overall structure of the thesis, a literature review (Chapter 2) of the existing theoretical and methodological frameworks relevant to the current thesis follows the present chapter. This involves a review of the current relevant research on multimodality, corpus linguistics, formulaic sequences, and gesture studies. Chapter 3 introduces the Nottingham Multimodal Corpus and rationale for using it as the data source for the thesis. It also details the development of a gesture coding scheme for segmenting all the gestures in both studies, following which is a description of the gesture transcription format designed for the thesis.

Chapter 4 and Chapter 5 report the approaches, analyses, findings and implications of the first and second study of the thesis, respectively. The three research questions for each study are addressed, and the multimodal units of meaning of the TFSs and the circular gestures are demonstrated in each chapter. Chapter 6 provides a discussion of the implications of the current thesis for the existing knowledge and debates in gesture studies. The discussion is unfolded from three perspectives in accordance with the research questions, i.e. the temporal relationship between speech and gesture, the meaning coordination between speech and gesture, and the functions of gestures. As the final chapter, Chapter 7 first revisits briefly all the research questions, findings and implications, and then proposes areas for future research. Implications and applications of the approaches and findings of the thesis for the broad field of applied linguistics are also outlined.

Chapter 2. Literature review

2.1. Introduction

This chapter aims to lay the theoretical and methodical foundations for the thesis, presenting the most relevant theories, concepts, analytical frameworks, previous research, major issues and debates. It begins with an overview of the field of multimodality in 2.2., clarifying the major research strands under this broad concept. It also specifies where the current thesis, a multimodal corpus-based research, sits in the field of multimodality and how it differs from others theoretically and methodologically. In addition, a demonstration and evaluation of the existing written corpora and multimodal corpora will also be included in this section, pointing out the noticeable lack of multimodal corpora that can be used for multimodal linguistic research.

Section 2.3. mainly concerns defining formulaic sequences, which forms the main research target for the first study, discussing two issues of major relevance, including the criteria for segmenting formulaic sequences and the processing of them in the mind. As gesture is the primary research interest, the next section 2.4. will first provide a specific definition, and then introduce the key categorisation and analytical frameworks that are relevant to the thesis. The latter part of 2.4. will contain some of the most significant research and debates regarding the temporal and semantic relationship between speech and gesture, followed by a brief mention regarding the challenges of analysing multimodal data and a review of existing multimodal tools.

As the second study of the thesis focuses on exploring the speech patterns co-occurring with a recurrent gesture, the circular gesture, section 2.5. will concentrate on reviewing previous research on the patterned uses of gestures of high occurrence from different theoretical perspectives. The research on the cyclic gesture that is most relevant to the second study will also be introduced

in detail in 2.5. Research on frequently occurring gestures not only indicates the possibility of identifying patterns for the circular gesture, but also provides valuable theoretical and methodological insights for the thesis.

A general introduction of the existing discussions of the functions of gestures from different perspectives will be presented in 2.6., following which is another vital issue concerning theories that can provide cognitive justifications for form-meaning associations in recurrent gestures. The final section will demonstrate previous studies that work towards the concept of the multimodal units of meaning from different theoretical backgrounds, with a primary focus on demonstrating two multimodal corpus-based investigations, i.e. multimodal construction grammar, and head nods and response tokens.

2.2. Multimodality

The first goal of this section (in 2.2.1) aims to define multimodality, describe the origin of the field and summarise the two largest strands of research of this nature in linguistics. These are written discourse analysis from the perspective of social semiotics, and gesture studies from numerous approaches. The current study follows a multimodal corpus linguistics approach as a recently emergent research paradigm in applied linguistics. The studies taking a multimodal corpus approach mostly draw on multimodal spoken corpora to investigate the interplay between speech and gesture. The second goal is to offer an overview of the characteristics of the existing monomodal/textual and multimodal/audio-visual corpora, comparing and evaluating them from the perspectives of representativeness, size, transcription and annotation. It will reveal that, whereas there is a sufficient number of balanced, large-size and publicly available monomodal corpora, there is a noticeable lack of such corpora for multimodal linguistic research. These four indexes are chosen here based on the four connotations of the word ‘corpus’, i.e. sampling and

representativeness, finite size, machine-readable form and a standard reference, that are mostly accepted by the corpus linguistics community (McEnery & Wilson, 1996).

2.2.1. Definition, origin and strands of research

Multimodality, as an area and an approach, celebrates inter-disciplinary research and has the potential of embracing a wide range of research interests in the broad field of linguistics and applied linguistics. The fundamental difference of taking a multimodal approach from a word-centred one is that multimodality not only investigates written/spoken words in meaning construction, but also other modes such as gesture, image, facial expressions, sound, etc. (Jewitt, 2013). That is, each mode is an indispensable part of a multimodal ensemble, and multimodality is interested in how and why people in a certain social context assemble a range of modes to represent meaning. Knight (2011) provides a more specific definition and underlines that multimodality is the investigation of the processes of meaning generation through the construction of different modes as well as their media, i.e. the physical representation of the modes (e.g. physical hand-and-arm motion is the media of gesture). That is, in examining one mode, such as gesture, the media that realise the gesture (the movements of the hand and arm) must inevitably be examined.

Multimodality originates from the field of social semiotics, i.e. the study of meaning-making processes of language and other semiotic systems in certain social and cultural contexts of communication (Halliday, 1978). Michael Halliday describes language as a social semiotic system in his pioneering work, *Language as Social Semiotic* (1978). Instead of seeing languages as static, he emphasises their social and cultural nature, i.e. how they are shaped by the ways in which people use them and their functions in

various contexts of communication. This stance is crucial for multimodality as it sees language as a choice made by users to represent and communicate meaning, a process that can be accomplished by semiotic systems other than language.

Following Halliday, scholars such as Kress and van Leeuwen in *Reading Images* (1996) extended their social semiotic research from words to other visual/textual modes, such as image, which lays the foundation for multimodality. They have shown that other modes like image can also deliver the three aspects of Halliday's meaning functions, which are ideational meaning (constructing contents), interpersonal meaning (managing social relationships) and textual meaning (creating coherence) (Halliday, 1978). These three aspects are important criteria for evaluating a mode, though what constitutes a mode is still a controversial issue. The most widely acknowledged modes include writing, image, sound, speech, gesture, gaze and posture, and linguistic terms such as 'grammar', 'coherence', etc. are transferred to describe these modes such as image (Jewitt, 2013).

A survey of the most recent published books reveals that, until now, the majority of multimodal research in the field of social semiotics predominantly pays attention to texts/words and image, providing a rich description of the meaning-making process. For instance, David Machin's book, *Introduction to Multimodal Analysis* (2016), has a sole focus on the multimodal analysis of image, typology and page layout in relation to semiotics and visual grammar. Similar works include two edited books entitled *New Discourse on Language: Functional Perspectives on Multimodality, Identity, and Affiliation* (Bednarek & Martin, 2010) and *Interactions, Images and Texts: A Reader in Multimodality* (Norris & Maier, 2014). In a few books, such as *Multimodality, Learning and Communication: A Social Semiotic Frame* (Bezemer & Kress, 2016), multimodal spoken face-to-face communication has, indeed, occupied much more space compared to many other published works in the field;

however, the scope of datasets is very limited in the sense that they focus on providing thick descriptions of some examples of communication in various real-life contexts (e.g. classroom interaction, operating theatre, job interviews, etc.).

Whereas multimodal research in social semiotics primarily focuses on the detailed analysis of the meaning-making processes of texts and image in both printed and online media, there have been a great number of multimodal studies on speech and gesture in spoken language. Works as such cover a wide range of theoretical and methodological backgrounds, including pragmatics (Kendon, 2004), conversation analysis (Goodwin, 2000, 2007; Tsuchiya, 2013), first and second language acquisition (Gullberg, 2006, 2011; Kita & Özyürek, 2003; Mayberry & Nicoladis, 2000), gesture and metaphor (Cienki & Müller, 2008a; Forceville & Urios-Aparisi, 2009; Mittelberg, 2008), bilingual studies (Gullberg, 2010; Nicoladis, Pika, Yin, & Marentette, 2007), psycholinguistics (McNeill, 1992, 2005), and gesture and grammar (Fricke, 2013; Harrison, 2013; Müller, Bressem, & Ladewig, 2013). Among these approaches, research in language acquisition, bilingualism, and psycholinguistics follow the tradition of psychology and rely heavily on quantitative methods, aiming to discover the numerical differences in gesture uses between separate groups of participants. They draw on data collected from labs, assigning controlled tasks to participants and then conducting quantitative analysis. On the other hand, research in pragmatics, conversational analysis, and metaphor attach immense importance to the qualitative analysis of the meaning and function of gestures based on discourse contexts. They tend to provide thick descriptions of a handful of selected episodes of communication in real-life contexts.

In view of the above discussion, previous multimodal research on speech and gesture predominantly either depend on statistics, paying little attention to the discourse context, or on qualitative analysis, lacking scope in data.

However, multimodal corpus linguists develop a paradigm that differs from most existing gesture studies as they attempt to combine both qualitative and quantitative approaches (Adolphs, 2008; Adolphs & Carter, 2013; Knight, 2011). They interpret the meaning variations of gestures through discourse analytical methods and count the frequency of occurrences of these variations, or emerging patterns of behaviour, following corpus-based methods. In addition, as the corpus linguistic approach uses authentic, naturalistic data to assist linguistic investigation (Sinclair, 1991), multimodal corpus linguistic researchers think very highly of this principle and strive for collecting and drawing on corpora that are as real and natural as possible. In a word, multimodal corpus linguistic research usually draws on a large-scale corpus of real-life spoken language and includes an extensive number of instances in the analysis, in attempt to generate empirical evidence for certain research queries.

Notable examples of multimodal corpus-based research in the area of speech and gesture include the studies on the functional coordination between head nods and response tokens (Knight, 2011), which annotated and analysed a substantial number of instances of head movements and the co-occurring speech from the Nottingham Multimodal Corpus. Based on a large global news broadcast (the Red Hen corpus), another study also selected and examined a huge number of examples, exploring the multimodal constructions of gesture and speech in four motion events (*circles*, *spin around*, *zigzag* and *all the way from X PREP Y*), i.e. the percentage of each event accompanied by gestures) (Zima, 2014). Similarly, drawing on a large multimodal corpus, the Bielefeld Speech and Gesture Alignment Corpus, Kasper Kok (2017) examined the statistical preferences of words and grammatical categories for being accompanied by gesture. Taking a more fine-grained approach to gesture forms, Ladewig (2014b) classified patterns of form-meaning variations of cyclic gestures (i.e. rotational movements) based on more than

400 instances from a German corpus. Due to the great relevance of the multimodal corpus-based research on speech and gesture relationship to this thesis, most of the aforementioned research will be demonstrated in more detail in the remainder of the chapter.

2.2.2. Corpora: monomodal vs. multimodal

Although the original meaning of the word ‘corpus’ means ‘body’, referring to a collection of texts of any size, it has gained certain connotations with the advent of corpus linguistics (McEnery & Wilson, 1996). These connotations include sampling and representativeness, finite size, machine-readable form and a standard reference. However, there is a high degree of acceptance nowadays that these facets are more than connotations; rather, they are criteria by which to classify corpora and judge their quality and usefulness. The following sub-sections will discuss the issues regarding the four facets of the criteria, and, in accordance with the four aspects, describe and compare the state of the art of the current monomodal and multimodal corpora in terms of their representativeness, size, transcription and annotation.

2.2.2.1. Representativeness

Sampling and representativeness means the building of a corpus should be based on certain sampling techniques to make it representative of a language variety of interest (McEnery & Wilson, 1996). To achieve this, the selection scheme should take into account both demographics (i.e. the statistical characteristics, or sociocultural variables, of human populations such as gender, age, income, educational background, etc.) and genres (e.g. fictions, casual conversations, political speeches, etc.), and the scheme of a corpus reflects the type of linguistic research that the corpus can properly support

(Adolphs, 2008). For instance, a corpus of casual conversation will not be able to assist research on formal speech. In addition, for a corpus to truly represent a language variety, corpus balance (i.e. approximately equal quantities of data in each variable in the selection scheme) is required (Hunston, 2002). However, delimiting demographic factors and genres itself is not so straightforward, and, until now, 'no definite list of genres exist[s]' (Hunston, 2002, p. 29). As pointed out by Kennedy (1998), 'The notions of representativeness and balance are, of course, in the final analysis, matters of judgement and can only be approximate' (p. 62).

A corpus can be general, selecting samples from different contexts, thus opening up a wide range of research opportunities, or it can be specific/specialised, designed to explore particular linguistic features in a certain context (Sinclair, 1991). From a different perspective, Hunston (2002, p. 30) makes a fair point that 'the real question as regards representativeness is how the balance of a corpus should be taken into account when interpreting data from that corpus'. This is the common practice for most corpus-based studies as the particular high/low occurrence of some linguistic feature(s) usually pertain(s) closely to high/low proportion of a certain genre in a corpus.

There are an extensive number of balanced written corpora available that can support textual analysis of different varieties of English languages (go to http://www.lancaster.ac.uk/fass/projects/corpus/cbls/corpora.asp for a detailed description of the existing monomodal corpora). They include multi-million general corpora such as the 100 million word British National Corpus (BNC, 10% spoken), representing British English, and the 560 million word Corpus of Contemporary American English (COCA, 118 million words spoken), representing American English (updated each year). The International Corpus of English (ICE) comprises collections of spoken and written English samples from different English varieties (broadly classified by countries), including Canada, East Africa, Great Britain, Hong Kong, India,

Ireland & SPICE Ireland Jamaica, New Zealand, Nigeria (written), the Philippines Singapore Sri Lanka (written), USA (written). The 4 billion word Bank of English corpus is a subset of one of the largest corpora in the world, COBUILD, with 450 million words (spoken and written) released to the public.

As written corpora are considerably easier to compile compared to spoken ones, monomodal spoken corpora are much modest in size; however, they include a relatively balanced sample of participants (usually selected by age, sex, social class, region, etc.) and represent an extensive variety of genres (casual conversations, meetings, lectures, telephone conversations, radio broadcasts, etc.). They are great assets for research on spoken language. Examples include the 118 million word spoken language from COCA, the 10 million word spoken part of the BNC, the 5 million word Cambridge and Nottingham Corpus of Discourse in English (CANCODE), the half million word London-Lund Corpus (LLC), and the 249,000-word Santa Barbara Corpus of Spoken American English (Parts 1-4 have been released), among many others.

There are also many written and spoken corpora that are representative of a specific genre. Examples include the 3 million word Switchboard Corpus (telephone conversations), the 1.2 million word Zurich English Newspaper Corpus (ZEN, newspapers texts in Early Modern English), the 1 million word Michigan Corpus of Academic Spoken English (academic English), the 1.6 million word British Academic Spoken English corpus, the half million word Bergen Corpus of London Teenage Language, the 2.5 million word International Corpus of Learner English, and the Hong Kong Corpus of Spoken English (four discourse types: conversation, business, academic and public).

Whereas most monomodal written and spoken corpora are systematically sampled and balanced corpora, representing language occurring in different

contexts of communication by a large number of speakers from all walks of life, however, this is not the case for most existing multimodal corpora (see Table 2.1 for an overview of existing multimodal corpora) (Knight, 2011). Apart from a few exceptions such as the Göeborg Spoken Language Corpus, the majority of the video recordings were collected from a very limited number of speakers in lab-based contexts with controlled tasks. For example, the MM4 Audio-Visual Corpus filmed four people interacting in a controlled meeting room in the lab with predetermined topics and contents. Similar instances include the Mission Survival Corpus 1 (MSC 1), the CUBE-G Corpus and the UTEP ICT Corpus, among many others.

In addition, the multimodal corpora are predominantly restricted to one specific context rather than collecting samples from different genres. One common setting is the meeting room such as the AMI Meeting Corpus, the MM4 Audio-Visual Corpus, the NIST Meeting Room Phase II Corpus, and the VACE Multimodal Meeting Corpus.

Table 2.1 An overview of existing multimodal corpora (Knight, 2011)

Name, Reference(s), Language	**Size, Composition and Additional Information**
AMI Meeting Corpus. Ashby et al., 2005. Corpus of European English.	100 hours of recordings taken from 3 different meeting rooms. This corpus was created for the use ‘of a consortium that is developing meeting browsing technology’.
CID (Corpus of Interactional Data). Bertrand et al., 2006; Blache et al., 2008. Corpus of French.	8 hours of dyadic conversations, comprised of 2 participants sat in close proximity of one another, each wearing a microphone headset. Participants were encouraged to chat informally, so with no directions on how to structure the talk.
CUBE-G corpus. Rehm et al., 2008. Corpus of Japanese and German.	Dyadic conversations involving Japanese and German speakers (this is a cross cultural corpus). One participant is an actor, whose contributions are scripted/instructed.

Czech Audio-Visual Speech Corpus/Corpus for recognition with Impaired Conditions. Železńy et al., 2006; Trojanová et al., 2008. Corpus of Czechoslovakian.	Developed to test and train the 'Czech audio-visual speech recognition system' (automatic speech recognition). The first corpus features 25 hours of audio-visual records from 65 speakers. The second has 20 hours of data across 50 speakers. In both each speaker was instructed to read 200 sentences, in laboratory conditions (50 common sentences; 150 were speaker specific).
D64 Corpus. Campbell, 2009. Corpus of British English.	4-5 people recorded over two 4 hour sessions across two days. Non-directed and spontaneous conversations in a domestic environment. Participants wore reflective sticky markers to track movement.
Fruits Cart Corpus. Aist et al., 2006. Corpus of American English.	104 videos of 13 participants (4-8 minutes each). Approximately 4000 utterances in total. Comprised of task-orientated dialogues in an academic setting. Designed to explore language comprehension, now used to analyse language production (NLP).
Göeborg Spoken Language Corpus. Allwood et al., 2000. Corpus of Swedish.	Small components of this 1.2 million word spoken language corpus have been aligned with video records. This corpus contains conversations from different social contexts with a range of different speakers talking spontaneously (i.e. non-directed or scripted).
IFADV Corpus. Van Son et al., 2008. Corpus of Dutch.	A free dialog video corpus composed of face-to-face interaction between close friends/ colleagues. This corpus is comprised of twenty 15 minute conversations (5 hours in total).
MIBL Corpus (Multimodal Instruction Based Learning). Wolf and Bugmann, 2006. Corpus of British English.	Human-to-human instruction dialogues, with one participant teaching a card game to the other. This corpus links speech to movement on the screens and is used to train service robots.

Mission Survival Corpus 1 (MSC 1). Mana et al., 2007. Corpus of Italian.	A meeting corpus which includes a range of short meetings, with up to 6 participants in each. The topics and tasks covered in the meetings are controlled but not scripted.
MM4 Audio-Visual Corpus. McCowan et al., 2003. Corpus of European English	Features 29 short meetings between 4 people filmed in controlled, experimental conditions. The majority of the meetings were scripted and cover specific, predetermined topics and tasks.
NIST Meeting Room Phase II Corpus. Garofolo et al., 2004. Corpus of American English.	Part of the NIST MDCL (Meeting Data Collection Laboratory). This corpus contains 15 hours of recordings from 19 meetings; comprised of both scenario-driven meetings and 'real' meetings.
NMMC (Nottingham Multimodal Corpus). Knight et al., 2009. Corpus of British English.	250,000 words; 50% single speaker lectures, 50% dyadic academic supervisions. Sessions were video and audio recorded, transcribed and aligned using DRS (the Digital Replay System).
SaGA (Bielefeld Speech and Gesture Alignment Corpus). Lücking et al., 2010. Corpus of German.	This corpus contains 280 minutes of audio/video recorded data comprised of 25 direction giving and sight description dialogues based in a Virtual Reality environment. The data is considered to be 'naturalistic' as content is spontaneous, though controlled/prompted.
SK-P 2.0 SmartKom multimodal Corpus. Schiel et al., 2002. Corpus of German.	This corpus contains 96 different single users who were video/audio recorded across 172 sessions. Each user carried out specific, prompted tasks and was recorded in public spaces such as cinemas and restaurants. This corpus is effectively HCI based.
SmartWeb Video Corpus (SVC). Schiel and Mögele, 2008. Corpus of German.	99 recordings of human-human-machine dialogue, with 1 speaker interacting with a human person and a dialogue system (i.e. the main participant is using a Smartphone, which records their face and they are talking to the other participant).

UTEP ICT. Herrera et al., 2010. Corpus of American English.	Cross cultural corpus involving task-based conversations between groups of 4 participants who are stood in a room and free to move around. 200 minutes of data.
VACE Multimodal Meeting Corpus. Chen et al., 2005. Corpus of American English.	Comprised of meeting room 'planning sessions'. Spontaneous talk in controlled task-based environments. 5 participants in each scenario, with 5 scenarios recorded.

The narrow representation of most multimodal corpora is mainly relevant to the aims of these, which is to enhance computational technology, especially for the purposes of audio-visual recognition (e.g. the Fruits Cart Corpus, MM4 Audio-Visual Corpus, NIST Meeting Room Phase II Corpus, etc.) and human-computer interaction (HCI) (e.g. MIBL Corpus, SmartKom Multimodal Corpus, etc.). Findings based on such corpora have limited implications for language use for the population at large and for other contexts. In addition to the research purpose, the features of the lab environment (e.g. quiet, good lighting, etc.) is conducive to setting up audio and visual recording equipment and capturing high-quality videos. There is a distinctive lack of recording devices that are well-tailored to the more complex real-world communicative settings (e.g. noise, bad lighting, people moving around, etc.). After all, it is easier to obtain high-quality video recordings in a controlled environment where participants are situated in the same position. Another factor that might cause the narrow representation of participants is the challenge of recruitment as it is easier to reveal private information in videos.

2.2.2.2. Size

The second connotation, finite size, refers to the planned number of words that a corpus is aiming to reach, determined at the onset of collection. It should be noted, however, that the intended target size may or may not be reached in the

end. The exception to this is monitor corpora (e.g. the COBUILD corpus), which intend to keep growing and monitoring language change (McEnery & Wilson, 1996). This connotation closely pertains to the controversial issue of the size of a corpus. Sinclair's (1991) recommendation is to make it as large as possible as the frequency of words (e.g. *the* is much more frequent than *school*) and their variations (*school* used as a noun and a verb) can be greatly imbalanced. Even mega-corpora will generate words that only occur once. Reppen (2010, p. 32) suggests considering the 'the representativeness (have I collected enough texts (words) to accurately represent the type of language under investigation?) and practicality (time constrains)'. Hence, to determine whether a corpus is large enough, one needs to ask, on the one hand, whether enough number of words have been collected in each sample in each category in each genre (Kennedy, 1998). One will also have to look at whether enough encounters of the research target have been reached in order to have robust evidence to support the findings. A small corpus may be sufficient for research on a very frequently occurring lexical item in a very specific context (e.g. investigating a specific item such as 'pronouns' in political speeches may not need a multi-million word corpus); however, a researcher aiming for the overall use of an item will need a much bigger and more general corpus (Adolphs, 2006).

In addition to representativeness, Meyer (2002) also underlines certain practical factors such as the financial, human and technological resources available to a project, and the time allowed for collecting, computerising and annotating them. Hence, spoken corpora are usually much smaller than written ones as they need considerably more time to transcribe and annotate (one hour of speech may generate 12,000 to 15,000 words and take about two days to transcribe depending on the granularity of transcription) (O'Keeffe, McCarthy, & Carter, 2007). Any spoken corpus above one million words is large, whereas any written corpus below several million is small.

As discussed in the last section, there is no lack of multi-million mega-size textual corpora available to assist a wide variety of purposes (e.g. compiling dictionaries and teaching materials, conducting linguistic and sociolinguistic research). However, according to the criteria set by textual corpora, the size of most multimodal corpora is not measurable as they have not been transcribed. Examples include the CUBE-G Corpus, the Czech Audio-Visual Speech Corpus, the D64 Corpus, the Fruits Cart Corpus, the Mission Survival Corpus 1, the MM4 Audio-Visual Corpus, the NIST Meeting Room, the Phase II Corpus, and the UTEP ICT. Moreover, many multimodal corpora are not large in terms of the length of the videos as most only have a few hours of video recordings, except for the AMI Meeting Corpus (100 hours), the Göeborg Spoken Language Corpus (1.2 million words), and the Nottingham Multimodal Corpus (more than 20 videos, 250,000 words). The current situation concerning the small size of most existing multimodal corpora is, again, relevant to their research purposes. They were mainly designed for developing computational technologies (e.g. machine training) rather than assisting in exploring linguistic queries. Also, like compiling spoken corpora, transcribing multimodal corpora is enormous in terms of time and money; even more so if other modes of communication (e.g. gesture, gaze) require to be annotated and transcribed.

Similar to monomodal corpora, there is no available standard of how ‘big’ a large multimodal corpus is; however, like the principle applied to monomodal corpora, the more frequent the research target is, the smaller the multimodal corpora need to be (Adolphs, 2008). With particular reference to gesture, for frequently occurring gestures like head nods, one hour can extract thousands of instances and, therefore, most research on head nods might not need a particularly large multimodal corpus (Knight, 2011). For less frequently occurring gestures such as the circular gestures (rotational movements of the hand), less than one thousand instances have been found in

10-hour videos (Chen & Adolphs, 2017), and a large corpus is required. Generally speaking, a large multimodal corpus will be needed to conduct research on gesture. In addition to the requirement to consider the extraneous variables, bear in mind that certain speakers just do not gesture frequently. The context and task of speaking can also affect the number of gestures. For instance, the increase of physical activities such as writing, drawing and typing may reduce talking and gesturing.

2.2.2.3. Machine-readable

That a corpus should be machine-readable basically means that it is in the format of electronic texts, which most corpus annotation and analytical software can readily process (McEnery & Wilson, 1996). Hence, online electronic texts will need to be cleaned (as they may involve codes, hyperlinks and images) and transformed into plain texts, which is the format that most word processing software works well with. Printed texts will need to be digitised and spoken corpora transcribed into electronic texts. These processes not only consume an enormous amount of time, but also affect the integrity of meaning delivered through the original media format, mainly due to the fact that many components (e.g. image, prosody and bodily movements) cannot be transformed into texts (Adolphs, 2006). Nevertheless, digitisation and transcription are necessary and inevitable processes in most circumstances at this stage of technological development.

Most corpora will also need to go through the process of annotation to enable further automatic retrieval and analysis (McEnery & Wilson, 1996). Annotation is the process of adding extra information to the texts at different linguistic levels (e.g. words, phrase). The most commonly encountered sub-categories of annotation, or mark-up, are tagging and parsing: the former particularly refers to adding a code to each word in a corpus to specify the

part-of-speech information (known as POS tagging), the latter specifies grammatical, syntactic structure (Hunston, 2002). Tagging and parsing can be done automatically (which is the most efficient and prevalent way, and has a reasonable rate of accuracy), or manually (which is extremely time-consuming but with higher accuracy). The software used for automatic tagging and parsing is called a tagger and parser, respectively. In addition to tagging and parsing, there are numerous types of mark-up, including semantic annotation (word relationships or meanings), discourse annotation (discourse management elements e.g. hedging, politeness, anaphora annotation), prosodic annotation (e.g. intonation), and nonverbal annotation (e.g. laughter), etc. Researchers usually design their own mark-up schemes in accordance with their research purposes, and an annotated corpus is available for various automatic quantitative analyses of linguistic features, adding much value to the corpus compared with an unannotated one.

An issue that is closely associated with annotation is the standardisation of textual mark-up information among different corpora, which is crucial for enhancing the interchangeability of machine-readable texts (McEnery & Wilson, 1996). Efforts in this direction include the establishment of several standardised encoding systems including the Text Encoding Initiative (TEI) used to annotate the entire British National Corpus (BNC) (Sperberg-McQueen & Burnard, 1995), the Corpus Encoding Standard (CES) applied in corpora of European languages and beyond, and XML (Extensible Markup Language). All three standards are SGML-conformant (Standard Generalized Markup Language), usually composed of the basic structure of two angle-bracketed tags, a start tag and an end tag (e.g. <p> marks-up the start of a paragraph and <p/> marks-up the end of the paragraph). All these mark-up standards provide annotation frameworks, or metalanguage, enabling different projects to flexibly apply to their own corpora, and XML, as the most recent and popular one, has been incorporated into both TEI and CES systems.

The increasing proliferation of XML is, on the one hand, due to its advantage of being used on web pages, which tend to be the major medium of corpus distribution (XML is mainly used for web design) (Meyer, 2002). On the other, it provides the option of separating annotations from the texts so that the users have the choice of hiding annotations. This feature contrasts XML from others and enhances readability, especially for those heavily annotated corpora (McEnery & Wilson, 1996).

Almost all the written corpora available are in electronic form and have been annotated with at least the basic textual information such as structural annotation (e.g. headings, paragraphs, etc.) and part-of-speech annotation. The International Corpus of English (ICE) is annotated with textual structure, parts of speech and syntax (i.e. sentence structure is represented by a parse tree). The American National Corpus has structural mark-up (e.g. sections, chapters, paragraphs), sentence boundaries, part of speech tagging, lemma (i.e. the canonical form of a set of words, the lemma of *work* includes *works*, *worked*, *working*), noun chunks, verb chunks, and named entities (e.g. person, location, organisation, date, etc.). Moreover, with all the electronic texts online, it is easier than ever for researchers to develop and annotate their own corpora. For instance, the Sketch Engine[2] computer programme has the function of automatically extracting texts on the web by typing in certain key words/topic (e.g. global warming). Many tools are also available for automatic annotation, such as the widely used CLAWS Part-of-Speech (POS) tagger (Fligelstone, Pacey, & Rayson, 1997; Garside & Smith, 1997), USAS semantic tagger (Garside & Rayson, 1997), and Wmatrix (both CLAWS and USAS can be accessed through the Wmatrix web interface) (Rayson, 2008).

All the monomodal spoken corpora have been transcribed and annotated at various levels of granularity in prosody (e.g. intonation) and discourse (e.g.

[2] https://www.sketchengine.eu/

speaker turns, overlaps). For instance, the London-Lund Corpus of Spoken English (LLC) and the Santa Barbara Corpus of Spoken American English (SBCSAE) provide prosodic annotation (audio available in both), as do the Intonational Variation in English (IViE) corpus and the Hong Kong Corpus of Spoken English (HKCSE) (audio not available). The SEC and the Switchboard corpus have been most comprehensively annotated with details about intonation (including markings of tone unit boundaries, stress placement and tone choice) and rhythm (audio available in both). The spoken component of the International Corpus of English has been marked-up for pauses, overlaps, hesitations, false starts and speaker turns.

It has to be mentioned that, for most researchers, manual annotation is an indispensable and crucial part of analysis as existing corpora do not usually generate satisfactory results (e.g. they contain irrelevant instances) or have the level of annotation required by each study (Smith, Hoffmann, & Rayson, 2008). Moreover, current annotation tools do not allow for the automatic mark-up of the types of annotations mentioned above (e.g. discourse, prosodic and nonverbal annotations). Hence, as mentioned above, researchers will usually have to develop their own mark-up schemes according to their own research questions and annotate manually in corpus tools (e.g. WordSmith) or other software (e.g. Excel), which will allow for further qualitative analysis to be conducted (refer to Smith, Hoffmann, & Rayson, 2008 for a thorough discussion of manual annotation).

Unlike the well-developed monomodal corpora, many multimodal corpora have not been transcribed or annotated. When they are, they tend to have very different annotation schemes with differing levels of detail. For instance, the IFADV Corpus has rich components of textual and prosodic annotation (including orthographic transliteration, POS tagging, word alignment, phoneme alignment, conversational function, etc.), but it only contains gaze direction as a nonverbal element in the scheme. The Mission

Survival Corpus 1, on the contrary, includes no textual annotation, but has marked-up speech rate, pitch and energy, and head/hand/body fidgeting. Whereas the SmartWeb Video Corpus has textual and prosodic information, the UTEP ICT Corpus has not been transcribed, but contains annotations concerning proxemics (i.e. physical distances between participants), gaze and turn-taking. Although the Nottingham Multimodal Corpus and the Göeborg Spoken Language Corpus are among the largest existing multimodal corpora that have been fully transcribed, they have not been nonverbally annotated.

The CID Corpus and the SaGA Corpus have the most comprehensive annotation systems. The CID corpus (with phonetic and orthographic transcription) covers prosody (units, intonation, contours), morphosyntax, syntax, discourse (markers, turns, etc.) and gestures (hands, head, arms). Among all these aspects, the aspect of gesture alone has been annotated in various dimensions including symmetry (both hands symmetrical, both hands asymmetrical), phase (Preparation, Stroke, Hold, Retraction, an analytical framework will be explained in 2.3), gesture types (Adaptor, Iconic, Metaphoric, Deictic, Emblem, definitions will be provided in 2.3), orientation (Palm up, Palm down, Palm towards self, Palm away from self), region (Center center, Center, Periphery, Extreme periphery), coordinates (Right, Left, Upper, Lower, Upper right), contact (Forehead, Hair, Cheek, Chin, Eyes, Eyebrow), trajectory (Upper, Lower, Right, Left, Upper right, lower right), velocity (Normal, Fast, Slow) and amplitude (Small, Medium, large).

The SaGA Corpus (with orthographic transcription) is more focused on gesture annotation compared to discourse information, which is in line with its goal of exploring the relationship between gesture and speech in direction giving and object description contexts. It is marked-up for only a few elements of the discourse context (e.g. interactive goal), but it has a fine-grained gesture annotation scheme, including the gesture phase (stroke), gesture types (deictic,

iconic, discourse), representation techniques (indexing, placing, shaping, drawing, posturing, sizing, counting and hedging), handshape (adopting the American Sign Language lexicon), palm orientation (e.g. forward, backward, left, right, up, down, up/right/forward, etc.), the orientation of the back of the hand, wrist position, movement direction, and movement trajectory.

There has been no standardised annotation system for gesture; however, in view of the similarities in the annotation schemes in the corpora such as the CID and SaGA corpus, it seems possible to develop one scheme that contains the most basic elements such as the gesture phases and gesture types. Both frameworks for annotating gesture phases and types adopted by CID and SaGA corpora are well-known in gesture studies and more frameworks as such (e.g. the McNeill's 1992 gesture space framework) can be included in a standardised gesture annotation scheme. The same principle applies to other modes such as face, eye gaze, eye brows, etc. For instance, Ekman and Friesen's (1969) facial scheme has been widely used in annotating facial expressions. As researchers continue to explore the best practices for developing multimodal corpora, standardised annotation schemes will be constructed which could be adopted by different projects, hence, enhancing comparability and transferability.

2.2.2.4. A standard reference

The last connotation of the meaning of a corpus, a standard reference, indicates the availability of a corpus of a language variety to a wide, international audience, acting as a common dataset to support all types of linguistic research (McEnery & Wilson, 1996). This is especially important for controlling the number of variables and increasing the comparability among different studies. High-quality corpora are also frequently used as reference corpora by researchers drawing on different corpora to compare

linguistic features. For instance, the BNC is often used as a reference corpus by other researchers to compare and foreground key linguistic features of their own corpora. Moreover, a reference corpus can also be used as a test bed for the validity and reliability of the findings based on a different corpus. For instance, Carter and McCarthy (1995) verified their findings of spoken language based on the 5 million word spoken corpus CANCODE (The Cambridge and Nottingham Corpus of Discourse in English) in the 10 million word spoken component of BNC.

However, for the majority of the existing multimodal corpora, being a reference corpus is neither the goal at the outset of design, nor an option, as they are not freely available to the public for ethical and/or legal reasons (e.g. requirements of the funding body) (Knight, 2011). One exception is the 100-hour AMI Meeting Corpus, which was released years ago with all the video recordings, transcriptions and annotations. However, users will need to bear in mind the limitations of this corpus. For instance, it is restricted to one genre, meeting communication. Also, whereas a certain part of the meetings in the corpus are 'real' as they were supposed to take place regardless of being recorded or not, many meetings are scenario-driven, though the interactions are not controlled or scripted. Despite all the limitations, the high-quality, transcribed and annotated AMI Meeting Corpus is a valuable asset to the multimodal community, which can be tailored to different research purposes.

Since the corpus development reported in Knight (2011), there have been numerous multimodal corpora built up in the area of computer science with the purpose of enhancing speech, face, gesture and emotion recognition technologies. Examples include the Multimodal Opinion Level Sentiment Intensity dataset (MOSI) (Zadeh, Zellers, Pincus, & Morency, 2016), the ELEA Audio-Visual Synchronized meeting corpus (ELEA AVS) (Sanchez-Cortes, Aran, Jayagopi, Mast, & Gatica-Perez, 2013), the DUEL corpus (a multi-lingual multimodal dialogue corpus for disfluency, exclamations and laughter) (Hough

et al., 2016), the Danish NOMCO corpus (multimodal interaction in first acquaintance conversations) (Paggio & Navarretta, 2017), the NoXi database (multimodal recordings of mediated novice-expert interactions) (Cafaro et al., 2017), etc. Similar to their precedents, these corpora were collected from lab-based contexts, using a variety of sound and motion capturing devices and a variety of verbal and nonverbal annotation systems according to the diverse purposes of each project.

This PhD thesis takes much advantage of the 250,000-word Nottingham Multimodal Corpus, which is one of the most carefully collected and rigorously evaluated large multimodal corpora. A full description of the NMMC will be provided in Chapter 3, 3.2., demonstrating why it is an ideal platform for the two studies to be carried out, as well as describing its limitations.

2.3. Formulaic sequences

As the first study of the thesis concerns building up a speech-gesture profile for formulaic sequences, this section will present some of the issues of most relevance to the current research. It is believed that native speakers have a lexical system in mind that contains a large number of ready-made, lexicalised multiword strings as wholes, which can bypass online grammatical analysis and be retrieved holistically and efficiently in language use (Pawley & Syder, 1983). Wray (2002, p.9) lists numerous terms in the literature that are used to describe the formulaic aspect of language, i.e. strings of words appear to be processed as wholes rather than individual words, and one of the most frequently used terms for those strings is ‘formulaic sequence’ (FS) proposed by Wray herself (2002, 2008). Formulaic sequences include, on the one hand, fixed expressions such as idioms, which contain semantically opaque and syntactically irregular bundles of words that have meanings that are different from their constituent parts, and whose meanings are established by usage or

habit in a certain socio-cultural community (e.g. *by and large*). On the other hand, they can also be semi-fixed, less opaque and more regular with slots for certain lexical items (e.g. *NP be-TENSE sorry to keep-TENSE you waiting*) (Pawley & Syder, 1983). A large quantity of recurring expressions in daily life, such as *for example*, *at the end of the day*, and *in order to*, etc., are all formulaic sequences. Whereas those lexical strings are ubiquitous in human interactions and highly conventionalised (Kuiper, 2004), the question remains whether gestures co-occurring with them are also patterned and conventionalised to some degree. This is one of the major reasons why the first study decided to use formulaic multiword strings as a starting point to explore the speech-gesture profile. A further reason is that, as formulaic language is significant in terms of delivering information (Kuiper, 2004), maintaining interactional relationship (Bardovi-Harlig, 2012; Burdelski & Cook, 2012), signalling individual and group identity (Burdelski & Cook, 2012), and assisting language acquisition (Bannard & Lieven, 2012; Wray, 2008), etc., establishing a speech-gesture profile will illuminate further implications of these important aspects.

Although different approaches have been used for identifying FSs, the current study follows a corpus linguistic approach and a lengthy multimodal expression that is on the top of the frequency list was chosen (refer to Chapter 3 for more details about the target formulaic sequences). Whereas the existing investigations and applications of formulaic language involve a wide range of areas including linguistics, applied linguistics, corpus linguistics, psycholinguistics, psychology, first and second language acquisition, computer science, etc. (Wray, 2002, 2008), the following sub-sections will address several issues that are most relevant to the current research. These are the definition of FSs in 2.3.1., the issues regarding demarcating FSs in 2.3.2. and the studies concerning the processing advantage of FSs as evidence of their holistic mental processing in 2.3.3.

2.3.1. Defining Formulaic Sequences

The working definition of a formulaic sequence is:

> *'a sequence, continuous or discontinous, of words or other meaning elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar (Wray, 2002, p. 9).'*

This working definition is not only theoretically neutral, but also flexible and inclusive, hence widely used in the literature to describe routinised language. The word 'formulaic' indicates the meaning of unity and custom/habit, and 'sequence' suggests at least two discrete meaningful elements that can be words or others. The elements can be linked or not, and they *appear* to be prestored as a whole in memory, but not necessarily. Before the formulaic sequence was proposed, Wray (2008) had coined another term, morpheme equivalent unit (MEU), which was a rather restricted concept that would not work with real data. She strictly defined MEU(s) as units pre-formulated in memory and automatically processed as wholes such as a single morpheme. Such a definition will not be compatible with many formulaic uses of language, and, as Wray (2008) points out, can only describe a subset of formulaic sequences.

Nonetheless, formulaic sequences and Morpheme Equivalent Unit share many features in structure: both refer to word strings composed of at least two single words (the MEU can also be a single word); both can be continuous or discontinuous; the forms of both can vary from fully fixed (e.g. *over the moon*) to partially fixed (e.g. It is + adj. + to + verb); the meanings of both can lie somewhere between untransparent (e.g. *pull one's legs*) and transparent (e.g. *NP be-TENSE sorry to keep-TENSE you waiting*). Admittedly, there is no guarantee that researchers will ever capture the truth of whether human beings

do avail themselves of a system of prefabricated lexical resources or not, and, if so, how the mind processes and retrieves formulaic language. Nevertheless, exploring formulaic language uses and applying them in language teaching and communication software can be of considerable value for L2 language learners and people with a language deficiency, because formulaic language has been found conducive to enhancing fluency due to their processing advantages (Ellis, Simpson-vlach, & Maynard, 2008; Simpson-Vlach & Ellis, 2010; Wray, 2000).

2.3.2. Frequency and formulaic sequences

It is well acknowledged that identifying formulaic sequences can be challenging and little agreement has been reached on the methodological issues regarding the most reliable way of detecting them (Schmitt, 2004; Wray, 2008). Wray (2008) identifies several criteria including frequency, phonology, form idiosyncrasies, spelling, intuition, published lists and corpora, mixed criteria and bespoke designations, and she has further emphasised that each approach has its own limitations. Such a list covers most methods used in different research areas, except for the eye-tracking experiments in psycholinguistics (Schmitt, 2004). The reality is that researchers tend to design and adopt different methods in line with their own theoretical and methodological traditions in their respective areas. Such a discrepancy in research methods makes it impossible to compare and contrast results.

Among all the approaches for extracting formulaic sequences from the lexicon mentioned above, frequency, which has its root in corpus linguistics, is clearly one of the most often used approaches. Corpus linguists often refer to those recurrent multiword expressions as bundles, chunks or clusters (Biber, 2009; Biber et al., 2004; Reppen, Fitzmaurice, & Biber, 2002). The rationale behind the criterion of frequency is that the most often used phrasal units are

more likely to be wholly prestored and retrieved more easily, although research has shown that this might not be the case (Schmitt, Grandage, & Adolphs, 2004). Corpus-based frequency analysis has been widely used to select target formulaic sequences for teaching and for psycholinguistic experiments (Schmitt, 2004).

However, although extracting frequency lists seems efficient and plausible, it is not without certain limitations. Firstly, the debate of how frequent a multiword unit needs to be in order to qualify for idiomaticity could be endless (Schmitt et al., 2004). Also, the high occurrence of some strings is largely attributed to the high occurrence of its components (Ellis et al., 2008). For instance, the high frequency of *what I mean* may due to the high frequency of *I mean*. Moreover, although frequency analysis is objective, human intuition has to come into play during the whole selection process, starting from deciding the length and the frequency threshold of the target word strings to the exclusion of irrelevant ones (Read & Nation, 2004).

Researchers are now fully aware that frequency alone is inadequate for pinpointing formulaic sequences and new methods are advocated (Wray, 2012). For instance, in addition to using concordance software such as Wordsmith and SARA to generate word strings of high frequency, it is suggested that the program *collocate* (Barlow, 2004) should be used to elicit or triangulate the frequency lists by measuring the Mutual Information score (MI). MI refers to the degree of co-occurrence or co-selection of the words in a lexical bundle and has the advantage of extracting more semantically and pragmatically salient frequency lists (Ellis, 2012; Ellis & Simpson-Vlach, 2009; Simpson-Vlach & Ellis, 2010). In addition to MI, other commonly used statistics for measuring co-occurrence and identifying collocations include chi-squared, log-likelihood, t-score and z-score (Gablasova, Brezina, & McEnery, 2017; Hunston, 2002; McEnery & Hardie, 2011).

Owing to the lack of consistency in methodology in formulaic language

research, the idea of adopting multiple criteria seems necessary (Read & Nation, 2004; Wray, 2000). It is against such a methodological background that speech-accompanying gestures, which constitute an integral part of utterances, may also have a role to play in identifying and demarcating formulaic sequences. That is, introducing a nonverbal criterion that goes beyond the speech medium seems rather tempting. If the close temporal, semantic and cognitive relationship between speech and gesture hold water (which will be demonstrated in the rest of the chapter), it is possible that gesture could become a new approach in identifying and demarcating formulaic sequences.

2.3.3. The processing advantage of formulaic sequences

Although there are still numerous unsolved questions concerning the processing of formulaic sequences (Wray, 2012), the processing efficiency of recurrent formulaic sequences of different types (e.g. idioms, collocations and word strings, etc.) have been confirmed by psychological research, though non-native subjects seem to benefit less due to their partial mastery of the target language (Bannard & Lieven, 2012; Conklin & Schmitt, 2012). The efficiency of processing in formulaic sequences is largely relevant to the *holistic processing* mode proposed by Wray (2002), which relies on the pre-formulated lexical strings stored in the lexicon, as opposed to the *analytic processing* mode that relies on grammatical rules to create novel lexical combinations. The holistic processing mode echoes Sinclair's (1991) *idiom principle*, and the analytic processing mode is similar to his *open choice principle*. The former results in the selection of two or more words that usually co-occur in previous uses, and the latter leads to the creative selection of individual words on the basis of grammatical rules. Language users switch between the two modes during language production.

The processing advantage is confirmed in one eye-movement experiment (Underwood, Schmitt, & Galpin, 2004), in which both native and non-native participants fixated less on the terminal words in the formulaic sequence contexts than in the non-formulaic ones. Nevertheless, the non-native subjects do not have the duration advantage compared to the native subjects. That is, the time that the non-native subjects looked at the final words of formulaic sequences was as long as the time they looked at the final words in non-formulaic sequences. In another experiment, using reading times as an indicator of efficiency (Conklin & Schmitt, 2008), both native and non-native subjects showed the processing advantage of formulaic sequences. However, one self-paced reading experiment (Schmitt & Underwood, 2004) does not confirm the processing advantage of formulaic lexical units. Researchers of the study attributed the results to the nature of the reading task in the experiment, which might have lured the participants into word-by-word reading instead of processing the target formulaic sequences holistically.

Nevertheless, it is important to recognise that whether such a processing advantage means the mental storage of formulaic sequences as a whole is still debatable as this efficiency may be a result of familiarity with formulaic sequences rather than pre-storage (Conklin & Schmitt, 2008). Again, the first study of the current thesis can provide some meaningful insights in this respect. That is, if the gestures co-occurring with the formulaic sequences tend to be a complete unit, they will become nonverbal evidence of the holistic mental processing of speech formulaic sequences.

2.4. Gesture: definition, frameworks and analysis

This section will provide a comprehensive introduction of some key theories, concepts and frameworks in gesture studies that relate closely to this thesis, and ends with a review of existing multimodal analytical tools. Section 2.4.1.

will specify different types of bodily movements and define what gesture means throughout the thesis. 2.4.2 and 2.4.3. are mainly concerned with demonstrating some established analytical frameworks for categorising gesture types and segmenting gesture phases. They will be followed by a brief illustration of the close temporal and semantic relationship between speech and gesture based on previous research in 2.4.4. The final section will describe the obstacles of realising automatic search, annotation and analysis of multimodal datasets, and then offer an overview of the most well-known software for assisting multimodal research with a brief introduction of their features.

2.4.1. Defining gestures

Human beings have different types of bodily movements, and it is necessary to distinguish gestures from others. Many previous gesture studies have defined gestures as the improvising movements of the hand and arm (Duncan, 2006; Kita et al., 2007; Kita & Özyürek, 2003; McNeill, 2005). This definition falls onto the gesticulation side of Kendon's Continuum proposed by McNeill (1992) in recognition of Kendon's major contributions to gesture studies. Along this continuum, McNeill divides gestures into four types moving from gesticulations to emblems, to pantomimes and to sign languages. In most published works, whenever gesticulations are the target, researchers usually call them gestures for convenience.

Emblems are conventionalised gestures that have a relatively stable meaning-form mapping across different contexts in a particular socio-cultural group and their meanings can be culturally conditioned (Brookes, 2005). For instance, the ring gesture with the thumb and index finger touching, and the other fingers extended or bended naturally has different dominant meanings in different cultures (Desmond, Collet, Marsh, O'Shaughnessy, & Morris, 1979):

it means ‘OK/good’ in China, but ‘zero’ in France. In general, the meanings of emblems vary geographically, socially and culturally (Brookes, 2005; Kendon, 2004). Emblems are the type of gestures that are known to all, i.e. whenever the notion of ‘gesture’ is mentioned, people usually give an example of an emblem.

Pantomimes or mimes refer to gestures performed to describe a story without speech; **signs** are linguistic components of sigh languages that have their own morphological, semantic, syntactic and grammatical systems like any other linguistic systems. Each sign language has their own system such as British Sign Language, American Sign Language, Spanish Sign Language, etc.

When moving from gesticulation to signs, the dependence on speech decreases. That is, whereas the interpretations of gesticulations mainly rely on the co-occurring speech, signs are entirely independent of speech (McNeill, 1992, 2005). Gesticulations are also believed to be performed more spontaneously and less consciously compared to the other three types. Although the degree of consciousness of gesticulation might trigger another debate (Kendon, 1996), this is claimed as the major reason why psycholinguists set their goals on them, regarding them as an external/visual window to the human mind (Beattie, 2003; McNeill, 2005). On the other hand, anthropological and cultural studies tend to concentrate on emblems as they are culturally specific and sensitive (Desmond et al., 1979; Kendon, 1995).

In view of the aims of this study (exploring gestures patterns co-occurring with the target formulaic sequences and speech patterns co-occurring with circular gestures), the targeted gestures are spontaneous hand-and-arm movements in interactions. That is, they are not pre-planned, but occur naturally while speaking. As far as the current dataset, i.e. real-life academic supervision meetings, is concerned, all the gestures produced in this context are naturally occurring rather than pre-planned, and hence can be

included in the analysis. Head movements can be regarded as a sub-category of gestures (Knight, 2011), but, under special circumstances, the head can be regarded and used as a third hand when both hands are not available (e.g. performing another gesture or holding something) (McNeill, 1992). Emblems are not intentionally excluded from the study as, theoretically, it is possible that they may co-occur with the target formulaic sequences. However, other bodily movements such as changes of postures, self-touching or any other self-adapting behaviours (e.g.: adjusting clothes) are excluded from the analysis.

2.4.2. Categorising gestures

There have been many categorisation frameworks put forward over the past 50 years. Kendon (2004) systematically reviewed different typologies of gestures proposed by early researchers in gesture studies including Efron, Freedman and Hoffman, and Ekman and Friesen. McNeill (1992) comments that the three schemes are fundamentally the same due to their common origin from Efron's (1941) classification system. More recent, contemporary classification schemes have also been proposed by researchers of distinctive theoretical backgrounds, among which similarities can be found in gesture description. Examples include McNeill's (1992, 2005) gesture dimensions and Kendon's (2004) pragmatic scheme, as well as others from influential gesture scholars such as Calbris (2011) and Streeck (2009b). The first two schemes from McNeill and Kendon will be introduced in detail in the rest of this sub-section as these are the two most established schemes, especially the former. The current thesis will also use their terms to describe the gestures occurring in the datasets.

McNeill's (1992) classification scheme is arguably the most widely adopted, and has been included in the annotation scheme of many multimodal

corpora such as the CID and SaGA corpus (refer back to 2.2.2.3). The scheme includes four major types of gestures, namely, iconics, metaphorics, deictics (or pointing gestures) and beats. This scheme was built up drawing on participants' narrative descriptions of cartoon clips, and the initial goal was to explore the referential significance of gesture as a semiotic sign in relation to the verbal semiotic resources; each gesture type was defined with reference to their relationship with the synchronised speech part.

Iconic gestures refer to those motions performed to describe the features of the co-occurring speech parts such as the shapes of concrete objects or movements of actions. A convenient example would be someone portraying the size or shape of an object (e.g. a book) or performing a motion (e.g. throw) when saying the words. Such a temporal and semantic synchrony between gesture and speech has been found in different languages such as English, Chinese, Japanese, Swedish and Spanish (Chui, 2009; Duncan, 2006; Talmy, 1975). Bodily movements not only can embody visual representations of the core meanings, or characteristics, of a given object or action, but also provide supplementary information that is too complicated to convey in speech (e.g. the size and shape of an object or the way a given motion is performed) (Beattie & Shovelton, 2002; Holler & Beattie, 2002, 2003). Therefore, iconic gestures can significantly improve the effectiveness and efficiency of online communication (Holler, Shovelton, & Beattie, 2009).

Metaphoric gestures may also include elements or images resembling the core features in the co-expressive parts in speech; however, unlike iconic gestures, those contents are usually abstract such as knowledge, concept, genre, proper name, argument, etc. McNeill (1992, 2005) presents some *conduit gestures* analogous to *conduit metaphor* in speech. One metaphoric conduit gesture is the Palm Up Open Hand gesture (PUOH), where a speaker holds the hand in front of the body with the palm facing up, forming a relatively bounded area like a cup. Another example is two hands with the

palms facing each other forming a bounded space between them. He suggests that the abstract idea presented is metaphorically comprehended as a concrete object that can be supported and transported by the hand.

Deictics, or pointing, are those movements executed with a body part pointing to an object, a location or a space. Such body parts are usually the fingers, but can also be others such as the lip, the head, the nose or the chin, etc. (Kendon, 2004; McNeill, 1992, 2005). The objects referred to by deictics can be something present in the context of communication or not, and the objects can be either concrete or abstract (e.g. an argument).

Beat gestures are up-down or left-right movements of the finger or the hand. They differ from other gestures in the sense that they have no overt referential correspondence in speech; in other words, beats do not contain any referential meaning. In terms of function, they are mainly used by speakers to mark-up discourse relationships by making certain segments of speech more salient by adding gestural emphasis to them (Duncan, Cassell, & Levy, 2007; McNeill, 2005). In this sense, beats can be regarded as metaphoric and their major functions tend to be more metalinguistic and pragmatic rather than referential (Gullberg, 1998).

To sum up, as mentioned above, these types are not mutually exclusive, and one type of gesture has the potential of simultaneously belonging to more than one type. This can cause confusion in coding as reported by Seedhouse and Knight (2016). McNeill (2005) later slightly revised his theory and proposed that each type is indeed a dynamic dimension rather than a static and exclusive type. The concept of dimension means one gesture can belong to more than one dimension at the same time, which is normally not allowed by the concept of ‘category’.

The current thesis will continue to use McNeill’s classification scheme to refer to different gesture types. It is not only a well-established coding system, but also an applicable one. The physical form of the deictic/pointing and beat

gestures are easy to distinguish, and so are the speech contexts of iconic (co-occurring with concrete objects/actions) and metaphoric (co-occurring with abstract lexical items) gestures.

Taking a pragmatic, context-based approach, Kendon (2004) proposes the necessity of investigating gestures, especially conventionalised gestures, in social communicative contexts and sees gestures as social bodily actions that are employed to accomplish the speaker's momentary communicative goals. Kendon (2004) puts forward three ways that gestures can be used, namely, pragmatic actions, deictic actions (pointing) and representational actions. Although from a different theoretical stance, it can be inferred from the categories that Kendon's classification shares many similarities with McNeil's (2005) framework.

The name, *pragmatic actions*, mostly reflects Kendon's view of gestures as social actions designed to achieve communicative goals, and mainly includes those gestures that are performed as *speech acts* in order to complete some communicative function such as 'negation', 'confirmation', 'request', 'offer', etc. *Deictic actions* are basically the same as those of McNeil's pointing gestures, orientating the recipient's attention by pointing towards a concrete or abstract object in the real or imagined world using different body parts such as fingers, lips, heads, etc. Additionally, the pointing gestures may have some cultural variations such as the lip pointing in Australian aborigines and South-East Asians, and the 'nose-winkle' in New Guinea (Kendon, 2004). *Representational actions* can be regarded as the combination of iconic and metaphoric gestures in McNeil's description, in which the gestures represent some features of the concrete or metaphoric objects, motions or relations.

As admitted and stressed by Kendon (2004), these three variations of gesture actions are not categorical and mutually exclusive. One can easily imagine a speaker pointing to an object with a Palm Up Open Hand gesture and simultaneously performing two actions, pointing and pragmatic

(presenting). Overall, whereas deictic actions and representational actions are relatively easy to distinguish, pragmatic action is such a broad concept that it can overlap with the other two actions. Kendon's scheme is not as widely used as McNeill's, but the concept of representational gestures (referring to both iconic and metaphoric gestures), as opposed to beats and deictics, is commonly seen in the literature.

At the end of this sub-section is a brief mention of those researchers who attempt to link gesture categories to Pierce's semiotics: symbol, icon and index (Enfield, 2013a; Mittelberg, 2008). In semiotics, symbols are those signs whose signifier-signified mappings rely on socio-cultural conventions rather than the here-and-now interpretation of the context. Icons refer to those signs that have similarities, or resemblance, between the signifier and the signified. In index signs, there exist certain contiguity between the signified and the signifier such as causal relationship. For instance, viewing gestures as signs, Mittelberg (2008) proposes that those gestures (such as emblems) that hold a conventional form-meaning correlation across contexts in a given speech community are symbols. Iconic and metaphoric gestures that share some similarities between the gesture forms and their co-gestural speech are icons. Gestures such as pointing gestures can be considered as index signs. However, caution needs to be taken when simply mapping the three aspects of sign onto certain gesture categories as the decision of whether a gesture is an icon, a symbol or an index needs to be made by virtue of its singular instance of use (Hanna, 1996).

2.4.3. Frameworks for segmenting gesture phases

To examine the alignment and interplay between speech and gesture, researchers have developed frameworks for demarcating gestures into different phases (i.e. gesture phases) just as speech can be segmented into

words, phrases and clauses. This sub-section will demonstrate in detail the schemes proposed by Kendon (2004), McNeil (2005) and Kita's team (Kita et al., 1998). These frameworks were developed based on the close observation and analysis of the changes of the hand and arm movements (e.g. changes of the hand shape, speed intensity, trajectory, etc.), and the gesture's relationship with the co-occurring speech. They were chosen mainly because they are the most widely adopted and influential frameworks which have proved particularly valuable for gesture segmentation and analysis.

Kendon (2004) mentions five phases, i.e. some position of rest or relaxation, preparation, stroke, post-stroke hold and recovery. The position of rest or relaxation is the position where the hand is resting or in a relaxed condition. Stroke is the phase with the most laborious motion where the meaning of a gesture is completed. It is the only compulsory part that is usually co-expressive with the co-occurring speech. Following the stroke, there may be a post-stroke hold phase where the hand, as the name suggests, remains still and maintains at the end of the stroke phase. The post-stroke hold phase is the prolonged part of a stroke, and the stroke and post-stroke phases together form the nucleus of a gesture, where the meaning is expressed. The movement coming before or preparing for the stroke (e.g. formulating the handshape, getting more strength in the hand, or moving the hand away from a rest/relaxed position, etc.) is the preparation phase, during which there might be pauses or holds. Such a hold is different from the post-stroke hold. The movement after the stroke where the hand relaxes or returns to a rest position is the recovery phase.

McNeill (2005) proposes a similar framework, including preparation, pre-stroke hold, stroke, stroke hold, post-stroke hold and retraction. The preparation, stroke, post-stroke and retraction (recovery in Kendon's framework) have the same definitions as those of Kendon, except for pre-stroke hold and stroke hold. Pre-stroke hold refers to the temporary pause

of the hand as it reaches the starting point of the stroke. McNeill gave an example of a pre-stroke hold, underlining its function of maintaining the temporal alignment between the stroke and the co-expressive speech ("he bends it way back"). The stroke movement of the gesture is a pulling down to the rear movement, which is semantically coordinated with the speech of "bends it way back". However, the hand reaches the starting point of the stroke during "he" and before "bends". The speaker then holds the hand in that position, which triggers a pre-stroke hold, and pulls his hand to the back while he says "bends". Using the same illustration, McNeill demonstrated a post-stroke hold during "ck" as the speaker can finish saying the word "back". Hence, in line with Kendon (2004), he claims that the post-stroke hold plays the same role as the pre-stroke hold in synchronising the stroke movement and the associated speech part. In other words, because of the pre-stroke and post-stroke hold phases, the stroke of pulling back can co-occur with its co-expressive speech "bends it way back". The stroke hold is a special type of stroke, which is motionless, but has effort and meaning. An example provided by McNeill is a hand held up, imitating the upper floor of a building.

Kita's team (Kita et al., 1998) establishes a similar but more detailed scheme for segmenting movement phases for both signs and co-verbal gestures, which contains preparation, stroke, holds (including dependent holds such as pre-stroke hold and post-stroke hold, and expressive independent holds similar to the stroke hold in McNeill's framework), retraction and partial retraction. They also specified that preparation might contain 'an optional *liberating movement*, which makes the hand free from a constrained position such as undoing the interlocking of fingers by two hands. The *locational preparation* brings the hand to the starting position of the expressive phase, and the *hand-internal preparation* is a hand-internal movement, in which the hand shape and the orientation of the hand are set to the starting values of the expressive phase (Kita et al., 1998)' (the expressive phase refers to the stroke).

They point out that ‘the location preparation and the hand-internal preparation overlap in time in the majority of the cases. Furthermore, the onset of the hand-internal preparation does not precede the onset of the location preparation’ (Kita et al., 1998). In addition, they distinguish partial retraction from full retraction (the hand returns to a rest position) to designate a situation in which the retraction phase is interrupted, and the hand starts another phase, such as a preparation or a stroke. It is also worth mentioning that they clarify the definition of the *resting position*, which ‘is part of the body or the furniture where the hands can be supported’ (e.g. lap, folded arms resting on the chest, a chair, a table, etc) (Kita et al., 1998). The resting position ‘could also be the position where self-adapting body movements take place’ (e.g. combing the hair with the hand, adjusting clothes, manipulating objects, etc.) (Kita et al., 1998).

As can be seen from the above discussion, the three frameworks are essentially the same in terms of being composed of four types of phases, i.e. preparation, stroke, hold and retraction/recovery, though differ slightly in details of description. Among those phases, only the stroke phase is obligatory for a gesture to exist. This thesis also adopts the phases proposed by them, and these will be introduced in the next chapter. There are certain other coding schemes available such as the one focusing on the movement features of different gesture phases (Bressem & Ladewig, 2011), and another distinguishing the boundaries between static (i.e. hold and rest position) and dynamic (i.e. gesture phases such as preparation, stroke and retraction) hand status based on the clear-blurry transitions of images (Seyfeddinipur, 2006). However, they are not relevant here. McNeill’s (2005) coding scheme for gesture space (see Figure 2.1 below) is also very widely used, but, again, is not the concern of the current study.

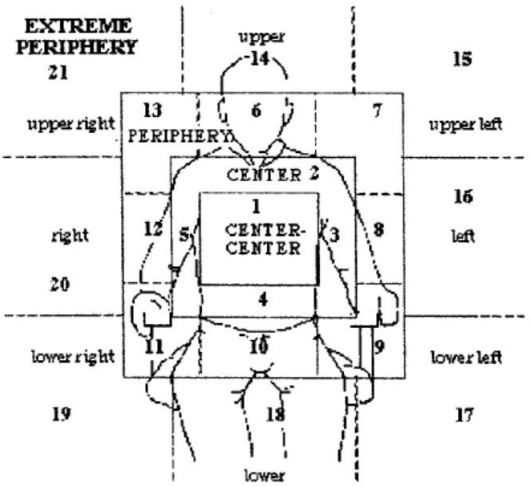


Figure 2.1: Framework for segmenting gesture space (McNeill, 2005, p. 274).

2.4.4. Temporal and semantic relationship between speech and gesture

Previous investigations into the synchrony between gesture and speech has generally uncovered the very close temporal proximity between the two, which becomes a crucial factor for arguing for the integral relationship between the two modalities in the creation of utterance meaning (Holler & Beattie, 2004; McNeill, 1985, 2005). The most expressive part of a gesture, or stroke, usually co-occurs with the segment of speech with which it is closely associated in meaning. The co-gestural segment of speech is called the 'lexical affiliate' (Schegloff, 1984), or the 'idea units' (Kendon, 2004). Using the example above, the "bends it way back" is the lexical affiliate, or the idea unit, of the pulling back stroke. Such a timely alignment between speech and gesture is fundamental to their simultaneous contributions to the utterance meaning (McNeill, 1992, 2005). The meaning of gesture can be the same as the co-occurring speech, but also complementary or supplementary (Goldin-Meadow, Alibali, & Church, 1993; Wagner, Malisz, & Kopp, 2014). Complementary gestures contain additional information that has been referred to in speech. This could be the case when someone is saying "The ball is this

big" and simultaneously depicting its approximate size. Supplementary gestures encode information that is not referred to in their accompanying speech. For instance, a child can point to candy while saying "I'm hungry". The mismatch of information in speech and gesture has been explored in child learning in psychology (Goldin-Meadow, 2015).

Due to the match in time between speech and gesture, Kendon (2004) claims that speech and gesture are integrated parts of utterances, both of which must have been planned together before the speaker utters the words and performs the gesture. For Kendon, gestures are social moves, or actions (Austin, 1975; Manning, 2013), and, to accomplish a communicative action, the speakers usually manage to manipulate and mediate between the two modes into a speech-gesture *ensemble*. From a psychological point of view, McNeill (1992, 2005) proposes the Growth Point theory which holds that speech and gesture grow out of the same idea unit, i.e. they are co-expressive of the same mental image/idea, though in quite different ways. For McNeill, gestures stand for imagery and synthetic thinking and speech is analytic. Hence, according to the Growth Point theory, the synchronisation between the two channels are deeply rooted in the mind rather than at a superficial level. Even when gestures embody an aspect of meaning that is similar to the speech component, they are never redundant as they represent separate parts of the same idea unit.

Such a deep cognitive relationship between speech and gesture can become even more revealing in psycholinguistic studies on the semantic and syntactic coordination between gesture and speech (Chui, 2009; Duncan, 2006; Duncan et al., 2007; Kita et al., 2007; Kita & Özyürek, 2003). Taking a well-known cross-linguistic study as an example (Kita & Özyürek, 2003), where participants were required to retell the same story from the same cartoon video (Talmy, 1975) which includes the motion event *swing*, Japanese and Turkish speakers were less likely to produce the arc trajectory in gesture

than English speakers. This is because the Japanese and Turkish motion verbs do not encode the path of arc as the English verb *swing* does. That is, the production of gestures tends to be in accordance with the meaning of lexical items prestored in the mind.

Although the integral relationship between speech and gesture is undeniable, this thesis raises two issues concerning the speech-gesture temporal relationship which awaits further investigation. Firstly, despite the consensus that strokes and their lexical affiliates co-occur, there is insufficient empirical evidence regarding where exactly the onset and the end of strokes co-occur in relation to the lexical affiliates. Schegloff (1984) claims that gestures tend to precede their lexical affiliates, but he has neither reported any empirical evidence of this, nor does he clarify the gesture phase (e.g. preparation, stroke) to which he is referring by using the term 'gesture'. More recent investigations have been conducted by a few researchers such as Chui (2005a) and Ferré (2010) which were designed to address the question of where exactly the strokes start and end in relation to their lexical affiliates, and their findings differ considerably. Whereas 182 strokes out of 303 iconic gestures (60.1%) tended to synchronise with the lexical affiliates in Chui's research, 72% of the 244 iconic gestures in Ferré's tended to precede their lexical affiliates. Ferré (2010) also examined the end points of the strokes and their lexical affiliates, leading to the result that the former ends slightly after the latter.

Secondly, although previous research on the temporal relationship between speech and gesture mostly relies on iconic gestures that describe concrete objects and actions (refer back to McNeill's categorisation), there has yet sufficient research regarding the ways in which other types of gestures (e.g. metaphoric, deitic, beat gestures) align with their associated speech. For instance, Harrison's (2010) research on the temporal relationship between negation and the Palm Down gesture (a recurrent metaphoric gesture encoding

the meaning of negation) generated four types of synchronisation, caused by factors such as grammatical structure, discourse environment and the intention of the speaker. Taking two types of context as demonstrative, the strokes co-occurred with the node of negation (such as *not* in *don't*) when there was no speech following the node (e.g. *I don't*); however, when there were Negative Polarity Items such as *at all*, *any* or *single*, which can reinforce or emphasise the sense of negation, the strokes of the Palm Down gestures tended to accompany them. This research indicates that the temporal coordination between metaphoric gestures might be much more complex than expected, and more research is needed to address the issue.

The two issues concerning the temporal relationship between speech and gesture discussed above will be explored in detail in Chapter 6, along with a demonstration of some valuable implications offered by the two studies of the current thesis.

2.4.5. Analytical software

This section aims to discuss the challenges of analysing multimodal data and realising automatic search and annotation. In addition, a brief review of the currently available multimodal analytical tools will also be provided.

Owing to the labour-intensive nature of multimodal research, the realisation of automatic, or even semi-automatic, data capturing and encoding is vital for enhancing the level of standardisation, objectivity, quality and quantity of research of this nature. Whereas recording, tagging and analysing visual representation of textual data has reached a high standard of sophistication, the development of tools for multimodal data with parallel functions is somewhat insufficient. As the majority of tools in gesture research stem from human-computer interaction (HCI) and Sign Language recognition, they have mostly been developed and applied in a relatively rigid,

laboratory-based environment, which make them less robust when being applied to datasets of natural interactions (Adolphs & Carter, 2007).

In theory, with the current machine learning technologies such as the Hidden Markov Model (HMM) and algorithms, it is likely that automatic real-time gesture tracking and coding of any type of bodily motions can be realised. However, due to the considerable complexity and idiosyncrasy of gesture forms in, and across, individual speakers, even the most cutting-edge multimodal projects to date have only managed to automatically detect those seemingly routinised gestural sub-categories, such as the head nods/shakes and beat gestures with a varying degree of accuracy (Kapoor & Picard, 2001; Knight, Evans, Carter, & Adolphs, 2009a; Seedhouse & Knight, 2016). The aid of the available technologies can significantly enhance the efficiency of research and pave the way for future technological development.

Although automatic speech and gesture search and annotation might be, to a great extent, unrealistic at the present time, tools available for multimodal synchronisation, alignment, annotation and analysis are numerous. They are created by the private sector and academic institutions and can be used either free of charge or with a reasonable fee. With the abundance of tools, it is, therefore, crucial for researchers to be aware of the theoretical and methodological assumptions underlying each tool, tailoring these to their own research. For instance, ELAN (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006) and Anvil (Kipp, 2001) are the most prevalently adopted free software in gesture studies, especially from the psycholinguistic and linguistic perspectives. They offer the functions of marking-up multiple tiers or tracks of verbal and nonverbal data, and support frame-by-frame observations of motions. Both ELAN and Anvil have the function of importing and exporting of data from Praat (http://www.fon.hum.uva.nl/praat/), which supports phonetic analysis. Following many gesture investigations, the current thesis also used ELAN as a free and powerful tool for gesture analysis (details of

analysis will be provided in Chapter 4 and Chapter 5). Transana (https://www.transana.com) is a powerful tool in terms of assisting speech transcription and synchronising different sources of data, allowing simultaneous assessment of speech transcripts and the corresponding audio/visual data. Therefore, Transana seems to be favoured by researchers in the field of conversation analysis. A fee is required to be paid for the use of Transana and its related products.

Other observational tools are also commercially available. They more or less function in the same ways as ELAN, Transana and Anvil, although some have special features. Examples include the multimodal observation tool, Interact (https://www.mangold-international.com/en/software/interact), which may prove particularly useful for coding online live data without downloading them, and clustering and examining behaviour patterns. The Observer XT (http://www.noldus.com/animal-behavior-research/products/the-observer-xt) is a specific tool invented not only to code human behaviour, but also that of animals. Diver (http://diver.stanford.edu/home.html) promotes panoramic video capture, featuring the observation of videos from all perspectives, and is particularly designed to overcome the issues caused by 2D, single-perspective observation and analysis.

However, until now, there is no software that provides automatic coding of the meanings/functions of gestures. This is mainly due to the fact that the meanings of most co-verbal gestures dominantly rely on the corresponding discourse contexts, which involve different sociocultural factors, objects, materials, etc. (Calbris, 2011). The complex and dynamic nature of multimodal communication means that its coding may never be completely automatic (Adolphs & Carter, 2007), but the existing multimodal analytical tools have already afforded current researchers unprecedented opportunities of exploring the patterning of multimodal human behaviour in a much more efficient way than previously.

2.5. Towards the meaning of gestures of high occurrence

The premise of exploring the multimodal units of meaning in speech and gesture is that, similar to speech, gestures are meaningful, and patterns of their use can be established. This section will be dedicated to reviewing research exploring such issues. The fact that gesture has not seen as pertinent to meaning as words, until the works from early gesture scholars such as Kendon (1981) and McNeill (1985) have been widely accepted, might be one reason why attention to gesture is long overdue in linguistics. This is understandable as it appears that information can be delivered, in most circumstances, solely by 'words'. Except for emblems, which tend to be recognised as most eligible for a position in a formal linguistic system owing to their stable form-meaning associations across contexts in a given culture, most co-verbal gestures are perceived as improvised and idiosyncratic both in form and meaning.

However, in great contrast to the neglect of gesture in linguistic studies, there is the interlocutors' constant use of gesture in meaning construction in communication, regardless of whether this is done consciously. Considering the four types of gestures proposed by McNeill (2005), iconic gestures often complement speech by portraying the particular size or shape of an object and enacting the trajectory/mode of a motion event (Chui, 2009; Holler et al., 2009). Metaphoric gestures such as the Palm Up Open Hand gesture have a sense of offering or requesting a response (Kendon, 2004; McNeill, 2005). Deictic gestures are prevalently performed by speakers to point to a specific entity that is absent from speech, making both the pointing gesture and the signified entity indispensable elements of the utterance (Fricke, 2013; Goodwin, 2007). Although beat gestures do not link to any lexical item, they emphasise the co-occurring speech and, thus, add some extra meaning of importance to the utterance (Rossini, 2012).

Despite the increasing acceptance of gestures as a significant mode of

semiotic resources in communication, systematic investigations of the meaning variations of certain gestures in different speech contexts is still insufficient. Conducting such a study is not as straightforward as conducting word-centred linguistic research, and the main obstacle is the seemingly idiosyncratic forms of most gestures. There appears to be no starting point to investigate the patterns of gesture use. However, beginning from those frequently seen and relatively recognisable gestures, efforts have been made to examine the meaning variations of these gestures in communication. Three strands of research of this nature will be reviewed in the following sections, including research on the use of emblems or quotable gestures (Brookes, 2005, 2011; Kendon, 1995; Payrató, 1993; Sherzer, 1991), gesture families (Kendon, 2004) and recurrent gestures (Ladewig, 2014a). The following sub-sections will report on each in sequence.

2.5.1. Emblems

For a gesture to be an emblem, the core features mentioned in the existing literature generally include a high level of conventionality of the mapping between gesture form and its meaning, a broad geographical distribution within certain socio-cultural groups and the potential to be a substitute for speech as a lexical component (Brookes, 2004; Ekman & Friesen, 1969; Hanna, 1996; Kendon, 2004). Early research on emblems tends to concentrate on establishing a repertoire of emblems for a certain cultural community, i.e. constructing a list of verbal translation for the identified emblems (Brookes, 2004; Desmond et al., 1979; Payrató, 1993; Salvato, 2008). However, there is little research on the use of emblems in actual communication within a social and cultural context, except for a few studies mainly from Kendon (1992, 1995, 2004) and Brooks (2005, 2011). The exploration of emblems in interactions is crucial in understanding how they lend themselves to the online momentary construction of meaning, under which circumstances they tend to

emerge, how they relate to speech referentially and pragmatically, and the level of conventionality in different emblems, etc. (Brookes, 2005; Hanna, 1996; Kendon, 2015). The review here focuses on the meaning variations of emblems.

In Kendon's (1995, 2004) examination of four gestures in Southern Italians, there are two emblems, or quotable gestures, which are the Purse Hand (Mano a borsa) (Figure 2.2), the Praying Hands (Mani giunte) (Figure 2.3). Examples from video recordings show that the purse hand means questioning, criticising and reported questioning; the praying hand means begging someone to accept certain conditions expressed in the utterances or in relinquishing the speaker's responsibilities for doing something. Similarly, Brooks (2005) studied the uses of three emblems among black urban residents of a South African region and named them based on their meanings, these being the Drinking Gesture, the Clever Gesture and the Money Gesture.



Figure 2.2: The purse hand.



Figure 2.3: The pray hand.

Fine-grained interactional analysis of emblems shows that they can certainly co-occur with their lexical equivalents, but they can also modify the utterance

meaning. For instance, in one robbery instance of Mano a borsa (Kendon, 1995), the purse hand gesture was maintained throughout the time that the speaker was describing the physical environment of the crime. Whereas the speech part of the utterance embodies no intention of making an enquiry, this meaning is encoded in the purse hand. It is the gesture that informs the interlocutor that the utterances are not simple description, but questioning.

Furthermore, with the core meaning remaining the same, minor changes of the gesture forms could lead to minor variations in meaning. For instance, the differentiations in gesture space and direction in the purse hand gestures reflect different degrees of relevance of the action of questioning to the current speaker. Those purse hand gestures performed at the high/face level of, and closer to, the speaker's body are in line with the speaker's own question, whereas those at the lower level away from the speaker's body co-occur with reported questions (Kendon, 1995, 2004). In Brookes' research on the Drinking Gesture (Figure 2.4) (Brookes, 2005), the stroke of the gesture usually involves moving the forefinger of the left hand to the right hand. The amount of up-down movements or the intensity of the actions can signify various amounts of drinking: the more movements there are and/or the more forceful the movements, the greater amount of alcohol consumption is implied.



Figure 2.4 The Drinking Gesture (Brookes, 2005)

To conclude, Kendon and Brookes' research on these emblems demonstrates

their meaning variations within certain cultural communities. They also discuss the ways in which emblems could slightly vary in their gesture parameters (e.g. space, intensity, direction, etc.) in accordance with different discourse contexts. However, as both Kendon and Brookes took a more anthropological approach and only offered a detailed analysis of a few interactional episodes, neither has provided a numerical description of the gestures they have observed. Hence, we require stronger empirical evidence of how these emblems are actually used in their corresponding cultural communities.

2.5.2. Gesture families

The previous sub-section has introduced Kendon's (2004) research on two emblems based on recordings from Southern Italians. Drawing on the same dataset, he also reports four frequently occurring gestures that have stable form-meaning associations, which he terms gesture families. The four gesture families are 1). the *grappolo* ('bunch') gesture or the purse hand or the finger bunch (the G-gesture), 2). the ring gesture (the R-family), 3). the Open Hand Supine family (OHS, informally the palm up family), and 4). the Open Hand Prone family (OHP, informally the palm down family). As Kendon looked at their meanings/functions from a pragmatic perspective, he also names them pragmatic gestures.

A gesture family can be regarded as a group of gestures with at least one salient common element in the kinesic construction coupled with one semantic theme. The G-gesture family is characterised by the finger bunch or the purse hand gesture with all the digits extended and with a joint at the knuckles (the precision grip). The gesture family has the function of marking topics, asking questions and demanding justifications. Four variations of the G-family have been discovered depending on the changes of the purse hand (into or out of the

finger bunch) and the movement (upwards and inwards to the speaker, or downwards).

The R-gesture family (Figure 2.5), as the name suggests, has the same ring hand shape, with the joint of the index and thumb at the tips of the fingers and the other three fingers separated and extended (slightly flexed). Similar to the G-gesture family, the forms of the R-gesture family can vary slightly in terms of the changes of the hand shape (in or out of the ring hand shape) and movement patterns, but they share the same meaning of being exact or precise.

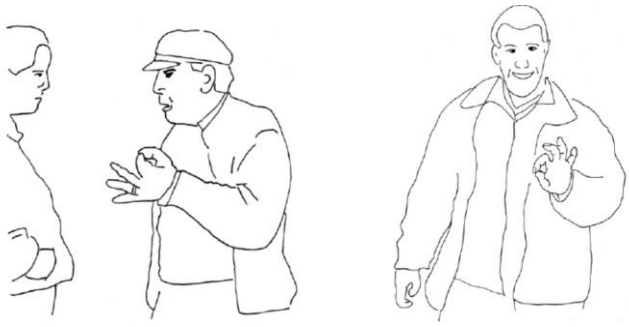


Figure 2.5: Two examples of the ring hand shape in the R-gesture family (Kendon, 2004).

The Open Hand Supine and Open Hand Prone family are not only prevalent in Southern Italy, but also in many other cultures including English, German, Chinese, French, etc. The Open Hand Supine (Figure 2.6) is also commonly known as the Palm Up Open Hand gesture (PUOH), a term proposed by McNeill (1992, 2005) as a conduit metaphoric gesture.



Figure 2.6: Open Hand Supine gesture (Kendon, 2004).

Numerous gesture researchers have discussed the palm up gesture and similar characteristics of this have been demonstrated regarding its form and uses in context (Calbris, 2011; Cienki, 2008; Gullberg & McCafferty, 2008; Streeck, 2009b). The form of a palm up gesture usually includes the action of opening the hand with the palm facing upwards in front of the speaker, forming a cup-like shape on the surface of the palm as if holding and presenting something. Speakers perform this gesture family when they are offering, giving, or receiving, usually something abstract such as a concept, a demonstration, a theme, etc.

The Open Hand Prone family (Figure 2.7) is another established gesture with the palm facing downwards or outwards. It is associated with the meaning of rejecting, interrupting, stopping, denying, etc. Whereas the semantic core of negation maintains the same, other parameters of the gesture can change the degree of intensity. Such changes can come from the form of the hand shape (floppy or fully extended), the velocity of the movement and the distance of the hand away from the speaker (Calbris, 2011). In short, the faster/stronger the movement is, the more well-formed the hand is, and the further the hand is away from the body, the stronger the meaning of negation.

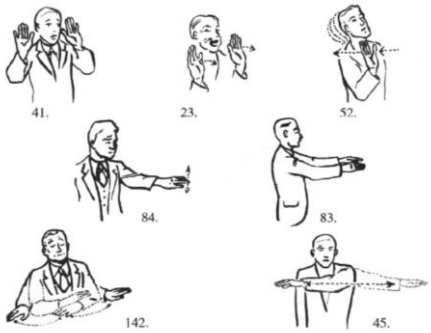


Figure 2.7: Examples of Open Hand Prone gestures (Kendon, 2004).

Nonetheless, as with the research on the emblems, although the findings of the gesture families above appear plausible, we still need further empirical

investigation of the ways in which these gesture families are used in real-life communication within different cultural groups.

2.5.3. Recurrent gestures

More recently, a group of German scholars has begun exploring the linguistic potential of some frequently occurring gestures, which they call recurrent gestures (Ladewig, 2014a). Recurrent gestures and gesture families are essentially the same. They take a form-based approach and hold that gestures can be segmented into kinesic components (Calbris, 2011), or *kinemes* like 'phonemes' and 'morphs' (Birdwhistell, 1970). These components can be the movements, space, handshapes, etc. of the motions of the hand, arm and finger, all of which can contribute to the utterance meaning and can be investigated as discrete semiotic units (Kendon et al., 1981; Ladewig & Bressem, 2013; Rossini, 2012).

Such a form-based stance can also be observed in Kendon and Brookes' research on the meaning variations of emblems and gesture families, although they have not articulated this overtly. This perspective is in contrast with the view that gesture meanings are non-combinational and not subject to demarcation, i.e. the meaning of each part is determined by the meaning of the whole gesture and all the parts merged into one single unit of meaning in the entire gesture (McNeill, 1992, 2005).

Drawing on German discourse, fourteen recurrent gestures with stable meanings are proposed, i.e. form-meaning associations, and the ultimate goal of such research is to establish a grammatical system of gestures, i.e. multimodal grammar (Bressem, Ladewig, & Müller, 2013; Calbris, 2011; Fricke, 2013; Ladewig, 2014a). Examples include movements of throwing away, sweeping away, cyclic, dropping of the hand, etc. However, the empirical scope of these gestures is insufficiently wide to make

generalisations, and future research is required to examine each of the recurrent gestures in different languages other than German.

One of the earliest and most systematic studies of recurrent gestures is the cyclic gesture in German speakers (Ladewig, 2011, 2014b). With the same cyclic movement, the two parameters including the position and size can contribute to semantic variations across contexts: 1). in the context of expressing the notion of on-going activities with the hand positioned in the peripheral space; 2). in the context of non-fluent speech signifying searching for words/concepts with the hand in the central gestural space; 3). in the context of encouraging or requesting the other party to continue an action with a large-size cyclic gesture by rotating the elbow rather than the wrist. These results are partly in concert with those of French speakers who use the cyclic motion to express the sense of the progression of time and rotation of events (Calbris, 2011).

However, a closer inspection of Ladewig's research on the cyclic gesture suggests certain drawbacks which affect its validity. First of all, not all the information of the chosen parameters has been provided, presumably space and size, in all three functions, but only gesture space has been mentioned in the first two functions, and gesture size in the last category. It is unclear whether the cyclic gestures in the first two functions are large, and those in the last function are in the central/peripheral gesture space. That is, it has not been specified why that specific gesture parameter for each category has been chosen, i.e. space for the first two, and size for the third.

Secondly, actual statistics of how many gestures in each function possessing that particular feature have not been given. For instance, it is realistically impossible that all the gestures in the first category are all in the central gesture space, and the researcher needs to provide a specific percentage of such gestures in that category. The same problem applies to the other two functions.

Finally, coding the space and size of gestures can be very challenging and subjective, which make the findings less convincing. As the second study of the thesis focuses on a similar gesture, the circular gesture, details regarding the problems of coding gesture space, size, velocity will be provided in Chapter 5, 5.2.1. These problems prevent the researcher from conducting a more fine-grained analysis of gesture parameters in each meaning variation. Moreover, it is also possible that, although rotational gestures can have minor variations in gesture parameters, such differences may not have a great impact on the general meaning variations of the recurrent gesture.

The conclusion to this section provides a brief comment on McNeill's (2005) illustration of the conventionality of gestures (see the continuum in Figure 2.8).

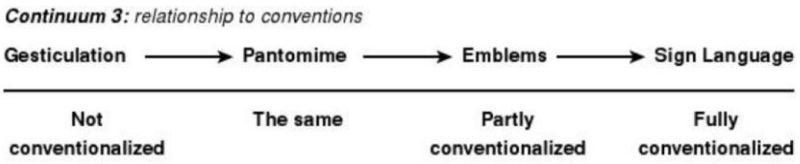


Figure 2.8: The continuum of relationship to conventions in gesture (McNeill, 2005, p. 10).

In his discussion of the degree of conventionality in gesticulation, pantomime, emblems and sign language, he placed gesticulation and pantomime on the left of a continuum as being not conventionalised, whereas emblems are placed towards the right end as partly conventionalized, and signs to the far right end as fully conventionalised. However, what has not been taken into consideration in this continuum are those gesticulations that have relatively stable form-meaning associations across contexts such as the gesture families and recurrent gestures. They are more conventionalised than those idiosyncratic gestures and less conventionalised than emblems. The major

differentiation between them and emblems is that they are not detachable from speech, or unable to be used as verbal substitutes (Brookes, 2011; Kendon, 1992; Parrill, 2008; Payrató, 1993). Hence, those frequently occurring gestures (which will be called recurrent gestures for convenience) with relatively stable form-meaning associations can be placed somewhere between gesticulations and emblems. Currently, it is this type of gestures that attract an increasing amount of attention.

2.6. The functions of gestures

In view of the discussion in the previous sections, it is without doubt that gesture plays a crucial role in interactions, and previous researchers have established their functions from a variety of perspectives. For instance, Calbris (2011) underlines the demarcative and referential functions of gesture. The demarcative function means that, as gestures change from one to the next, they segment utterances into units of different sizes from words, to phrases and to sentences.

3 *qui vont permettre de stocker l'eau glacée* *et de façon à la restituer* ||
[Both fists Right fist]
which will enable chilled water to be stored and in a way that it can be released

4 *pendant les les périodes* *d'élévation de température* ||
[Hands moving apart joining together]
during the the periods of high temperature

6 *provenir euh de* *de l'utilisation inadaptée euh* ||
[Open hands facing upwards alternating up-down movement]
arise er from from the ill-adapted utilisation er

7 *de de* *de l'énergie électrique* ||
[joined cupped joined]
of of of electrical energy

8 *et cetera* *Donc* *tout cela a été* *recherché derrière le décor* ||
[cupped close Both fists Left fist]
et cetera So all that has been researched behind the decor

Figure 2.9: Gestures segment idea units (Calbris, 2011).

Taking the above piece of discourse as an example (Figure 2.9), Calbris (2011) demonstrates in detail how the changes of the hand shape and movement are in accordance with the demarcation of semantic units from 3 to 8. For instance, the change of the hand from both fists to right fist divides unit 3, and the shift of movement demarcates unit 4 and 6.

To a great degree, the demarcative function of gesture is closely associated with their referential function. The referential function refers to the fact that gestures contribute to the meanings of utterances, a fact that has been widely discussed and acknowledged (Calbris, 2011; Kendon, 2004; McNeill, 2005). Gestures tend to co-occur with the speech components, or idea units, with which they are associated in meaning. As a result, as the idea units unfold in utterances, gestures tend to change accordingly. Hence, as pointed out by Calbris (2011), gestures usually play demarcative and referential functions simultaneously.

Gestures can perform pragmatic functions. As discussed in the previous sections, Kendon (2004) views gestures as actions. Because of this, he not only explores the referential functions of gestures, but also examines their pragmatic functions, in relation to his studies of emblems and gesture families in Southern Italy (refer to 2.5.2.). For example, the purse hand can perform the action of criticising/questioning; the praying hand can beg; the palm up gesture can present; and the palm down gesture can negate.

Gestures can also mark discourse relationships, foregrounding those components that are more important than others. McNeill (2005) found that beat gestures tend to co-occur with new or noteworthy information in discourse. Similarly, Kendon's (2004) research on the G-gesture family (refer to 2.5.2.) shows that the finger bunch usually synchronises with the topic of the speakers' discourse.

It should be pointed out that, first and foremost, the four functions mentioned above (demarcative, referential, pragmatic, and discourse marking)

are not mutually exclusive. For instance, gestures of any function are segmenting utterances into units, and a gesture that performs a pragmatic function is also contributing referentially. Secondly, the functions of gestures are not limited to the four discussed above, and this thesis intends to explore this issue further based on the current datasets. Finally, researchers have also discussed the role of gestures from other perspectives. For instance, there is a lengthy debate concerning whether gestures are performed mainly for the sake of speakers (assisting speech production) (Beattie & Aboudan, 1994; Butterworth & Beattie, 1978; Hadar & Butterworth, 1997) or listeners (assisting communication) (Alibali et al., 2001; Bavelas et al., 2008). Moreover, there are a few researchers dedicated to investigating gestures that are specifically addressed to the interlocuters in conversation (e.g. gestures seeking for agreement), which they term interactive gestures (Bavelas, Chovil, Coates, & Lori, 1995; Bavelas et al., 1992). These issues will be addressed further in Chapter 6.

2.7. Interpreting gesture-meaning association

Underlying the identification of the form-meaning associations in frequently occurring gestures is the issue of how to justify these associations (Enfield, 2009). One plausible approach is to turn to cognitive linguistic theories that examine the metaphoric relationship between speech and gesture. The following sub-sections will mainly introduce two similar cognitive theories concerning gesture and metaphor (2.7.1.) and the analogic links underlying gesture form-meaning associations (2.7.2).

Before presenting different approaches for interpreting the meaning of gestures and their parameters, it is necessary to point out that the meaning of gesture as a sign (a symbolic combination of form-meaning unit) is not as specific as that in speech, but better defined as ‘notions’ that could be realised

by different various verbal materials (Calbris, 2003, 2011; Enfield, 2009; Rossini, 2012). That is, when it is claimed that one variation of the cyclic gesture has the meaning of progression, it is not that the gesture is an illustrator of that specific word, *progression*, but that it occurs in a discourse context where such a meaning is indicated (e.g. "I'm trying"). Moreover, the notion of gestures cannot be properly comprehended without speech, i.e. the interpretation of the meanings of gestures rely heavily on the discourse context (Adolphs & Carter, 2013; Streeck, 2010).

2.7.1. Gesture and metaphor

There has been a growing area of research which investigates metaphor in gesture and explores the relationship between metaphor, gesture and cognition (Cienki, 2016; Cienki & Müller, 2008a). The study of metaphor is an important field of research in cognitive linguistics, and Conceptual Metaphor Theory predominantly concerns the metaphoric thinking of human beings, i.e. the ability to understand one thing/the source domain out of the other/the target domain (Lakoff & Johnson, 2008). Deeply underlying metaphoric conceptualisation is the body's recurrent motor-sensory experiences, or image schema, in the real world. As past research on metaphor mainly draws on written texts (including speech transcripts), the possibility of studying metaphor in gesture is mainly due to the fact that metaphor is modal independent. That is, in principle, metaphor can be represented in any modality such as speech, gesture, image, etc. either individually or concurrently (Forceville & Urios-Aparisi, 2009).

Although the ultimate goal of studying metaphor in gesture is to further our knowledge of metaphoric thinking and how common/individual momentary experiences of the body are abstracted and perceived in the mind, metaphor and image schema have proven valuable in justifying the form-meaning associations in recurrent gestures (Cienki & Müller, 2008b; McNeill, 2008). For example, the

Palm Up Open Hand gesture (McNeill, 1992) embodies the IDEAS ARE OBJECTS metaphor so that ideas can be held by hand. In other words, the association between the PUOH gesture and the meaning of presenting and receiving is rooted in the image schema (i.e. recurrent motor-sensory experiences) of giving and receiving objects by the hand in our world experience. Due to the prevalence of such behaviour, such a metaphoric association of this gesture is recognisable across many cultures. The cyclic gestures instantiate the schema of CYCLE, which comes from numerous real-world phenomena such as the rolling wheel, changing seasons, the sun rising and setting, etc. (Ladewig, 2011).

There are some points worth clarifying in terms of gesture and metaphor. Firstly, as our physical experiences can lead to the instantiations of certain metaphors in gesture, speech can also influence metaphoric representations in gesture (Gu, Mol, Hoetjes, & Swerts, 2017; Núñez & Sweetser, 2006). A case in point is the recurrent up-pointing gesture performed by Chinese speakers when they utter the phrase "last week" (Shang Zhou 上周), as the two Chinese characters are equivalent to "Upper Week" (Gu et al., 2017). Secondly, whereas gestures are robust in terms of projecting the source domain by describing appearance (e.g. shape and size), enacting actions (e.g. beating, weighing), using directions (e.g. pointing up/down) and assigning space (e.g. bounding space) (Chui, 2011), they are not appropriate for reifying some abstract domains, such as the COLOUR metaphor (WHITE IS RIGHT, BLACK IS WRONG). To represent the COLOUR metaphor, speakers are more likely to use the hand to separate the space in front of them, metaphorically referring to the two moral concepts (Cienki, 2008). Thirdly, although the examples showed are all metaphoric gestures (in McNeill's categorisation, refer to 2.4.2.), there is no necessary link between metaphoric gestures and gestures with metaphor (Cienki, 2008). It is possible that there are metaphors in some iconic gestures and no metaphor in some metaphoric gestures (e.g. beat gestures). Finally, as the study

of metaphor originates from written language, the list of schemas available (e.g. GOOD IS UP, IDEAS ARE OBJECTS, etc.) may be insufficient to describe metaphors in gestures. New schemas will need to be generated based on metaphors emergent in gesture. For instance, as most English speakers write from left to right, they tend to refer the left to the past and the right to the future, which gives rise to the schema of LEFT IS PAST, RIGHT IS FUTURE (Calbris, 2008).

2.7.2. Analogic links in gesture elements

The theories and approaches regarding identifying and justifying the analogic links between gesture forms and their associated meanings are proposed and demonstrated by Calbris (2011) and are essentially the same as the theories regarding gesture and metaphor. Taking a cognitive semiotic approach to gesture, the most significant contribution of Calbris' works is the innovative and systematic detection and description of the meanings of gesture parameters (Calbris, 2003, 2008, 2011). She stresses the role of gesture as a self-reliant symbolic system that interacts with speech and mind, and she regards bodily movements as the representations of the mind, which provides a window for investigating the ways in which thought is rooted in the perceptual experiences of the body in the physical world (Chui, 2009; Kita, 2009; Simpson-Vlach & Ellis, 2010; Stam, 2006). Hence, similar to theories in gesture and metaphor, the meanings of gestures stem from recurrent day-to-day bodily experiences. However, Calbris' paradigm is more fine-grained as she segments gestures into components and examines the analogic links between gesture elements and their semantic variations based on context analysis (Calbris, 1990, 2008, 2011).

Calbris (2011) coins the term action schema, referring to recurrent chains of actions in the real world consisting of a subject (the actor), a tool involved,

an action and an object. Action schemas can be projected to concrete actions, which the schema may originally come from, or from another domain metaphorically. In each action, she emphasises the existence of multiple form-meaning analogic links, which can be activated by context and the close relationship between the constituents in each action. The instance, the action schema derived from the activity of fishing (Calbris, 2011) is claimed to be composed of different elements including the person, the rod, the action of holding the rod and fishing, and the fish being drawn to the bait attached to the fishing line, etc. Based on the context, the gesturer can potentially refer to any element in this schema while performing the fishing action.

Taking another more frequently seen gesture as an example, Calbris (2011) illustrates the multiple analogical links in the transverse movement of the level hand (Figure 2.10). The transverse movement of this is composed of five elements, i.e. the movement, the fingertip that draws, the palm that resists, the edge that cuts and the palm that covers.

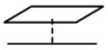
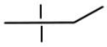
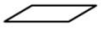
Gesture	Transverse movement of the Level Hand				
Relevant feature	Movement makes relevant:	the fingertip that draws	the palm that resists	the edge that cuts	the flat palm that covers
Analogical links	*Everywhere*	*Straight line*	*Obstacle*	*Cut*	*Horizontal plane*
					
Meanings	Quantity Totality	Directness	Stop-refusal	Cutting	Flat surface
Semantic derivations	Superlative Perfection	Determinism Certainty Obligation Frankness	Negation Refusal End		2nd Surface (same) Level Equality Stability

Figure 2.10: The analogic links of the transverse movement of the level hand (Calbris 1990: 140, see in Calbris, 2011).

The five elements correspond with five analogic links and five semantic variations respectively. Depending on the discourse context, different form-semantic links can be triggered. For instance, in the case of the utterance

"you're absolutely right", the first analogic link between the movement and meaning of totality is activated. However, if a speaker is saying "cutting", then the fourth analogic link would be the most relevant.

In conclusion, the conceptual metaphor theory and action schema theory are fundamentally the same. They both take a cognitive perspective and hold that the form-meaning associations in recurrent gestures are deeply rooted in the real-world experiences of the body. The findings of the current thesis can also provide some implications on this, which will be presented in Chapter 6.

2.8. Towards the multimodal units of meaning

As the main goal of this thesis is to use multimodal corpus-based approaches to explore the multimodal units of meaning, the final section of this chapter will introduce some studies which explore the concept of the multimodal units of meaning in speech and gesture from different theoretical perspectives. For instance, advocating a multimodal approach to grammar, Fricke (2013) establishes a useful criterion for distinguishing multimediality and multimodality, with the former simply referring to the simultaneous occurrence of more than two modes, whereas the latter requiring structural and functional integration between different modalities with a deep embedding relationship rather than just flat, sequential combination. She emphasises that the prerequisite of speech-gesture grammar is the identification of recurrent, gesture-meaning correspondence in various uses of contexts. To illustrate, Fricke (2013) demonstrates her findings of the syntactic and functional integration of speech and gesture at an integrating point provided by the German article *son* ('such a'). She explains how the gestures (mostly iconic and deictic gestures) accompanying *son* not only instantiate the qualification of the noun phrases, but also formulates an integral part of the syntactic structure. Based on such observations, she establishes a grammatical profile

for the German article *son*.

From the ethnographical perspective and based on extensive observations of real-life communication, Enfield (2013a) proposes the concept of 'Composite Utterances' as the products of cooperation and negotiation in social interaction, and he specifies *move* in as the basic element in the unfolding of discourse. A multimodal move could be in the form of any descriptive unit rooted in user-based, context-sensitive theoretical and methodological tradition such as a turn-construction unit in conversational analysis, a growth point (McNeill, 1992) or a multimodal ensemble (Kendon, 2004), etc. Enfield (2013a) emphasises the hybrid nature of composite utterances as combining conventional and non-conventional signs based on his research on pointing gestures. He exemplifies this with a multimodal deictic utterance composed of a conventional deictic gesture, a conventional pronoun and a non-conventional, context-based object.

It is worth noting that Enfield's concept of *move* as a multimodal utterance differs from the long-standing term 'move' in discourse analysis, originally developed by Swales (1990) to help non-native students analyse the different structures of texts of different genres (e.g. newspaper articles, journal articles). Adopting a top-down analytical approach, move analysis focuses on the construct of the different move types of texts in a genre. In other words, a text is explored in terms of the sequence of 'moves', with each move representing a stretch of text and serving certain communicative function. For instance, in research articles, texts may begin with a move that specifies the topic and reviews existing research, followed by a move that identifies a gap, followed by a move that states the goals of the current research, followed by a move that demonstrates the research methods, and so on. Such move types will be different in another genre, such as newspaper reports, advertisement or classroom student-teacher interactions, etc. In a spoken text, a move can be comprised by an utterance, such as a teacher asking a question in class. However, in general,

Swales' (1990) well-established concept of move contains a much more complex and broader sense than Enfield's. It can be applied to both spoken and written texts, and also to texts of various lengths. Moreover, repetitive observation and analysis of many texts are required in order to identify and describe the patterned uses of moves in each genre, and the identified sequence of moves can then be applied as a framework for analysing texts of the same genre.

Differentiated from the two concepts above, the research on Multimodal Construction Grammar (Steen & Turner, 2013) adopts a corpus-based approach. Rooted in cognitive linguistics, the concept of 'construction' is defined as a pre-packaged unit of form-meaning association stored in the mind and usually composed of multiple modalities including texts, gestures, prosodic features, etc. (Steen & Turner, 2013). This concept is discussed by them (Steen & Turner, 2013) drawing on a large dataset of global news broadcasts, namely the Red Hen corpus, in which mainly speech and gesture have been annotated and analysed in order to demonstrate the ways in which they blend into one unit and contribute to meaning. In the same vein, a study of multimodal constructions was carried out by Zima (2014) based on the Red Hen corpus. Her research focuses on the coordination of gestures with four motion events, including *in circles*, *spin around*, *zigzag* and *all the way from X PREP Y*. Results show that 60% to 80% of the instances in the four types of constructions were associated with certain recurrent gesture forms, the percentage of which are claimed as a sufficient indication of the entrenchment of those multimodal units in the mind.

Zima (2014) correctly points out that the issue of finding a threshold for the frequency of speech-gesture co-occurrences in order for them to qualify as a pre-stored meaning construction requires further research. Also, the relatively small number of instances in each construction type (slightly more than 100) indicates the need for further research with a greater number of

instances in order to replicate the research and assess the findings. Despite those concerns, a corpus-based, comprehensive exploration of the interplay of modalities in multimodal constructions is of considerable value when it comes to capturing a more nuanced picture of Multimodal Construction Grammar and the multimodal units of meaning in general. A more fine-grained analysis could include whether and how the variations of gesture parameters (e.g. space, handshape, trajectory, velocity, etc.) correlate with the variations of meaning constructions.

Until the present, one of the most carefully designed corpus-based research on the functional relationship between speech and gesture is that relevant to head nods and response tokens (Adolphs & Carter, 2013; Knight, 2011; Knight, Evans, Carter, & Adolphs, 2009b). The research categorised all the head nods co-occurring with response tokens into five typologies based on their motions: small nods of short duration (type 1); small, multiple nods of longer duration than type 1 (type 2); intense nods of short duration (type 3); intense and multiple nods of longer duration than type 3 (type 4); multiple nods, comprising a combination of type 1 and 3, of longer duration than type 1 and 3 (type 5). The co-occurrence between the five types of head nods and the four functions of speech response tokens can then be analysed. The four functions of speech response tokens were derived from the classification scheme developed by O'Keeffe & Adolphs (2008) based on a large spoken corpus, which includes Continuers (signalling the speaker to continue the turn), Convergence (showing agreement), Engaged Response Tokens (emphasising emotions in response) and Information Receipt Tokens (marking shifts of topics in discourse).

Results show that type A nods tended to co-occur with those response tokens as Information Receipt Tokens, whereas Type B nods were likely to coincide with those Convergence Tokens and Type D with Continuers and Convergence Tokens. The researchers suggested that such correlations

between head nods and response tokens indicate the value of head nods in disambiguating the function of poly-functional response tokens. They also suggested defining the function of multimodal response tokens on a continuum rather than on discrete poles as the intensity of head nods could alter their function in discourse. In addition to the notable implications above, individual differences in the uses of head nods and response tokens have also been found in the participants, which indicates the idiosyncratic feature of speech and gesture production.

The current thesis adopts a similar multimodal corpus-based approach to the research on head nods and response tokens, and attempts to take the challenge further by examining hand gestures. Compared to head nods, hand movements are more complex in form and more challenging in terms of annotating, analysing and identifying patterns. Investigating speech and hand movements can not only further our knowledge of the interplay between speech and gesture, but also shed new light on the relationship between the hand and human cognition (Cienki, 2008; McNeill, 2005).

2.9. Summary

This chapter reviewed existing theories, concepts and research that are particularly associated with, and insightful for, the current thesis, i.e. exploring a multimodal corpus-based approach to investigate the multimodal units of meaning in speech and gesture. It began with a general overview of the two predominant domains of research in multimodality, i.e. multimodal written discourse analysis of texts and image, and research on speech and gesture. It showed that the current thesis differs from most previous research paradigms in terms of combining qualitative and quantitative approaches. Following the introduction of multimodality was a comprehensive evaluation of spoken multimodal corpora, which is vital for supporting multimodal

research on speech and gesture; however, the assessment showed the limitations of the existing multimodal corpora in representativeness, size, machine-readability and openness, hence being unable to assist in a wide range of multimodal research.

Mainly relevant to the first study (i.e. exploring gesture patterns co-occurring with formulaic sequences), the section on formulaic sequences primarily focused on providing a rationale for using frequency as the method for identifying formulaic sequences, and its limitations. It also reviewed research that provided evidence of the processing advantage of formulaic language, which points towards their holistic processing in the mind. In particular relation to the two issues, the first study can shed some new light on how gesture can demarcate formulaic sequences and give evidence for the holistic processing of them. Details about these will be provided in Chapter 4.

With particular relevance to the basic theories and frameworks in gesture studies, the chapter also demonstrated how gestures have been defined, categorised (McNeill's framework) and segmented (e.g. gesture phases) by previous researchers. The other relevant issues reviewed were the ways in which gestures are related to speech temporally and semantically, and the availability of multimodal software for analysing speech and gesture. The definitions, frameworks and theories demonstrated in this section are particularly significant as they lay the foundation for the description and analysis of gestures in this thesis. On the other hand, this thesis contributes to the existing knowledge by building up a more comprehensive analytical framework (see Chapter 3, 3.3. and 3.4. for details of gesture coding scheme) and providing implications for the temporal and meaning coordination between speech and gesture.

The introduction of the existing research and knowledge regarding gesture studies was followed by a review of research investigating the variations of form-meaning associations of frequently occurring gestures, i.e.

emblems, gesture families and recurrent gestures. The research on the cyclic gesture is particularly relevant to the second study of the circular gestures and has been fully introduced and evaluated. The review of the three strands of research indicates the lack of strong empirical grounds for their findings, and the current thesis, i.e. a multimodal corpus-based research, can provide some answers in that regard.

The chapter also reviewed four major functions of gestures, including demarcative, referential, pragmatic, and discourse marking. Based on the four functions, the thesis intends to establish a more comprehensive framework for describing the discourse functions of gestures in relation to speech contexts. Moreover, Chapter 6 will further discuss in more detail the ways in which the thesis can provide implications for other debates concerning the functions of gestures, i.e. whether they are primarily for speech production or communication, and the interactive nature of some gestures (termed interactive gestures).

The chapter also addressed the issue of how to justify the form-meaning associations in recurrent gestures based on cognitive linguistic theories, i.e. gesture and metaphoric thinking, and analogic links in gesture elements. The current thesis can also be insightful in this aspect, which will be illustrated in Chapter 6. The final section reviewed previous research on different linguistic concepts which also explored the multimodal units of meaning in speech and gesture. These are multimodal grammar, Composite Utterance, Multimodal Construction Grammar, and multimodal response tokens. The review of these concepts showed that, although the multimodal units of meaning have been investigated previously, the current research is novel in terms of using formulaic sequences and circular gestures as the main research targets.

Chapter 3. Data and framework development for speech-gesture alignment

3.1. Introduction

As the thesis intends to explore the emerging patterns of the multimodal units of meaning in speech and gesture, it is crucial to segment gestures into phases and identify the exact alignment between the strokes (i.e. the phase in which the meaning of a gesture lies) and the speech. Following the review on the existing multimodal corpora, gesture analytical frameworks and tools in the previous chapter, this chapter aims to demonstrate the datasets used in the current study and the frameworks constructed for segmenting gestures into gesture phases, together with the format for transcribing speech and gesture in accordance with the research purposes of the thesis. As the two studies of the thesis use different aspects of multimodal units as the point of departure, i.e. the first study uses recurrent speech and the second uses recurrent gesture, it is necessary to adopt different analytical procedures and tools, these will be described in the approach section in Chapters 4 (4.2.) and 5 (5.2.), respectively. However, the two studies share the same corpus (the Nottingham Multimodal Corpus, NMMC), the same gesture coding scheme and the same transcription principles, which will be introduced here.

3.2. An overview of NMMC

The primary purpose of section 3.2. is to provide key information regarding the multimodal corpus used by the thesis, the Nottingham Multimodal Corpus (NMMC), and justify the reasons for it being a suitable data source for the current research. It will begin with a description of some details of the two sub-corpora (lectures and supervision meetings) of the NMMC in 3.2.1., and

then explain why supervision meetings were chosen as the target of analysis. The sub-section 3.2.2. will introduce the natural and uncontrolled set-up of the supervision meetings, including the location, the equipment layout, the participants and the researcher. Information concerning existing speech transcripts in the NMMC and how it serves the purposes of the current thesis will be illustrated in 3.2.3., which will be followed by the ethical issues concerning the NMMC in 3.2.4. It will be demonstrated how the current PhD benefits greatly from the speech transcripts offered by the NMMC, and the ethical approval received by the end-user.

3.2.1. The sub-corpora of NMMC

As is widely acknowledged, compared to monomodal corpora, constructing a multimodal corpus of real-life communication is particularly challenging and time-consuming (Adolphs & Carter, 2013; Knight, 2011). The NMMC is one of the few existing large audio-video collections of natural human communication, although it has not been released to the public for copyright reasons (refer back to Chapter 2, 2.2.2. for an overview of the current multimodal corpora). It was the outcome of the combined effort of many researchers and postgraduate students at the University of Nottingham over three years from 2005 to 2008. As a part of an ESRC-funded project (reference PTA-030-2004-00942), the construction of NMMC was particularly designed to support the examination of speech, head nods and hand gestures, as well as exploring good practices of multimodal corpus construction. It was intended to expand corpus-based research from monomodal to multimodal by developing new paradigms, analytical tools and research opportunities for (applied) linguists.

The entire corpus comprises 250,000 words, the amount of which is almost evenly contributed by 13 monologue lectures and 13 dyadic

supervision meetings. As is presented in the tables below (Table 3.1 and Table 3.2), the length of the lectures is about one hour on average, containing around 5,500 and 12,500 transcribed words. The length of the supervision meetings mostly ranges from about 30 minutes to 90 minutes (except for video 10), with about 5000 to 16,000 words.

Table 3.1: An overview of the lectures.

Video	Length	Word count	Participants	Subjects
1	58:03	5588	1 male lecturer	Sociology
2	1:14:13	9655	1 male lecturer	Engineering
3	40:58	6538	1 male lecturer	Drama
4	1:00:43	9137	1 female lecturer	Marketing
5	53:31	10092	1 male lecturer	Business
6	1:21:34	12596	1 male lecturer	Linguistics
7	55:08	9054	1 female lecturer	Linguistics
8	58:33	8436	1 male lecturer	Literature
9	53:54	9958	1 female lecturer	Literature
10	1:09:31	13141	1 male lecturer	Linguistics
11	1:02:03	11591	1 female lecturer	Literature
12	1:06:17	10061	1 female lectures	Literature
13	1:01:42	9527	1 female lecturer	Linguistics
Totals	13:22:70	125,374		

The NMMC corpus can be regarded as a high-quality, specialised/special corpus (Thompson, 2010; Tognini-Bonelli, 1996) of academic talk. The lectures contain 13 participants with 7 male and 6 female speakers. The supervision meetings were conducted mainly between PhD students and supervisors in the School of English at Nottingham, including 8 male and 10 female participants. As can be inferred from the numbers, four speakers

participated more than once in the supervision meetings corpus. With the exception of the MA and PhD students, most participants tend to be middle-aged or older.

Table 3.2: An overview of the supervision meetings in NMMC.

Video	Length	Word count	Participants	Subjects
1	35:52	9,218	1 female supervisor + 1 male PhD student	Linguistics
2	1:00:27	11,875	1 male supervisor + 1 male PhD student	Linguistics
3	41:41	9,513	1 male supervisor + 1 female MA student	Linguistics
4	58:33	10,103	1 male supervisor + 1 male PhD student	Linguistics
5	1:28:22	16,560	1 male supervisor + 1 male PhD student	Drama
6	47:52	11,705	1 female supervisor + 1 female PhD student	Literature
7	1:16:31	6,755	1 female supervisor + 1 female PhD student	Linguistics
8	1:41:05	6,284	1 female supervisor + 1 female PhD student	Linguistics
9	57:17	8,073	1 female supervisor + 1 female PhD student	Linguistics
10	14:10	4,642	1 female supervisor + 1 female PhD student	Linguistics
11	43:03	11,726	1 male supervisor + 1 female PhD student	Linguistics
12	55:58	11,963	1 male supervisor + 1 female PhD student	Linguistics
13	37:58	10,194	1 male supervisor + 1 female PhD student	Sociology
Totals	11:93:09	128,611		

It has to be admitted that, as the major goal of the NMMC was not to support sociolinguistic research, it is limited in terms of its representativeness of the population in their genders, ages and social backgrounds, etc. As an academic spoken corpus, it mainly represents lecturers and supervision meetings as two types of academic talks in arts and humanities (mostly linguistics) at the University of Nottingham. Hence, while the NMMC does not represent other types of academic interactions (e.g. seminars), disciplines (e.g. science) or

universities. This is especially true for the supervision meetings sub-corpus, most of which were collected from the School of English at Nottingham between PhD students and their supervisors in the field of linguistics. Hence, the NMMC can support an investigation of academic communication in such context, such as lectures in humanities or PhD supervision meetings in linguistics, it would not be suitable as a data source for studies of interactions in other contexts of academic interactions. Despite its limitations, the NMMC is an ideal resource for multimodal research, such as those investigations conducted by its founders and the current PhD project. That is, as the goal of the thesis is not to generate findings applicable to a large population, but to explore new approaches to investigate the speech-gesture relationship, the NMMC is an ideal dataset in this respect.

As a corpus with high sound-and-image quality, the NMMC has played a significant role in exploring multimodal corpus-based approaches for conducting linguistic research that integrates both verbal and nonverbal modes of communication, including speech, prosody, head nods and gesture (Adolphs & Carter, 2007; Carter & Adolphs, 2008; Knight, 2011). Since its construction, numerous researchers and PhD students at Nottingham have drawn on it, either as a major dataset or a comparative corpus. For instance, the developers of the NMMC used it as the basis for exploring the functional relationship between head nods and response tokens, meanwhile, developing a multimodal analytical software and automatic head movement tracking equipment (Knight, 2011; Knight & Tennent, 2008). Following this, a further two PhD projects examined the holistic processing of formulaic sequences from the perspectives of pauses (Dahlmann & Adolphs, 2009) and prosody (Lin, 2012). Subsequently, in another PhD project, some videos from the NMMC were compared with another dataset collected from academic meetings between British supervisors and Japanese students (Tsuchiya, 2013). The general goal of that thesis was to run a comparative analysis of the ways in

which native British speakers and Japanese learners adopted response tokens, head nods and gestures as strategies in the transitional time period from listeners to speakers.

Despite all the original works undertaken by the previous researchers, the NMMC has not yet been fully explored, especially its value for multimodal speech and hand-and-arm gesture analysis. None of the aforementioned projects have adopted such a fine-grained analysis as is being undertaken here. The current thesis uses the supervision meetings owing to their superior sound-and-image quality, i.e. the lectures understandably contain a higher level of noise compared to one-to-one supervision meetings. A further reason for not using lectures is relevant to the research interest of dyadic communication. Therefore, in the following sections, more information regarding the supervision meetings will be provided.

3.2.2. Setup of NMMC

As corpus linguistics generally emphasises the study of language in real-life use (Kennedy, 1998; McEnery & Wilson, 1996; Sinclair, 1991), the level of naturalness is an important element for both monomodal and multimodal corpora as, in principle, corpora are supposed to reflect authentic language uses in various socio-cultural contexts. As many previous speech-gesture studies were conducted for psychological or computational purposes, they tended to obtain the datasets from a lab-based environment (Gullberg & Kita, 2009; Kita, 2009; McNeill, 2005), sometimes even with equipment such as data gloves/stickers, eye tracking devices, etc. In addition, the language elicited in most previous research was usually the product of controlled tasks such as narrating a pre-assigned video or playing a game. Hence, the naturalness of the datasets generated from those tasks are largely compromised. Despite the valuable implications drawn from that

psycholinguistic and computational research, the implications for speech and gesture uses in natural human communication is rather limited.

The NMMC is one of the few corpora that strives for the compilation of a dataset that reflects real-life human communication. The lectures were all delivered to a group of listeners in real seminar or lecture rooms, and the majority of the supervision meetings took place in real offices. Although there are a few videos of supervision meetings recorded in a lab, it could be argued that the authenticity of the supervision meetings might not be significantly affected providing the contents of the conversations have not been controlled. It is not uncommon nowadays for people to conduct supervision meetings out of the office for various reasons, such as in a coffee bar or a lab. Whereas the issue of how the change of environment could change the dynamic of interactions between the supervisors and students is still unknown, it can be assumed that the authenticity of those supervision meetings might not be greatly compromised by the change of environment.

A greater issue for the naturalness of the data concerns the Observer's Paradox (Labov, 1972). Although the researchers of the NMMC were not present and observing during the lectures or supervision meetings, the fact that there were a camera and a microphone recording facing the participants may well affect their behaviour to some degree. It is understandable that participants might feel uneasy and/or uncomfortable in front of the camera. It is hoped that the participants might become more accustomed to the research environment as the recording continues. In addition, as hand-and-arm movements are normally hard to control, although the participants might say something for the sake of the camera, it would be difficult for them to control their gestures. It is possible that people might gesture in a larger size or more frequently under pressure, but size and frequency are irrelevant to the current research questions. As the thesis aims for specific gesture forms that reflect the semantic elements of the gestures, they are much less likely to be changed

as the meaning of the gestures are deeply rooted in human cognition (Calbris, 2011; Ladewig, 2011; Lakoff & Johnson, 2008; Mittelberg, 2014).

As the construction of the NMMC aims for allowing the study of not only head movements, but also gestures, the set-up of the camera also reflects this. Figure 3.1 below illustrates the positioning of the participants, the two DV cameras and the microphone (Knight, 2011). The participants sat facing each other with their upper-front body recorded by an individual camera, which provided a clear view of the hand-and-arm movements. The microphone was placed between the speakers to obtain high-quality sound. The video files captured by the two cameras and the audio file recorded by the microphone were then all synchronised in Adobe Premiere with the final aligned image demonstrated in the Figure 3.2 below.

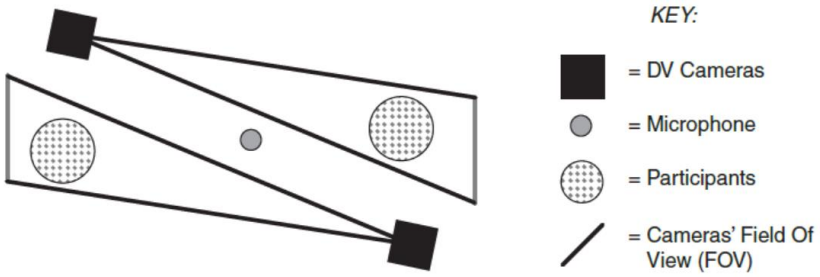


Figure 3.1: The set-up of camera in NMMC (Knight, 2011).



Figure 3.2: A screen shot of image in NMMC.

The NMMC did not use any tracking devices such as the SMI eye-tracker

(https://www.smivision.com/eye-tracking/software-downloads/) to gain data on eye fixation on gestures (Gullberg & Kita, 2009). Nor did it use a motion capturing device (Pfeiffer, 2013b) or data gloves (Pfeiffer, 2013a) to track trajectories, space, velocity and other performance indices. As the uses of these devices require the participants to wear sticky markers and/or wear gloves, they are extremely intrusive. Motion tracking technology is mostly applied in research concerning human-computer interaction, especially in the area of virtual reality research.

A point worth mentioning is the fixed tri-pods used in the NMMC, which do not allow the capture of movements of the participants out of the view of the cameras. Although this is not a problem for the current research context as participants barely move around in the supervision meetings, there was one instance in a meeting where the supervisor asked the students to draw something on a piece of paper. The student then knelt down and drew the picture on the floor, and the camera was unable to capture the drawing process or the picture. As the picture later became the focus of discussion, the understanding of that part of conversation is largely impeded.

On the whole, the NMMC might not be suitable for those research projects that target at eye-gaze and/or precise data of the gesture parameters (e.g. velocity, space); however, for the current thesis, the NMMC has a sufficient level of image/sound quality and authenticity.

3.2.3. Speech transcripts and representation in NMMC

The NMMC is one of the few large multimodal corpora that has been fully transcribed, which allows automatic textual search and analysis in any word processing software. As an eligible user of the NMMC, the researcher of the current thesis has access to the videos and speech transcripts of the corpus, although the gestures have not been annotated or transcribed. The lectures and

supervision meetings were systematically transcribed, checked, time-stamped and anonymised by highly trained researchers, and all this information was recorded and provided to the users of the corpus. As the process of speech transcription would require enormous time and effort, the transcripts became one of the greatest advantages for the current thesis.

In general, there are two ways of representing spoken language. It can either be transcribed orthographically or by the adoption of the IPA system (International Phonetic Alphabet) (Ball & Rahilly, 1999, p. 226). The NMMC chose the former approach for better readability, although the IPA approach is recognised as a more faithful and detailed representation of the phonetic and prosodic features of speech. Understanding the IPA conventions requires training and practice, which would make a corpus less accessible to many researchers and readers. For those researchers who are specifically interested in the phonetic aspects of speech, the IPA system might be more appealing, but, for most linguistic projects focusing on texts, the orthographic transcripts are simply much more efficient to work with.

There has been a long tradition of transcribing speech in conversational analysis to assist analysis, although transcripts cannot be seen as a substitution of actual audio data (Have, 1999; Richards & Seedhouse, 2007). Some transcription systems are widely adopted by numerous researchers, especially those of Gail Jefferson (2004) and Atkinson and Heritage (1984). Admittedly, transforming acoustic files into textual materials inevitably compromises the authenticity and loses some integral features of the spoken language; however, due to the limited development of automatic sound recognition and analysis technology, at the present time, transcription is a necessary step in order for a spoken corpus to be read and processed automatically.

However, those transcription systems developed in CA were not transferable to the NMMC (Knight, 2011) and the developers decided to follow the conventions set by the Cambridge and Nottingham Corpus of

Discourse in English (CANCODE) (refer to Appendix I for speech transcription conventions) (Adolphs, 2008, pp. 137–138). The CANCODE convention strikes a good balance between representing the spontaneity of interaction and readability. As the transcribing procedures of a large-scale spoken corpus such as the NMMC usually involve a number of researchers, the developers set up guidelines for the level of granularity and spelling at the very outset of transcription in order to ensure consistency. For instance, the spelling of vocalisations such as *mmm*, *mm*, *mmmm* and *mhm*, and expressions such as *going to/gonna* have to be decided beforehand (Knight, 2011). The consistency in such details in spelling is crucial for running corpus-based lexical search and analysis.

With a range of transcribing tools available (refer to Chapter 2, 2.4.5. for an overview of existing multimodal tools), the NMMC used Transana (https://www.transana.com/) for transcribing speech. As is noticeable from the episode below (Figure 3.3), interactions were time-stamped in the software on a turn-by-turn basis.

¤<14823><$M> +and I cross referenced that with a book by Leech and a few others and that it's literally just frequency lists from the BNC+

¤<23750><$F> Uh-huh.

¤<24257><$M> +and so I just looked at the McCarthy list looked at the Leech list and um looked picked out <$=> which were <\$=> which are the ones that would be indicative of spoken lexis.

¤<35648><$F> Uh-huh.

¤<36062><$M> <$E> sigh <\$E> The problem with that was the on= <$=> they didn't get very <\$=> I got far with it+

¤<42874><$F> Mmm.

Figure 3.3: An episode of the transcript in NMMC.

Admittedly, the time-based, linear mode of representation of daily communication has its limitations as people do not often speak in an as orderly fashion as is shown in the transcript. Features such as overlaps and interruptions (both marked as "+" in NMMC) are so prevalent in real-life interactions that this mode seems insufficient to represent these phenomena, especially when there are more than two interlocutors. However, for the current research questions, this does not present any problems as turn-taking is not the focus of the thesis. The level of detail provided by the NMMC is also sufficient for the current study, though minor revisions were made in the transcription reflecting the purposes of the thesis (i.e. marking up the exact length of pauses where necessary).

3.2.4. Ethics

The challenges regarding ethical issues are another major concern for corpus compilers and end users. It is of vital importance that every corpus linguist consider all the legal requirements and moral issues while compiling and working with a corpus (Enfield, 2013b; Kennedy, 1998; McEnery & Wilson, 1996). The developers of the NMMC propose a general structure of ethical guidelines at three levels, i.e. institutional, professional and personal (Knight, 2011). Institutional requirements may come from the Ethics Committee of a certain university; professional guidelines can be set by a certain discipline or funding body; personal concerns are largely related to the needs of individual participants and the researchers' perspectives. This is a good framework for any researcher to refer to when considering both legal and moral issues throughout the data collection, analysis and representation procedures.

Undoubtedly, it is more difficult to recruit participants that are willing to be recorded by cameras than those for audio recordings (Adolphs & Carter, 2013). Moreover, researchers working with multimodal corpora will face

more ethical problems due to the nature of the datasets. To compile a textual corpus (including spoken corpora that only provide written transcripts), it is relatively easier to receive consent from participants, authors or copyright holders to use the data for research and publication purposes. It is also easier to anonymise the identity of the participants in a textual corpus, which largely involves altering any names, phrases, topics, images, or any information that could potentially reveal the identity and/or privacy of the subjects.

The anonymising process is much more complicated for spoken multimodal corpora. To protect identities, the sound streams can be tempered and the images pixelated, however these methods will greatly compromise the naturalness and authenticity of the data. The same consequences can also result from many other anonymising methods such as using an actor's voice, hiding the eyes of the speakers or hiring someone to draw pictures of the screen shots. In addition, further problems can arise when the communication context involves other speakers who are not the participants in the research. The researcher will need their consent or anonymise them afterwards when using the data.

Owing to the potential difficulties of anonymising multimodal datasets, the most efficient approach is probably to receive permission from participants in advance for a variety of purposes (e.g. research, publication and presentation, etc.) (Knight, 2011). The idea is that, if the participants are willing to be recorded, there is a high chance that they will be open to the idea of exposing themselves to a wider audience. This is the approach adopted by the developers of the NMMC (refer to Appendix II). There were three different levels of consent for participants and almost all agreed to provide the highest level of consent. That is, the participants of the NMMC not only allowed the researchers to use the data for internal purposes, but also for them to present still and moving images with voice in public talks and any form of publication. With consent from one of the developers of the NMMC, all end

users have access to the same rights (refer to Appendix III for a signed consent form for researchers). The current thesis takes great advantage of these permissions, presenting video clips and images in international conferences, workshops and forthcoming publications.

3.3. Gesture phases and coding procedures

Chapter 2 (2.4.3.) provided an overview of the most important frameworks for distinguishing and segmenting gestures into different phases, mostly proposed by Kendon (2004), McNeil (2005) and Kita's team (Kita et al., 1998). This section will now demonstrate how these frameworks were adapted in the current thesis, and the overall procedures for segmenting gesture phases. The current thesis used the ELAN[3] software (Wittenburg et al., 2006) for both studies in order to segment and annotate speech and gesture. More details will be provided in Chapter 4 and Chapter 5.

The current thesis used a five-phase framework, namely, preparation, pre-stroke hold, stroke, post-stroke hold and retraction. During the segmentation processes, the gestures were mainly segmented in mute in ELAN without the transcript in order to avoid the influence of speech (details of the process will be provided in the latter part of this section). This decision was made based on the criticism from McNeill (1992), concerning the problem of circularity in the processes of segmenting gesture phases with reference to speech and then investigating the speech-gesture relationship. In other words, in order to generate more reliable findings of speech-gesture alignment, the segmentation processes of gestures need to avoid the influence of speech as far as possible. Prior to the analysis of the targeted gestures, 5 randomly selected gestures (25 in total) were segmented in mute from

[3] https://tla.mpi.nl/tools/tla-tools/elan/download/

approximately the beginning, middle and end parts of the chosen supervision meetings (V2, 3, 4, 11, 12) to test the feasibility of the method. The results show that, once the target gesture is located, it is highly possible to demarcate the gesture phases correctly by observing the movements in mute.

Following previous researchers (Kendon, 2004; Kita et al., 1998; McNeill, 2005), this thesis defines the stroke phase as the only phase that is compulsory for a gesture to exist and is closest to the apex of the entire movement. In other words, the main criteria for coding the stroke for the current research was the efforts in the movement rather than the discourse context as the researcher was coding in mute. The phase of stroke hold termed by McNeill (2005) (e.g. holding the hand to imitate a floor) is not used here as coding a stroke with an unmoved hand would require the assistance of the co-occurring speech. If such an instance did occur, the preceding movement would be coded as the stroke (e.g. lifting the hand from a lower position), and the holding part would be a post-stroke hold phase (this will be explained later). Moreover, Seyfeddinipur (2006) proposes the term, interrupted stroke, but this is used here. If the movements were interrupted by other gesture phases and end abruptly, these would be coded into two separate strokes.

The preparation phase is the phase where the moves lead the hand to the starting point of a stroke (Kendon, 2004; Kita et al., 1998; McNeill, 2005). It can depart from a rest position (where the hand is relaxed such as on the lap, the arm of a chair or a desk, etc.), or begin instantly after the completion of a stroke when the hand is only partially relaxed before going to a potential rest position. It can even begin immediately after the last stroke with no other gesture phases occurring in between, i.e. the hand performs continuous strokes. As mentioned by Seyfeddinipur (2006), a preparation phase can be interrupted, which can be termed an interrupted preparation; however, the phenomenon did not occur in the coding and, therefore, the concept has not been adopted. Normally, if a preparation phase is interrupted (e.g. the hand relaxes or

retracts), the preparation phase can be coded starting from where it resumes. Probably for this reason, other researchers have not included an interrupted preparation phase in their coding frameworks either.

The holding phase was described and coded differently among previous researchers. Kendon (1980) only used the term, hold, to account for the stage where the hand remains still at the end of the stroke. Later, Kita refined Kendon's framework by providing a more comprehensive description of holding and proposing pre-stroke and post-stroke hold phases (Kita et al., 1998). They acknowledge that the hand might not be completely frozen during the holds; it might be slightly drifting, as we are not machines. The pre-stroke hold phase (Kita et al., 1998) emerges when the hand remains unmoving after reaching the starting point of a stroke, i.e. the delayed period from the end of the preparation phase to the moment when the stroke movements actually start. On the contrary, if the hand reaches the end of a stroke, but holds for a while, a post-stroke hold phase (Kita et al., 1998) occurs. The current thesis adopted the definitions of pre-stroke and post-stroke holds as defined above, which works well with the current analysis. Such a coding does not require the consideration of the speech context such as Kita's (Kita et al., 1998) another type of hold, the dependent hold (which is similar to McNeill's stroke hold above).

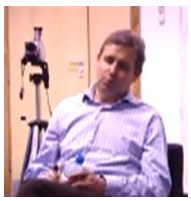
The retraction phase has been defined differently by from previous researchers. Here, it is the phase where the hand relaxes and/or goes into a resting position after the stroke. Kita et al. (1998) propose partial retraction to account for those retractions that only temporarily relax and do not reach an actual rest position. Considering the research questions of the thesis, such a distinction would serve no purposes here. Hence, all partial retractions were coded as retractions.

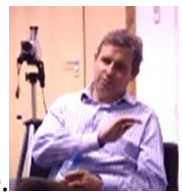
To illustrate the five phases, a randomly selected example from Video 2 is provided below:

Rest position (a): both hands stay in a rest position, resting the hands on the lap.

a. 

Preparation (a-e): The left hand remains in the rest position. The right hand gradually raises to the left side of the chest while opening the palm until it gains more tension with the palm facing down. Position e is the end of the preparation phase and the starting point of the stroke.

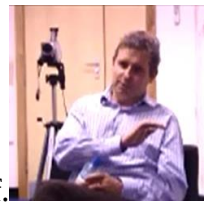
a.  b.  c.  d. 

e. 

Pre-stroke hold (e-f): Both hands maintain the same position before the execution of the stroke.

e.  f. 

Stroke (f-i): The right hand rapidly moves outwards towards the up-right space of the body departing from the left side, maintaining the same palm-down hand shape.

f.  g.  h.  i. 

Post-stroke hold (i-j): The hand is held at the end of the stroke.

i.  j. 

Retraction (j-n): The hand relaxes and returns to the rest position. The retraction phase ends at the moment the hand arrives in the lap where the hand will be resting.

j.  k.  l.  m. 

n. 

The overall coding procedures involved the three general steps listed below; however, the actual segmentation process usually involved repetitive back-and-forth observation and revision among the three following steps:

Step 1: Locate the target gestures

Step 2: Segment the stroke phase and then other phases in mute

Step 3: Check segmentation with speech

The first step of coding was to play the video with sound at both normal and slower speeds to roughly locate and annotate the gesture(s) that needed to be segmented, before the gesture phases could be precisely identified. Reducing the speed is necessary in terms of helping the researcher gain a clearer picture of the gestures. In addition, it is crucial to observe a few gestures preceding and following the target gestures co-occurring with the TFSs in order to see the movements clearly.

Based on the approximate location of the target gesture, the next step was to segment the stroke phase from the other gesture phases in mute without the transcript. Once the stroke had been segmented, the preceding phases (i.e. preparation, pre-stroke) and following phases (i.e. post-stroke and retraction phases) could be segmented. This step mainly requires painstaking observation of the changes of the handshape, arm, velocity and strength, etc. by manually moving the mouse back and forth in the annotation area in ELAN. The final step was listening to the speech and merging separate strokes which were supposed to be combined into one stroke (e.g. movements describing a single entity/action). However, no such revision has been made in the target gestures.

Large-scale research projects usually recruit more than one analyst to segment gestures, and check and double-check annotations, in order to ensure consistency and inter-rater reliability. However, the current research only has one coder and such a process is not possible.

3.4. Gesture coding schemes

The majority of previous gesture studies have defined the gesture phases in their frameworks; however, the actual coding processes in each phase involves numerous important decisions and few publications have offered detailed descriptions. Hence, the current section aims to provide specific illustrations of the different movement patterns and phenomena occurring in each gesture phase and the ways in which they were coded exactly. The following coding schemes were established and provided useful guidelines during the analytical process of the first study, i.e. coding the gestures co-occurring with the TFSs, and were further applied to the coding of the circular gestures in the second study (refer to Chapter 5, 5.2.1 for more details of coding circular gestures). Each following sub-section will be dedicated to illustrating the coding

principles in each gesture phase, from coding the preparation phase, to the stroke, and to the pre-stroke/post-stroke. It is hoped that this coding framework will not only clarify some issues that have been under-estimated before, but also help enhance the transparency and consistency of gesture analysis.

3.4.1. Coding preparation phase

The section will demonstrate the guidelines for demarcating the preparation phase, entailing fleeting preparation coded as no preparation, lengthy preparation, no preparation between successive strokes and preparation following post-stroke hold phases.

3.4.1.1. Fleeting preparation coded as no preparation

At times, the onset of a gesture can be so abrupt and rapid that the preparation phases can be too fleeting to observe. Therefore, no preparation phases were coded in those gestures. For instance, the pictures below, from a to h, represent a bi-directional symmetric stroke, where the supervisor moves the hand outwards to the right (a-e) from the rest position and moves back (e-h) to a post-stroke hold position close to the rest position. The movements are so rapid that the possible preparation phase might occur somewhere between picture a and b, the duration of which is only 0.002 second. Owing to the extremely fleeting nature of the movement, the images from the rest position to preparation are too close and blurry to observe. Although, theoretically speaking, a gesture starting from a rest position should have a preparation phase as the hand needs to prepare for the hand shape and location for the stroke, cases like this were exceptions and were coded as gestures with no preparation phases.

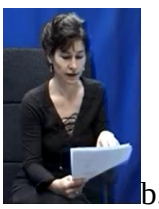
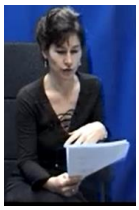
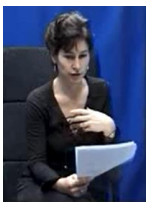
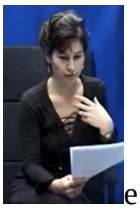
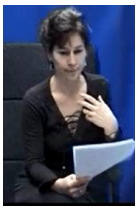
a.  b.  c.  d. 

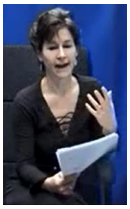
e.  f.  g.  h. 

3.4.1.2. Lengthy preparation

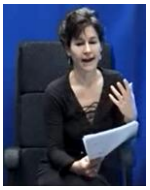
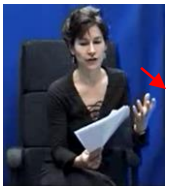
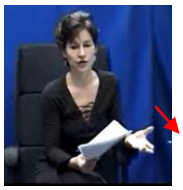
Sometimes the preparation phase could be somewhat gradual and lengthy before being followed by a more strenuous stroke phase. For instance, the following preparation from a to f occupies 2.3 seconds before a fast, downward stroke (f-i). The stroke lasts for only 0.3 of a second excluding the pre-stroke and post-stroke hold phases. The hand moves a long way until it arrives the starting point of the stroke with the required hand shape (an open hand with digits extended and separated), during which time the hand has gained more tension than at the beginning of the preparation.

Preparation:

a.  b.  c.  d.  e. 

f. 

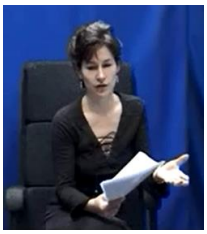
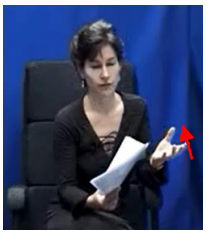
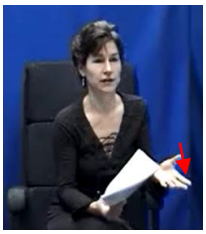
Stroke:

f.  g.  h.  i. 

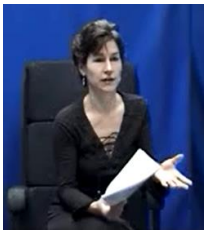
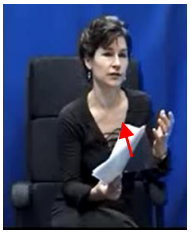
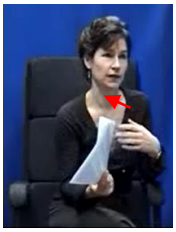
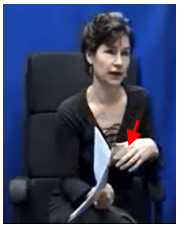
3.4.1.3. No preparation between successive strokes

Preparation phases are not mandatory for a gesture to exist, and they tend to be absent between successive strokes as exemplified below. Pictures from a to c represent an up-down stroke immediately followed by a different stroke from c to g.

Stroke 1: bidirectional movements (the downward part can be slightly faster probably due to gravity, but still coded as part of stroke).

a.  b.  c. 

Stroke 2: turning the palm of the left hand from facing upwards to downwards by rotating the wrist.

c.  d.  f.  g. 

3.4.1.4. Preparation following post-stroke hold

When movements are closely linked to each other, especially if the hand is constantly held between gesture boundaries, it can be challenging to identify

the starting point of the preparation phases as the changes of hand shape and position can be very subtle and gradual. On those occasions, instead of merely looking at the hands, the researcher observed the changes of the arm, the elbow and the wrist because movements may start from the arm and/or the wrist rather than directly from the hand. Below is an example of such an instance。

Post-stroke hold phase: the hand shape is mostly maintained with only slight changes.

a.  b. 

Preparation phase: the changes in hand shape are subtle; preparation starts when the arm moves towards the outside.

c.  d.  e.  f. 

Stroke: a rapid outward and downward movement performed.

g.  h.  i. 

As can be seen from the pictures above, due to the subtlety of the changes of the hand shape from the post-stroke hold phase (see a-b) to preparation (see c-f), the starting point of this preparation phase was coded as starting from the movement of the arm and elbow outwards. The motions of the arm and elbow

enable the wrist and the hand to rotate and move slightly inwards, which then allows the hand to perform a sudden and forceful outward and downward stroke.

A further important point concerns the boundary between retraction and preparation. Although the boundary was usually clear, when there was a difficulty, the transitional movement would be coded as a preparation phase rather than a retraction phase. This is also a method adopted by Kita's team (2007) as is more consistent with the forward-moving nature of the gesture rather than the backward-moving.

3.4.2. Coding the stroke phase

The stroke phases have a wide range of movement types and the following sub-sections will describe the framework used for coding them. They are the strokes in the Palm Up Open Hand gesture, in the bounded-space gesture, in the beat gesture, in bi-directional symmetric movements, in bi-directional asymmetric movements and in continuous movements.

3.4.2.1. Palm Up Open Hand Gesture

Palm Up Open Hand Gesture is a recurrent gesture that has been widely assessed; however, previous researchers seem to pay little attention to the specific description of how this gesture was coded (Calbris, 2011, p. 68; Kendon, 2004, p. 59; McNeill, 2005, p. 17). From their description, whenever the hand is open with the palm facing up, it can be termed as a PUOH. However, the coding is far from straightforward and there is insufficient description and agreement in the existing literature regarding the ways in which it should be coded and distinguished from gestures with similar features. For instance, the cyclic gesture (Ladewig, 2014a, 2014b) may also involve the

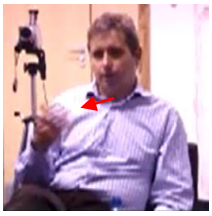
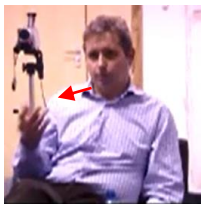
motion of rotating and opening the hand. Also, numerous gestures could end with an open hand facing upwards in their post-stroke hold phases. Therefore, the refinement of the coding and definition of the PUOH gesture is vital, and the current coding scheme intends to provide some answers to the aforementioned concerns.

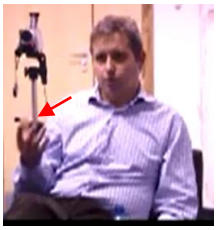
In the current study, for a particular gesture to be a Palm Up Open Hand gesture, it must be an independent gesture and the kinesic properties need to include: *1)* in the stroke phase, the palm of the hand being open and facing upwards in front of the gesturer with the fingers extended either partially or completely, either spread or flexed, and either touching or separated (however, the fingers cannot be folded, touching the palm and becoming a fist); *2)* the palm of the hand faces upwards with the back of the hand forming an approximately 0-45 degree angle to the horizon; *3)* the PUOH needs to have a separate stage where the hand rotates at least partially to open and/or extend the hand forwards; *4)* the coding of the stroke phase of the PUOH gesture largely depends on the velocity, which usually contains more strength as the hand slightly extends forwards and outwards. If the stroke depicts a full circle with the same speed, it would be coded as a circular gesture rather than a PUOH. The coding of the stroke starts from the moment when the PUOH is formed, at which point the hand may need to rotate and extend further to reach the final point. It is also worth mentioning that a post-stroke hold phase might occur where PUOH would freeze at the end of the stroke. An example is provided below:

Preparation: the hand, which is already facing upwards, moves away from the rest position and gains more tension, forming the Palm-Up-Open-Hand hand shape.

a.  b. 

Stroke: the hand moves upwards and then downwards, opening up the hand.

b.  c.  d.  e. 

f.  g. 

Retraction: the hand moves back to the rest position.

g.  h.  i. 

3.4.2.2. Bounded-space gesture

A further frequently mentioned gesture in the literature (McNeill, 2005) and in the current corpus is termed the bounded-space gesture in this thesis, based on the shape of the hands in the stroke. The hand shape of the stroke is illustrated in the picture below.

 The hand shape in the bounded-space gesture.

As can be seen from the picture, the hands are held in parallel, facing each other with the palms open, therefore, forming a bounded space between them. The fingers of the hands can either be extended or tilted, can be separated or touching each other. Most importantly, in the bounded-space gesture, whenever the hands move, they should move in synchrony and symmetrically, maintaining the bounded space between them.

When coding the bounded-space gesture, similar to the Palm Up Open Hand gesture, the stroke phase starts from the moment when the bounded space has been formed with the hands facing each other in parallel. This can take place before the hands reach the furthest point.

Below is a demonstration of how a bounded-space gesture was coded in the current study:

Preparation: the hands rise from the lap until they are open in parallel with the palms facing each other.

a.  b.  c.  d. 

Stroke: both hands continue to move to the centre and towards each other and then move away slightly.

d.  e.  f.  g. 

h.  i. 

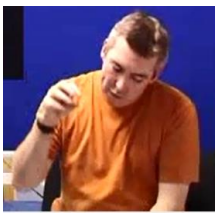
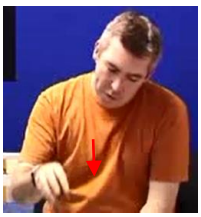
The following stroke: the hands change the trajectory of movements by moving the hands up and down in the opposite directions. Thus, the up-down movements were coded as a different stroke.

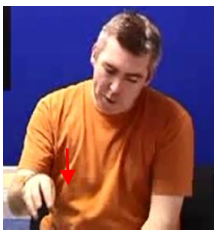
j.  k.  l.  m. 

The Palm Up Open Hand gesture and the bounded-space gesture are frequently occurring gestures, both of which might involve a long period of holding. According to the description of the previous researchers, it is not clear whether they coded the holding period as part of the stroke or as a post-stroke hold phase. For the current study, it was coded as a post-stroke hold phase.

3.4.2.3. Beat gesture

Beats are mostly formulated by up-down, in-out and left-right flicks of the hand or finger (refer to Chapter 2, 2.4.2. for further information) (McNeill, 2005). In the current study, continuous beats were coded as one single stroke, whereas discontinuous beats with pauses between were coded as separate strokes. Also, if the up-down motions cover a certain space such as , depicting a curved route, they were not coded as beats. Following is an example of three continuous up-down beats (a-g), which were coded as one stroke:

a.  b.  c.  d. 

e.  f.  g. 

Although there are some variations among the size of the beats, they are not only continuous, but also have a high degree of similarity in tension and tempo. Therefore, continuous beat movements were coded as one single stroke. Below is a further example of two continuous beats coded as one stroke (a-e), which move in right-left trajectories.

a.  b.  c.  d. 

e. 

3.4.2.4. Bi-directional symmetric movements

Bi-directional symmetric and continuous bodily motions with a similar speed, tension and movements were coded as one stroke. However, because of gravity, the downward or outward movements may appear faster than their

upward or inward counterparts. Below is a case of outward and inward symmetric bi-directional gesture with the hand moving out in an arc trajectory from a to e and then goes back to where it departs from e to h.

a.  b.  c.  d. 

e.  f.  g.  h. 

The above example is only one form of the bi-directional symmetric stroke involving a simple in-out motion. Sometimes the movement trajectories could be much more complex. Nevertheless, as long as the movement in each direction maintains symmetry, bi-directional movements were coded as one stroke. Moreover, the above instance of the bi-directional gesture is performed by one hand, but it could also comprise two hands as exemplified below. The two hands perform an in-out movement in turn symmetrically and continuously four times, and the entire movements were coded as one stroke.

Preparation (a-b): the right hand moves slightly outwards and the left hand inwards.

a.  b. 

Stroke (c-j): two hands perform rapid, large-size, outward-inward symmetric movements in the opposite direction twice.

c.  d. 

c-d: left hand moving outwards; right hand inwards.

e.  f. 

e-f: right hand moving outwards; left hand inwards.

g.  h. 

g-h: right hand moving inwards; left hand outwards

i.  j. 

i-f: left hand moving outwards, right hand inwards

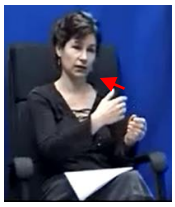
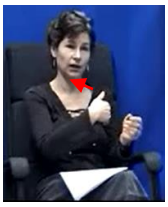
Retraction (k-m): the hands relax and move to the lap.

k.  l.  m. 

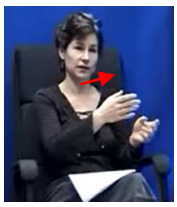
3.4.2.5. Bi-directional asymmetric movements

As illustrated in the previous section, bi-directional symmetric movements were coded as one single stroke; however, this did not apply to asymmetric bi-directional movements. In those, more often than not, the former half (see a-c in the instance between) is usually performed in a weaker form with less tension and/or at a lower speed than the latter half (c-f). Hence, the former half was coded as the preparation phase and the latter as the stroke. Below provides such an example.

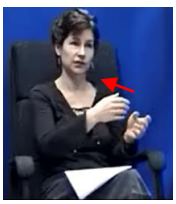
Preparation: Right hand moves slightly and gently inwards while bending the palm towards the speaker.

a.  b.  c. 

Stroke: The right hand rapidly moves outwards in the same trajectory as in the preparation phase above.

c.  d.  e.  f. 

Retraction: Both hands relax and move to touch each other.

f.  g.  h.  i. 

3.4.2.6. Continuous movements

Continuous movements with similar actions (straight, curved, rotating/circular,

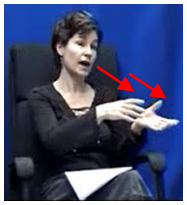
etc.), tension and velocity were coded as one stroke; however, continuous motions with **different** movements were **not** coded as the same stroke. The example below illustrates *two strokes* with a downward movement followed by two curvy continuous actions, which can be illustrated by this simplified conceptual routine: .

The first stroke in the following example is a downward movement (see c-d), which is followed by a continuous curved movement (d-m). These were coded as two separate strokes. The first part of the second stroke (d-h) has a similar trajectory and velocity as the second part (h-m) without any pause. Admittedly, there are some subtle variations between the handshapes and the routines between the two parts of the strokes; however, in view of the overall similarity and continuity of the two parts, they were coded as one stroke.

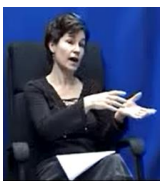
Preparation:

a.  b.  c. 

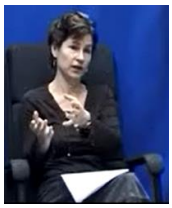
Stroke 1:

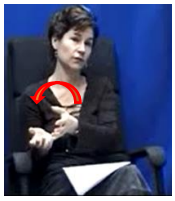
c.  d. 

Stroke 2 (Part 1):

d.  e.  f.  g.  h. 

Stroke 2 (Part 2):

h.  i.  j.  k.  l. 

m. 

Below is a further example of a downward movement (see c-f) followed by one curved movement (see f-j) in order to demonstrate how changes of movements lead to different strokes:

Preparation: raising the hand from the rest position and forming an open hand with the palm facing up and the digits flexed.

a.  b.  c. 

Stroke: downward movement performed by rotating the wrist, palm and fingers downwards.

c.  d.  e.  f. 

Stroke: curved movement performed by moving the hand to the right.

f.  g.  h.  i. 

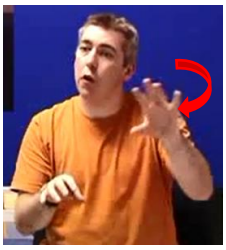
j. 

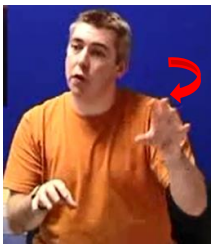
The principle of similarity in movement, tension and speed in continuous motions also applies to other motion types. For example, in continuous multiple circular movements, minimal differences in the shapes of the circles such as  depicted by the naturally moving hand would not affect the coding; however, changes of trajectory (from horizontal to vertical, etc.) and/or tension and/or velocity would lead to different strokes.

3.4.3. Coding the pre-stroke and post-stroke hold phases

The section will describe the principles of coding the pre-stroke and post-stroke hold phases. Strictly speaking, pre-stroke and post-stroke hold phases should still be conceived as part of stroke as they are the holding period after the hand reaches the start and end points of the strokes. However, there seems to be a lack of attention to the length of holding. One observation made in the current thesis is that the different lengths of holding can lead to different discourse functions (refer to Chapter 4, 4.5.2 for details). In the current study, only those holds that exceed 0.1 second were deemed to qualify as individual pre-stroke or post-stroke hold phases. This decision was made based on the observation that most intervals between movements continue for about 0.06-0.08 seconds, which were usually not visually salient. Only when the holding time reached 0.1 second and beyond would the hold become noticeable. Sometimes the hand was not completely frozen in the hold phases, i.e. it might drift slightly, which did not affect the coding. Below is an example illustrating this type of coding:

Stroke: the left hand outlines two continuous downward curved movements.

a.  b.  c.  d. 

e.  f.  g. 

Post-stroke hold: the left hand remains steady at the end of the stroke.

a.  b.  c.  d. 

Retraction: the left hand relaxes and retracts.



3.5. Transcribing and representing gestures

This section will illustrate the ways in which the gestures were transcribed in the current thesis, together with speech, which serves the observation and analysis of data. Similarly, as the orthographic transcription of speech compromises the authenticity and integrity of the data, any transcription of gestures will have the same effect. Similar to speech transcription, there are no agreed conventions for gesture transcription, but some essential elements are commonly seen in the published works. The figures below represent some of the most popular approaches of transcribing and representing speech and

gesture alignment. Kendon's (2004) conventions as demonstrated in Figure 3.4 and Figure 3.5 below are possibly one of the earliest proposed multimodal transcription guidelines for gesture researchers.

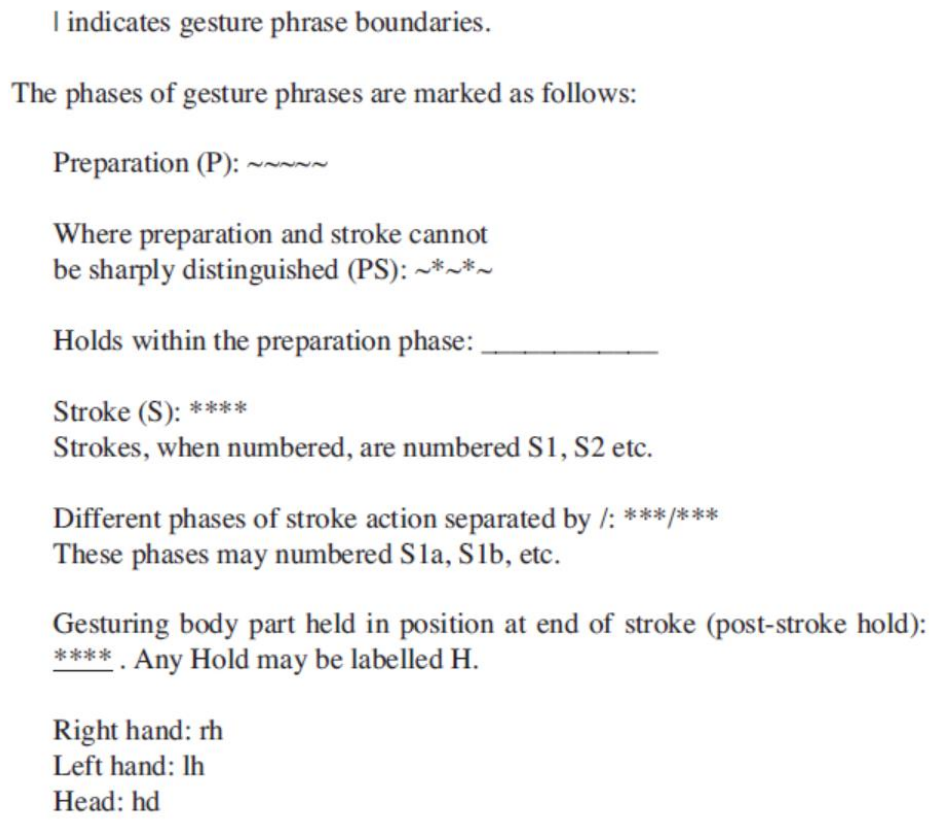


Figure 3.4: Kendon's (2004) gesture transcription convention.

Different symbols were used to represent gesture phases with their start and end points aligned with the speech co-occurring with them. Pictures were drawn to demonstrate the gestures with arrows indicating the movement trajectory. Immediately below the transcription is usually a more detailed description of the gestures, providing any relevant information that cannot be shown by the pictures and symbols.



Fig. 12.2 Example 67. S closes his open hand (dotted line) to *grappolo* rotating it in towards himself as he says "*Tenë n'atu nummërë*" ('he had another number'). This action is repeated twice more in S's utterance, as shown in the transcription below.

Example 67 Bocce I 7.42.37.
~~~ preparation; *** hand as *grappolo*; <<< hand opens.

he had another number, that was twenty-three, I don't know who he was telephoning
tenë n'atu nummërë, chillë erë vintitrè nun saccië a chi telefonavë
|~~~**********|<<<<<<*********|<<<<<<<<<<<*********|
[1a] [1b] [2a] [2b] [3a] [3b]

Figure 3.5: An example of multimodal transcription (Kendon, 2004).

Kendon's method has been adopted, with slight alterations, by numerous researchers working with gestures. One problem of using this method is the issue of becoming familiar with all the symbols during transcription and reading. Typing, aligning and formatting those symbols can be inefficient and tedious, especially when the sentence(s) run(s) across several lines. Also, this representation is suitable for representing single instances, but might not be ideal for comparing and contrasting several. As adopting a corpus-based approach means observing many instances and identifying emerging patterns simultaneously, this mode of representation might not be transferable to the current thesis.

For the second study, automatic software is available to extract concordances for all the speech that co-occurs with the target circular gestures (refer to Chapter 5, 5.2.4. for details); however, for the first study, no such technique is currently available for concordancing all the gestures co-occurring with the TFSs. Of course, the ELAN software can be extremely useful for observing the patterns in all the instances after all the gestures have been annotated, but written transcription is still necessary for a clearer view of
~~~

the movement patterns and speech-gesture alignment. Therefore, the current thesis designed an approach that can manually align all the target TFSs and their co-occurring gestures in the following format. In this way, a speech-gesture concordance is then formatted with the targeted speech and gesture aligned in the middle, which has been proven necessary and valuable for pattern analysis.

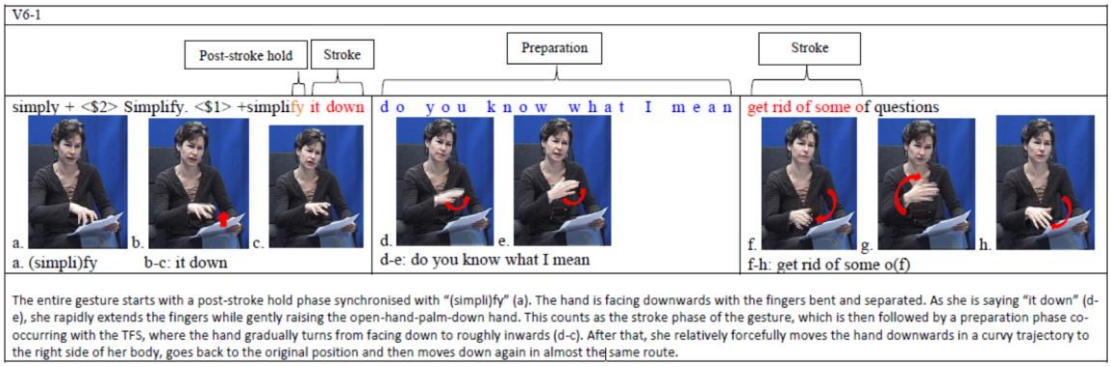


Figure 3.6: An example of gesture transcription co-occurring with a TFS.

As can be observed from Figure 3.6 above, the first row contains the number of instance of the TFS: V6 refers to Video 6 and 1 refers to the first instance. The second row contains all the gesture phases with the brackets marking the beginning and end of each phase. Divided by three columns, the TFS and the co-occurring gestures are positioned in the middle of the third row, with the speech/gesture preceding and following at the left and right side. The gesture phases were all coloured, which turned out to be rather helpful for analysing speech-gesture alignment: one can easily spot where the gesture phases start and end by simply referring to the colour coding. However, even when colour-printing is not an option for some publications or readers, the brackets and the names of the gesture phases in the second row can still assist the readers in distinguishing and observing the gesture phases without undergoing the difficulties of mapping out the different symbols and their corresponding gesture phases. Finally, a detailed explanation of the movements co-occurring

with the TFSs is given, providing information that is missing from the images. Such an explanation is vital for both the researcher and the reader in order to have a clear idea of what the movements look like.

A similar approach has been taken for all the instances of circular gestures in the second study (refer to CD-ROM data disk). As can be seen from the instance in Figure 3.7, the first column contains the speech at the left side of the speech co-occurring with the circular gesture. The middle column presents the speech co-occurring with the circular gesture (marked by CG# #), followed by any speech after it. The pictures are the stroke phase of the gesture (the beginning to the end), with the number of circles (2 circles in this instance) and the direction of rotation (clockwise or anti-clockwise) described below. The arrows in the pictures show the approximate plane (e.g. horizontal, vertical).

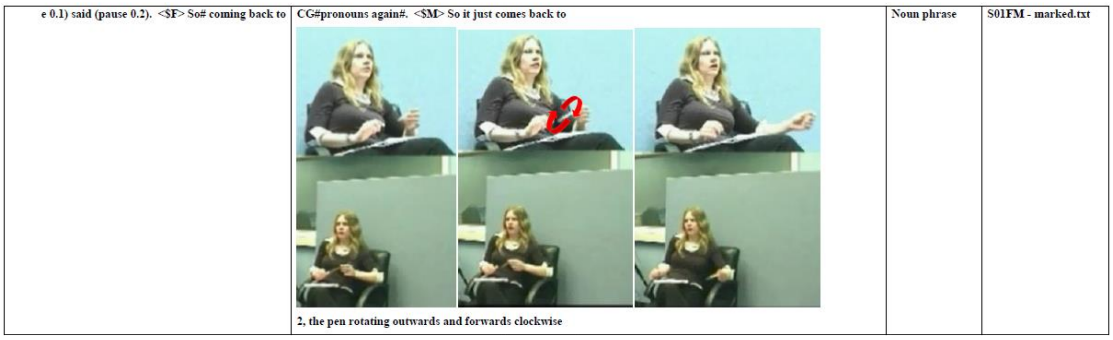


Figure 3.7: An example of gesture transcription for circular gestures.

3.8. Summary

To conclude, the current chapter described the data and the construction of gesture frameworks for segmenting gesture phases. It first provided relevant details regarding the NMMC, supporting the decision for using this as an appropriate, high-quality corpus for pursuing methods to examine the multimodal units of meaning in gesture and speech. As compiling a multimodal corpus normally faces numerous challenges in various respects

such as designing, recruiting, transcribing and annotating, using the NMMC allowed the researcher to devote more time and energy into the gesture part of analysis, which is extremely painstaking and time-consuming.

The latter part of the chapter defined, justified and exemplified the gesture coding schemes developed by drawing on the current dataset, which comprise preparation, pre-stroke, stroke, post-stroke and retraction phases. In each gesture phase, more fine-grained guidelines were set up in accordance with the different movement types emerging in the data. As previous gesture segmentation schemes scarcely clarified their coding in such detail, these guidelines are vital for enhancing the consistency and transparency of gesture analysis.

The final section briefly reviewed and evaluated some of the most prevalent modes of gesture transcription in the literature, and then illustrated the unique format of representation designed particularly for this thesis, which involved manual construction of speech-gesture concordances. This novel format of speech-gesture transcription is well-tailored to the analytical needs (i.e. identifying emerging speech/gesture patterns) and the multimodal corpus-based nature of the thesis.

Chapter 4. Towards a Speech-Gesture Profile of Formulaic Sequences: A Corpus-based Approach

4.1. Introduction

This chapter reports on the results of the first empirical study, which largely addresses the issue concerning which gesture patterns co-occur with meaning variations of a group of target formulaic sequences (TFSs), "(do) you know/see what I mean". Drawing on one particular meeting from the Nottingham Multimodal Corpus (NMMC), the research explores a multimodal corpus-based approach with the ultimate outcome of establishing a speech-gesture profile of the TFSs. The results of the case study are then compared with a small group of instances from other speakers within the NMMC, which generates new findings and implications for further research.

The corpus approach has demonstrated the differentiations in the frequencies of different meaning variations of multiword expressions in different genres (Sinclair, 1991). Despite a more robust description of the formulaic nature of language afforded by corpus linguistics, there has been a distinct lack of research on the role of gesture in formulaic sequences. In view of this, this study proposes the notion of the 'speech-gesture profile' of formulaic sequences as an analytical construct that underpins the description of the multimodal units of meaning. Using the TFSs "(do) you know/see what I mean", this chapter demonstrates the development of a speech-gesture profile and shows tendencies of gesture uses that co-occur with formulaic sequences in speech.

The nuances of meaning that are a direct result of gesture co-occurring with speech are readily observable in everyday interaction. For example, in the NMMC, a multiword expression such as "(do) you know/see what I mean"

often functions as a discourse marker, i.e. a functional expression that is primarily used to connect utterances, usually at an initial position, and contribute to the coherence of speech, but does not contribute to the syntax or actual meaning, or contents, of speech (Schourup, 1999). However, when accompanied by particular types of hand gestures, such as beat gestures, the meaning of checking understanding is foregrounded. Taking gestures into account in the description of units of meaning can thus shed new light on the form and function of multiword expressions while at the same time raising key questions concerning the ways in which gestures are organized in relation to multiword expressions. The questions addressed here include:

1) What is the temporal relationship between the TFSs and the co-occurring gestures?
2) What is the unit of meaning of the TFSs in accordance with their co-occurring gestures?
3) What are the discourse functions served by the gestures co-occurring with the TFSs?

Drawing on the NMMC, the approach developed in this chapter supports a new understanding of the relationship between language and gesture in multiword expressions. It also offers a multimodal framework for analysing meaning in discourse, as well as new insights into the nature of multiword expressions. The chapter will begin with a description of the approach used in the first study in 4.2., including a detailed description of the data, the frameworks for coding speech functions and the segmentation of gestures in the ELAN software. This will be followed by the presentation of results of the case study (4.3.) and the extended analysis (4.4.), focusing on the temporal relationship between speech and gesture, and the gesture patterns co-occurring with the TFSs. The discussion (4.5) will focus on presenting and illustrating the speech-gesture profile of the TFSs and the discourse functions of the

gestures. The final section will provide a summary of the approach, the findings and limitations of the whole study.

4.2. Approach

The approach adopted in the first study involves, firstly, identifying the various contexts in which different multiword expressions occur from a discourse analytical perspective (Adolphs, 2008), and then describing the gestures that co-occur with them from a semiotic perspective, i.e. their form and different semantic derivations (Calbris, 2011). The annotation was carried out with the use of an established software for multimodal discourse analysis (i.e. ELAN). This section will first concentrate on describing and justifying the use of "(do) you know/see what I mean" from Video 6 as the basis for the case study, along with a brief overview of the existing research on similar word strings (4.2.1). Then, the framework for coding speech functions will be illustrated in detail in 4.2.2, followed by a demonstration of the use of ELAN software in speech and gesture segmentation in 4.2.3.

4.2.1. The video for the case study and the TFSs

The first study started with a search for a relatively long formulaic sequence in all the supervision meetings in the NMMC for the reason that the longer the word strings are, the more likely they will be accompanied by individual gestures. During the search of 3/4/5-word clusters in AntConc[4], a group of expressions, "(do) you know/see what I mean", were foregrounded, and more than 80% of them (64 out of 76) come from one speaker (Supervisor 6 in Video 6). Video 6 is a supervision meeting between one native

[4] http://www.laurenceanthony.net/software/antconc/

English-speaking, female supervisor, and one native English-speaking, female PhD student. During the 47:52 minutes of conversation (11,705 running words), the supervisor used "(do) you know/see what I mean" 64 times. The rest of the TFSs are contributed by several other speakers, each providing only a few instances (details will be presented in 4.4). The extremely uneven distribution of the instances led to the decision of using Video 6 as the base for exploring a multimodal corpus-based approach to the units of meaning. As the goal of the study is to establish a novel approach rather than generating results that are applicable to a large population, the relatively small numbers in the case study is not problematic. The remainder of the instances can then be used as the dataset for an extended analysis, a test bed for the validity of the findings from the case study.

A further reason for choosing this sequence of formulaic language, "(do) you know/see what I mean", is due to its high frequency in real-life communication and its significant role in interpersonal relationship. Carter and McCarthy (2006) explicitly state that frequently occurred multiword expressions such as *you know, I mean, you know what I mean* and *do you know what I mean* are all adopted to reassure the assumed common ground between participants in communication. As "(do) you know/see what I mean" often occurs as an extended version of *you know* and *I mean* (Adolphs & Carter, 2013), previous research has explored the various meanings of *you know* and *I mean* (Stubbe & Holmes, 1995; Tree & Schrock, 2002), the full TFS described here remains underexplored.

In Stubbe and Holmes' (1995) description of New Zealand spoken language, they specifically point out that, although pragmatic markers such as *you know* and *I mean* are so prevalent in everyday speech that they seem redundant and informal, they perform a vital role of 'oil(ing) the wheels of interaction' (63) and monitoring the common ground between the interlocutors. Similar comments have been made by Tree and Schrock

(2002), claiming speakers' use of *you know* and *I mean* were usually tied to speaker's clarification and modification of speech, inviting the addressees' involvement and confirmation of understanding. From a slightly different perspective, Woods' (1991) study of Canadian English underlines the pragmatic function of *you know* as an appeal for sympathy, indicating an important interpersonal role.

In view of the aforementioned findings, the current TFSs presumably are multi-functional, playing roles in terms of discourse cohesion, negotiation of mutual understanding and interpersonal affiliation. This assumption is confirmed by the discourse analysis of each instance in the current study, based on which a fine-grained descriptive framework of the speech functions of the TFSs is built up. This framework will be illustrated in the subsequent section (4.2.2.).

Potentially, there could be four different variations in "(do) you know/see what I mean", including "you see what I mean", "you know what I mean", "do you know what I mean" and "do you see what I mean". Hence, all the four variations have been searched in the speech transcription in Microsoft Word. However, only two forms have been found in Video 6, which are "do you know what I mean" with 29 instances and "do you see what I mean" with 35 (see Table 4.1 below).

Table 4.1: Frequency of TFS variations.

Variation of TFS	Amount
you see what I mean	0
you know what I mean	0
do you know what I mean	29
do you see what I mean	35
Total	64

It is important to clarify again that all those instances were found in the supervisor's speech (Supervisor 6 in Video 6) as the student did not use any of these variations. The figure suggests an approximately equal preference for both sequences by the supervisor.

4.2.2. Coding speech functions

After pinpointing all the instances of TFSs in the speech transcription, the next step is to identify the functional variations in them. Whereas the literal meaning of this multiword expression indicates the function of checking understanding on the part of the speaker, discourse analysis points to a more nuanced framework of description. It seems that the functions of the TFSs are on a scale from being discourse markers at one end to a means of checking understanding at the other. In other words, there are some TFSs serving the function of a discourse marker, some used to check mutual understanding and some performing both functions. To distinguish the three functional variations in the speech TFSs, it is necessary to pay attention to the discourse structure, i.e. the position of the TFSs in the utterance, and the wider discourse context (refer to File 4.1 in the CD-ROM data disk for the speech transcripts of all the instances in NMMC).

A relatively clear-cut example of checking understanding is provided in the following extract:

> *Supervisor: I think I think that might give you the focus better and then this chapter will become more a chapter that's setting up the theorisation of space do you see what I mean+*
> *Student: Yeah.*

The key criteria of coding the function of checking understanding is to identify a speaker-eliciting-and-listener-confirming sequence. The confirmation from the student can be verbal, but also nonverbal, usually head nods. A multimodal discourse analysis of the functions of the TFSs reveals that, without having access to the nonverbal elements in interaction, it is unlikely to achieve a comprehensive understanding of the dynamic functional variations in them.

For those instances of TFSs that are discourse markers, they usually occur at the beginning of an utterance or in dysfluent speech, marked by false starts, repetition, incomplete utterance, etc. Moreover, the TFSs as discourse markers do not contribute to the referential meaning of utterances in the sense that they do not trigger any verbal or nonverbal response from the other party, but signal the forthcoming of more utterances. The following example is an illustration of such a circumstance, in which the TFS is not followed by any response from the student, but more utterances from the supervisor. Besides, in view of the fact that the student interrupts (marked by '+') the supervisor by saying "*Right"* before the TFS, it seems unlikely that the supervisor intends to check the understanding of the student.

> *Supervisor: +and utility of space theory for a literary critic as opposed to the chapter being about hotels +*
> *Student: Right.*
> *Supervisor: + <u>do you see what I mean</u> and why are hotels an important issue cos I think you've you've got to make a decision are you using space theory as a mechanism to help you understand what's going in the hotel+*
> *Student: Mhm.*

Another functional type of TFSs seems to combine both functions of checking understanding and as a discourse marker. They are usually inserted

in the middle of a sentence where the information is only partially delivered, but followed by instant or delayed response. The position of the TFS indicates the function of a discourse marker, whereas the wider discourse structure and contents point to the function of checking understanding.

Supervisor: it's a it's a transition which is taking place do you know what I mean (Student: Head nods) over a long longish period of time just as hotel buildings are taking place over a longish period of time (Student: Head nods) so I think if you …(turn continued)

Results (see Figure 4.1 below) show that more than half of the TFSs (34 out of 64) are adopted by the supervisor as a way of ensuring the understanding of the student on a particular issue, whereas only 11 are discourse markers. The remaining 19 instances appear to have both functions.

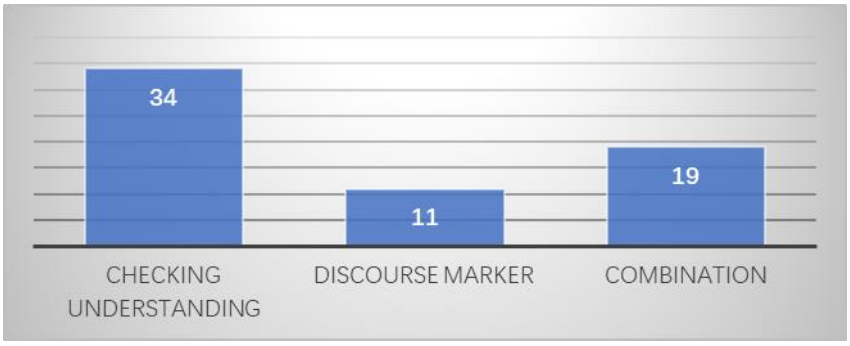


Figure 4.1: Distribution of speech functions in TFS.

4.2.3. Segmenting gestures in ELAN

The investigation into the multimodal units of meaning naturally requires a fine-grained description of the temporal alignment between gesture and speech. Therefore, it is crucial to segment gestures into analytical units so that it is possible to observe the gesture patterns co-occurring with the TFSs. This

is when the five-phase gesture analytical framework comes into play, with gestures being demarcated into preparation, pre-stroke, stroke, post-stroke and retraction (for a detailed description of gesture phases refer back to Chapter 3, 3.4.) (Kendon, 2004; Kita et al., 1998; McNeill, 2005).

The speech part of the corpus has already been transcribed, with exact time stamps, by previous researchers in Transana (Adolphs, Knight, & Carter, 2011; Knight, 2011; Knight et al., 2009a), however, Transana is no longer freely available, so it was decided to use the CLAN software (http://dali.talkbank.org/clan/) in the current thesis in order to pinpoint the locations of the TFSs in the video files using the time stamps. The TFSs were then all annotated in CLAN and the annotated files exported to the ELAN software with all the instances already marked-up in the time line. ELAN's multimodal annotation function is vital and efficient for analysing gesture and speech in line with the purposes of the current study.

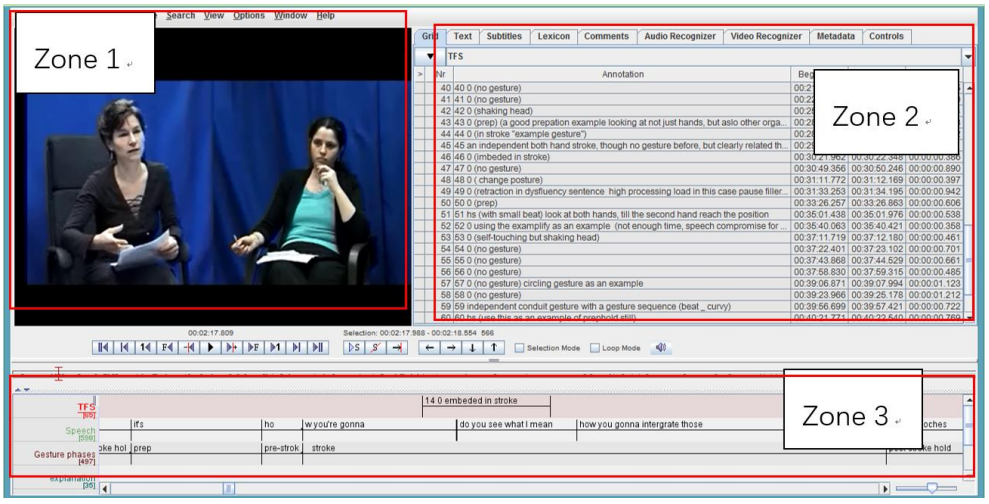


Figure 4.2: A screen shot of ELAN

Figure 4.2 is a screen shot showing an example of both gesture and speech when they have been annotated in ELAN. As is shown in the picture, Zone 1 is the video area where gestures can be observed, and Zone 2 is the area where all the annotations are listed and numbered. The researcher can select any

instance for analysis with a click of the mouse. Zone 3 is the annotation area, where the TFSs, speech and gestures were annotated in separate tiers. The exact speech and each gesture phase were aligned, transcribed and annotated in the second and third tiers. The tiers for speech and gesture transcription can be switched easily by dragging the tiers. It is necessary to point out that, as the speech sound is always linked together, it is not uncommon that one syllable can run across the boundaries of gesture phases. Therefore, at times it can be difficult for the researcher to decide which gesture phase such a syllable belongs to; however, one syllable of error would barely affect the results of analysis.

4.3. Results of the case study

This section mainly reports the analytical results of the case study drawing on Video 6. The sub-section 4.3.1. will illustrate six types of gesture phases co-occurring with the TFSs, based on which, the gesture patterns co-occurring with each functional variation of the TFSs can then be analysed (4.3.2).

4.3.1. Temporal relationship between TFS and gesture

After segmenting all the movements into gesture phases, the temporal relationship between gesture and speech can be assessed. The analysis generates six types of relationship: TFSs co-occur with (1) independent strokes/gestures, (2) preparation, (3) motions embedded in extensive strokes, (4) post-stroke hold, (5) post-stroke hold + retraction and (6) post-stroke hold + preparation. It is necessary to clarify that one TFS may co-occur with different gesture phases and the classification here is based on the predominant gesture phase that co-occurs with each TFS. Whereas the majority of the TFSs only coincide with one type of gesture phase, there are

two instances, in which the first halves of both TFSs co-occur with a post-stroke hold phase, and the second halves with a retraction phase and a preparation phase, respectively. The two instances will be illustrated at the end of this sub-section.

It also needs to be explained that, when a TFS co-occurs with a stroke phase and any other phase(s), such as preparation, retraction, etc., it was coded as an independent gesture. This is because the stroke phase is the most significant and only obligatory gesture phase where meaning is generated, i.e. the existence of one stroke equals to one gesture. For instance, half the TFS in Figure 4.3 below co-occur with a stroke phase, and the other half with a retraction phase.

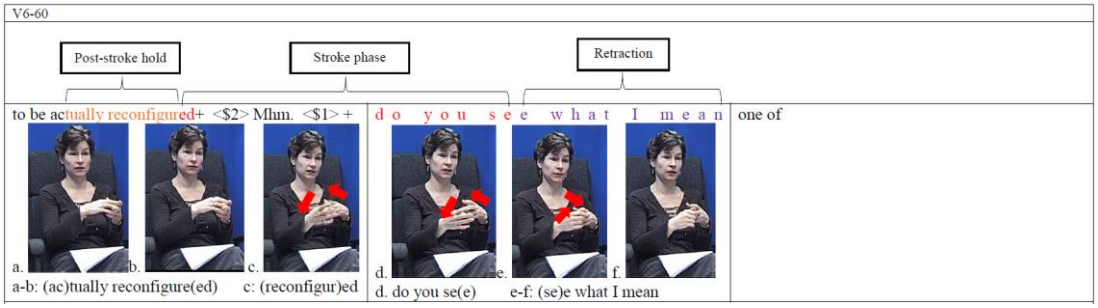


Figure 4.3: A TFS co-occurring with a stroke phase with a long retraction phase.

The hand remains in a post-stroke hold phase when she is saying "(ac)tually reconfigur(ed)[5]" (a-b), at the end of which she abruptly starts to move the right hand outwards and the left hand inwards (c). This stroke mainly co-occurs with the first half of the TFS "do you se(e)" (d). In line with the second half of the TFS "what I mean" is a retraction phase of the stroke where the hands start to relax with the fingers gradually bending (e-f) until the right hand holds the left one. Although the TFS is aligned almost evenly

[5] The letters in the parenthesis means they do not co-occur with the gesture phase.

with the stroke phase and the following retraction, this instance was coded as a TFS co-occurring with a stroke phase, or an independent gesture.

However, except for the five instances co-occurring with a relatively long post-stroke hold phase (V6-11, V6-19, V6-21, V6-22, V6-29), most of the strokes accompanying the TFSs start and complete in close proximity with the beginning and the end of the TFSs. Figure 4.4 below is a demonstration of such a phenomenon.

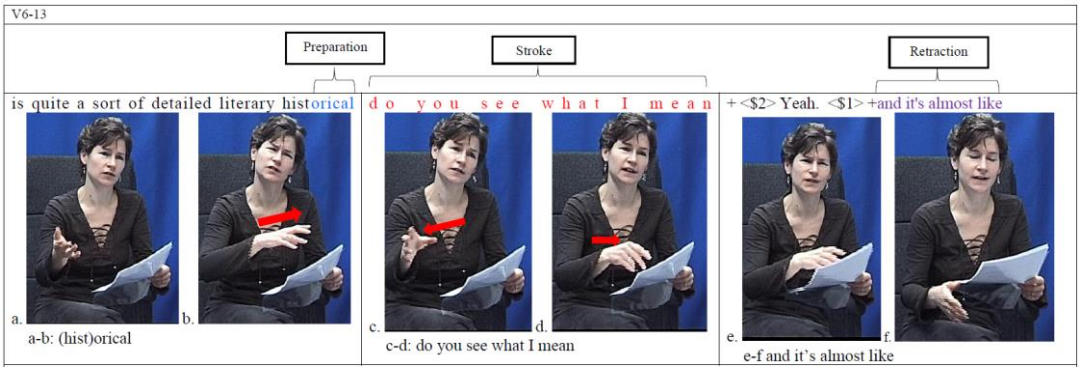


Figure 4.4: A TFS co-occurring with an independent stroke/gesture.

As the left hand is holding the papers, the speaker turns the right hand from facing up to downwards as gently moving the hand from a peripheral position (a) to a more central position closer to her body (b). After this preparation phase co-occurring with “(hist)orical” (a-b), during the stroke synchronised with the TFS, she rapidly moves the hand outwards and rightwards (c). The movement is so sudden and strong that the hand bounces back slightly (d). Immediately after the TFS, the hand starts to prepare for the next stroke as she says “and almost like”.

A TFS may also be accompanied by movements that constitute a preparation phase as is shown in Figure 4.5. below. The hand remains still (a-b) when the supervisor is saying “plac(e)”, which is immediately followed by a preparation phase that is almost in precise temporal synchrony with the entire TFS (c-e). During this preparation phase, her hands, wrists and elbows

rise gently and gain more tension during the process, in attempt to prepare for the subsequent stroke. The stroke in this gesture is an outward-inward movement performed by both hands, depicting large curved routes simultaneously and symmetrically (f-h). The stretched distance between the two hands (g) during the apex of the stroke seems to be analogic to the idea of 'long' communicated in the coordinated speech component, i.e. "over a long longish period".

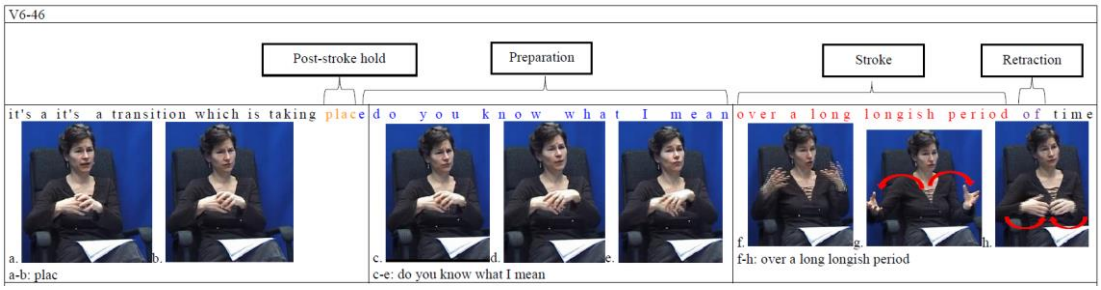


Figure 4.5: A TFS co-occurring with a preparation phase.

The following example (Figure 4.6) demonstrates a TFS-accompanying gesture that is embedded in a long stroke phase, starting from the utterance preceding the TFS and continuing after it. In this case, the stroke is temporally and semantically synchronised with the meaning delivered in "are gonna integrate those two approaches", with the TFS "do you see what I mean" inserted in the middle. Specifically, this stroke is a series of two-handed movements consisting of outward-inward motions in turn by rotating the wrists (c-h): the right hand moves outwards, the left inwards, and vice versa. The intertwined fingers are iconic of the meaning of "integrate the two approaches", which means unifying separate things.

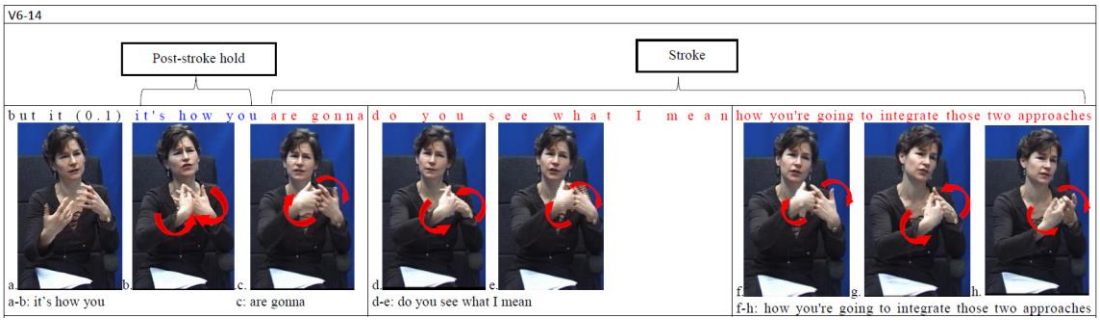


Figure 4.6: A TFS embedded in an extensive stroke phase.

The following example (Figure 4.7) shows a TFS-accompanying gesture that co-occurs with a post-stroke hold phase, where the hand remains steady with the palm facing upwards (d-e) after some up-down movements in the stroke (c), which is temporally in line with 'interpretive'. The post-stroke hold phase begins immediately before the onset of the TFS (see the hand form in pic. c), and is maintained over nearly the entire TFS (d-e), finishing during its final word, "mean" (f), where she instantly executes the subsequent stroke.

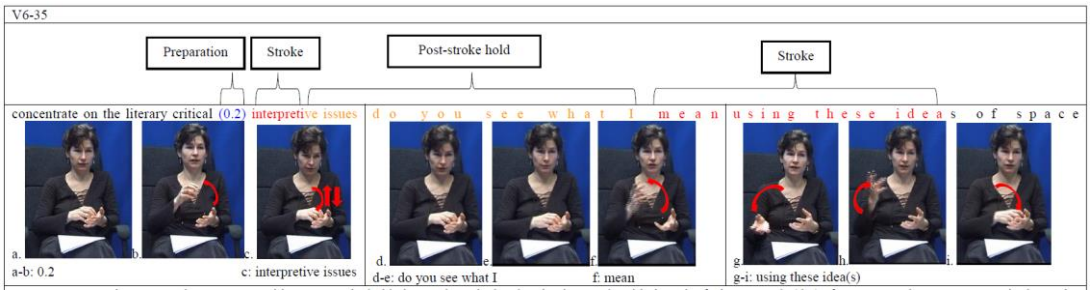


Figure 4.7: A TFS co-occurring with a post-stroke hold phase.

The instance in the next figure (Figure 4.8) is a TFS synchronised with a retraction phase. After an up-down movement stroke that is temporally coordinated with 'too much' and a 0.3-second pause (see movement sequence in pic. c: down-up-down-up-down), the hands instantly lose their tension and withdraw to a rest position right at the beginning of the TFS. In this rest position, the elbows are placed on the arms of the chair, with the left

hand clasping the right (e-f). This retraction phase is completed during most of the TFS, “(d)o you see what I”.

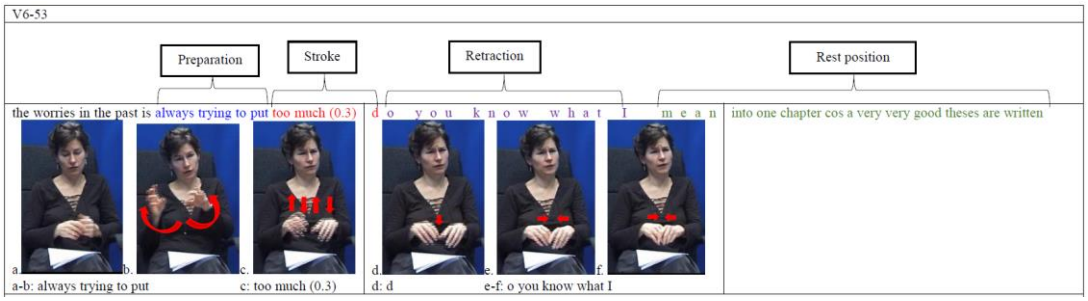


Figure 4.8: A TFS co-occurring with a retraction phase.

Whereas most TFSs of the 64 instances identified for this particular supervisor dominantly co-occur with one gesture phase, the instance below (Figure 4.9) is one of two instances that almost evenly co-occur with two gesture phases, these being a post-stroke hold phase and a retraction. As can be observed in the illustrations, both hands raise up during “so you could” (a-b), followed by a two-handed up-down beat stroke co-occurring with “do” (c) as a repetition directly before the fluent TFS. The former half of the TFS, “do you see”, is then synchronised with the post-stroke phase of the beat gesture (d), and the latter part, “what I mean”, with a retraction (e-f).

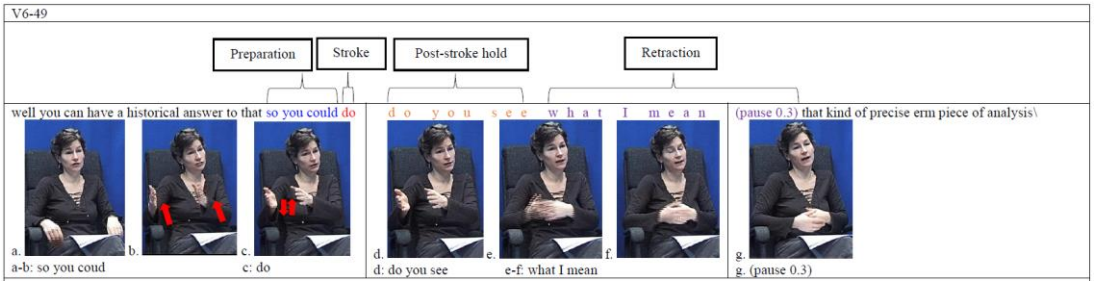


Figure 4.9: A TFS co-occurring with a post-stroke hold phase and retraction.

After presenting an instance co-occurring with a post-stroke hold phase and a retraction phase, Figure 4.10 is the other TFS composed of two gesture phases:

a post-stroke hold phase and a preparation phase. Both hands are held before the TFS when she is saying “(a)cturally a chapter on class” (a-b) with the left hand facing inwards and the right hand facing upwards. This position continues until the former half of the TFS “do you see” (c), which is then followed by a preparation phase in the latter half of the TFS “what I mean” (d-e). As the preparation phase starts, she uses the right hand to hold the small finger of the left hand while raising the hand to a higher position. This prepraration lasts a little after the TFS during “and th=” (f).

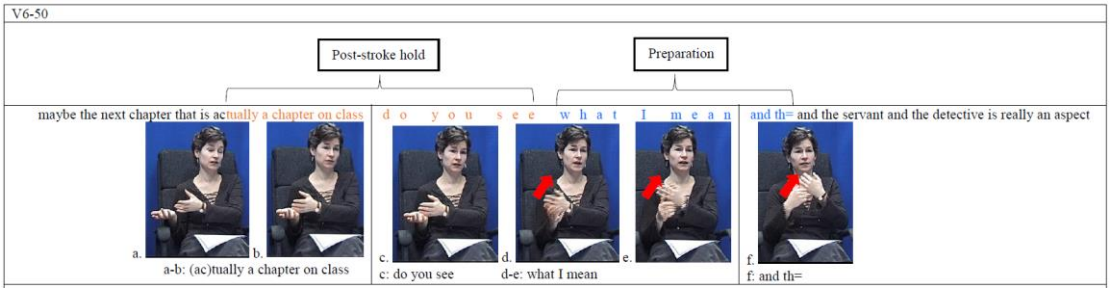


Figure 4.10: A TFS co-occurring with a post-stroke hold phase and preparation.

After demonstrating all the major categories of gesture phases coordinated with the TFSs, an overall view of their occurrences can be observed in the figure below:

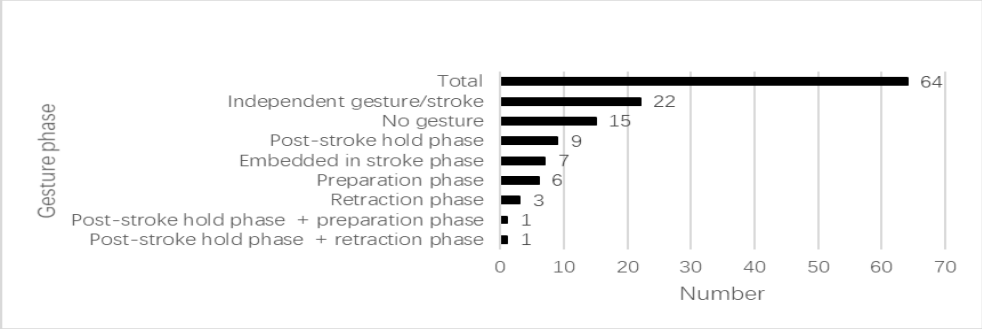


Figure 4.11: Distribution of the temporal relationship between TFSs and gestures in Supervisor 6.

On the whole, it shows a rather high rate of co-occurrence of both gesture and speech (49 instances out of 64). Among all the instances accompanied by a certain form of hand-and-arm movements, those that co-occur with independent strokes/gestures occupy the largest share with 22 instances out of the total of 49 (approximately 45%). A reasonably good number of instances are identified in the other three gesture phases, although the numbers drop considerably compared to those co-occurring with strokes. These are the post-stroke hold phases (9 instances, 20%), embedded in stroke phases (7 instances, 14%) and preparation phases (6 instances, 12%). The TFSs occurring with the other three types of gesture phases are extremely minor. Only three TFSs co-occur with retraction phases, one with a post-stroke hold + retraction phase, and one with a post-stroke hold + preparation phase.

4.3.2. Identifying gesture patterns based on functional variations

The temporal analysis of speech and gesture phases demonstrated above affords the basis for a more detailed investigation of the relationship between TFSs and the types of gesture phases in alignment with them. This section will demonstrate the findings of the tendency of certain speech functions co-occurring with certain gesture phases. In the current case study, three main categories of such relationships have been identified, which include:

a). TFSs as checking understanding with beat/beat-like strokes, beat/beat-like strokes with long post-stroke holds and post-stroke hold phases;

b). TFSs as a combination of checking understanding and discourse marker and embedded in stroke phases;

c). TFSs as discourse markers with no gesture.

The following sub-sections will be devoted to demonstrating each type in sequence. All the gestures co-occurring with the TFSs in Video 6 were transcribed and can be observed in File 4.2 in the CD-ROM data disk.

4.3.2.1. TFSs as checking understanding with beat/beat-like strokes, beat/beat-like strokes with long post-stroke holds and post-stroke hold phases.

Table 4.2 below demonstrates all the gesture phases in line with the TFSs as a means of checking understanding.

Table 4.2: Gesture phases co-occurring with the TFSs as checking understanding.

Beat/beat-like stroke	**11**
Beat/Beat-like stroke with long post-stroke hold	**5**
Post-stroke hold phase	**7**
Circular stroke	2
Preparation phase	1
Retraction phase	1
No gesture	7
Total	34

The results show that the TFSs as checking understanding tend to co-occur with particular types of gestures. Except for the 7 instances unaccompanied by any gesture, it is somewhat evident that a dominant number of those TFSs checking understanding are synchronised with beat/beat-like strokes (11), beat/beat-like strokes with long post-stroke holds (5) and post-stroke hold

phases (7). Only two instances are accompanied by circular strokes (i.e. rotating hand), one accompanied by a preparation phase and the other by a retraction phase. In this sub-section, four examples will be discussed in detail, among which the first three were chosen from the three types at the top of the table. The fourth instance will be a circular gesture as a recurring gesture in real life, though only co-occurring twice with TFS checking understanding.

Among all the instances of TFSs checking understanding co-occurring with the strokes (18 instances in total), beat/beat-like strokes/gestures are the dominant gesture types with a total number of 16 instances. 5 out of the 16 beat/beat-like gestures have long post-stroke hold phases. Beat/beat-like strokes refer to those motions that mainly involve up and/or down, left and/or right, inward and/or outward movements, etc., which can be performed single or two-handed. The movements and velocity of the hand are mostly symmetric in standard beat gestures, but asymmetric in beat-like gestures.

Speech episode (V6-6):

<$1> C= I mean would it be that that's what I was hoping cou= would it be better to start as your kind of introductory paragraph just just hang fire a bit on the space theory but to start the introduction paragraph by just saying why are ho= why are hotels as a particular kind of building useful to novelists+

<$2> Right.

<$2> + ***do you know what I mean (****<$2>: nodding****)*** *just forget the actual specifics of space theory and do you remember the ideas we were talking about in your other piece of work+*

<$2> Mhm.

Above represents a typical example of a beat gesture co-occurring with a TFS checking understanding. In this instance, the supervisor is specifying a correct

way of beginning a chapter by straightforwardly underlining the significance of hotels in writing this particular chapter (see the speech transcript above). The speech analysis reveals an interaction where the supervisor intends to check the understanding of the student by saying “do you know what I mean” after a detailed demonstration and the student confirms instantly by head nodding. Hence, what unfolds here is a clear speaker-checking-and-listener-confirming sequence of communication, which illustrates the function of the current TFS as a means of checking understanding.

Below (Figure 4.12) is the gesture in line with this instance. The hand holds in a post-stroke hold position when the supervisor is saying “novelists” (a-b). The right hand is open and held vertically with the fingers slightly bent, facing the left side of the body. With the same gesture form, she quickly moves the hand to the left and then to the right in a symmetrical fashion as she is saying “do you know what I mean” (c-d). This stroke counts as a typical left-right beat gesture. Instantly after she finishes the TFS, the hand relaxes and moves up to an upper position in front of her chest during “just for(get)”, which is the preparation phase of the next gesture.



Figure 4.12: A TFS checking understanding co-occurring with a beat gesture.

There are four instances of beat gestures co-occurring with the TFSs checking understanding, these are V6-6, V6-7, V6-15 and V6-17 (see Table 4.3 below).

A close temporal examination between gesture and speech among the four reveals a very close coordination between the verbal and gestural modes as the strokes tend to start immediately at the beginning of the TFSs and terminates close to their end. The strokes in the following table are colour-coded in red, or in bold for black and white print.

Table 4.3: Temporal relationship between the TFSs checking understanding and beat gestures.

Item	Speech
V6-6	useful to novelists+ <$2> Right. <$1> + **do you know what I mean** just forget
V6-7	anges space in different ways+ <$2> Mhm. <$1> + **do you see what I mean /just and**
V6-15	of work simply grown out of modernist spaces **do you know what I mean is it very s**pecific to explaining (pause 0.4) modernist
V6-17	the nineteenth century lead in? <$2> Mhm. (pause 0.1) <$1> **Do you know what I mean/in th/is par**ticular chapter

Among the 11 instances of TFSs co-occurring with beat/beat-like gestures, there are 7 beat-like gestures. Beat-like gestures differ from typical beat gestures as the movement in the two directions are not symmetrical, usually the latter half appearing stronger than the former. Hence, the former half was coded as the preparation phase and the latter half as the stroke. The preceding example of V6-13 is such a gesture, as well as the instance V6-8 (see Figure 4.13 below).

From the textual analysis (see the speech transcript below), the supervisor is demonstrating how the student can first present the themes of the thesis by describing how space becomes an interactive platform for people of distinctive social classes and identities. After a detailed and lengthy

explanation regarding this, she uses the TFS to check the mutual ground, which effectively triggers a positive response from the student. Such an interaction then suggests the function of checking understanding in this TFS.

Speech episode (V6-8):

> *<$1> +and th= the reason why the architectural space is interesting is primarily because of the way it brings together different classes and it's the interaction of different classes which in turn has effects on the identity of individuals+*
>
> *<$2> Mhm.*
>
> *<$1> +so y= you can kind of introduce almost like the big themes of your thesis* ***do you see what I mean*** *(<$2>: head nods) just just in the first ch= first I'm talking about the first hundred words or so*

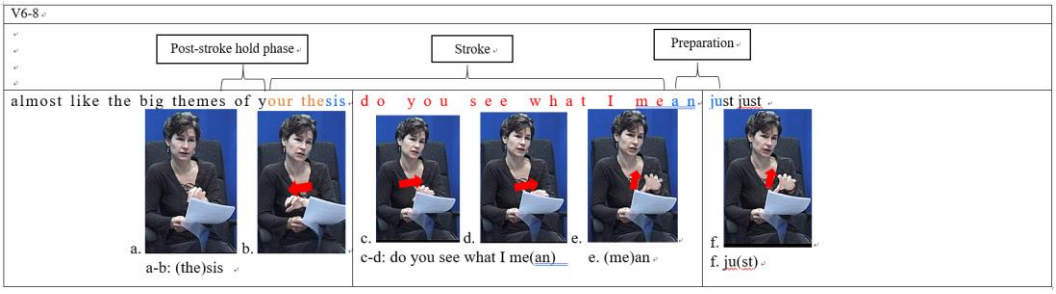


Figure 4.13: A TFS as checking understanding co-occurring with a beat-like gesture.

In terms of the gestures, the hand was in a post-stroke hold phase before the preparation (a-b) starts, with the thumb and the index finger of the left hand touching each other and the other three fingers bent and separated. Immediately before the TFS, she moves the hand slightly and gently to the right (a-b), which prepares for the more forceful movement towards the left (c-d) when she is saying “do you see what I (me)”. This fast-moving, one-direction beat motion resembles a typical beat gesture yet lacks the other

symmetrical half. Hence, gestures as such are called beat-like gestures. As she approaches the end of the TFS (e), she moves the hand gently upwards and outwards (e-f), preparing for the stroke movement of the subsequent gesture.

Examining the temporal alignment between the TFS and the beat-like gesture illustrated above, the close coordination between the two modes of expression can be easily noticed. As demonstrated in Table 4.4 below, such a temporal coordination between the stroke phase and the TFS checking understanding can also be found in most of the other 6 instances, except for V6-60. The strokes tend to start right at the beginning of the TFSs and terminate close to their end.

Table 4.4: Temporal relationship between the TFSs checking understanding and beat-like gestures.

Items	Speech
V6-5	tand what's going in the hotel+ <$2> Mhm. <$1> + do you **see what I mean** the hotel is your your primary focus of analysis
V6-8	almost like the big themes of your the**sis do you see what I mea**n just just
V6-13	is quite a sort of detailed literary historical **do you see what I mean** + <$2> Yeah. <$1> +and it's almost like
V6-18	to be fundamentally different <$2> Mhm. <$1> + **do you see what I/ mean from how** how environment and location is used
V6-54	the adequacy of the claim you're mak**ing do you see wha**t I mean in relation in relation to+ <$2> Yeah. <$1> +the evidence.
V6-59	can say how that whole debate is reconfigured **do you see what I mean** + <$2> Yeah. <$1> + via the hotel
V6-60	s to be actually reconfigured+ <$2> Mhm. <$1> + **do you see** what I mean one of

In addition to the beat/beat-like gestures, 5 instances of beat/beat-like strokes with long post-stroke holds have been identified. They are slightly different from the beat/beat-like gestures in that the beat movement only occurs at the beginning or the former half of the TFSs, with the rest of the stroke completed by a long post-stroke hold. Therefore, the temporal gesture-speech alignment in this gesture type is primarily achieved due to the post-stroke holds following the beat-like motions.

The following example represents such a type of gesture. Similar to the previous two examples, the speech context of this TFS suggests a function of checking understanding marked by a speaker-checking-and-listener-confirming sequence of speech unfolding as the supervisor is specifying the purpose of the chapter.

Speech episode (V6-11):

<$1> I think I think that might give you the focus better and then this chapter will become more a chapter that's setting up the theorisation of space ***do you see what I mean****+*

<$2> Yeah.

<$1> +and the historical issues of space I mean it's inevitably going to have to touch on the class issues a bit isn't it and then you can kind of+

<$2> Well that's how <$G?>+

<$1> +pick them up later.

Figure 4.14 below presents the gesture related to this. As can be observed, as the supervisor is saying “(theorisa)tion of spa(ce)” (a-b), the hand is open and held steady in a post-stroke hold phase. Directly before the TFS starts, she quickly moves the hand outwards and downwards (c), the movement of which takes place at the beginning of the TFS “do you”. This forceful but fleeting beat-like movement thus counts as the stroke phase of the gesture

co-occurring with the TFS. The rest of the TFS, “see what I mean” (e-f), is accompanied by a post-stroke hold phase. As the TFS finishes, she immediately lifts the hand (g-h) while saying “and” and prepares for the next stroke.

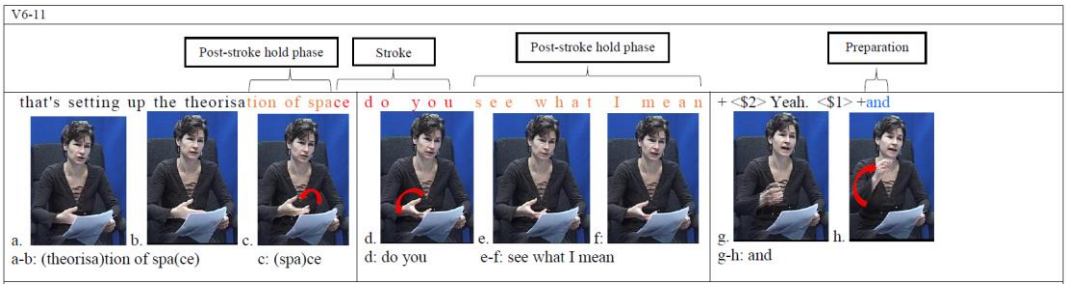


Figure 4.14: A TFS checking understanding co-occurring with a beat-like gesture with a long post-stroke hold phase.

Table 4.5: Temporal relationship between the TFSs checking understanding and beat/beat-like gestures with long post-stroke hold phases.

Items	Speech
V6-11	that's setting up the theorisation of spa**ce do you** see what I mean+ <$2> Yeah. <$1> +and
V6-19	s are fundamentally different+ <$2> Mhm. <$1> + **do you know** what I mean s so instead so instead of having
V6-21	make generalisations that will apply right across the period do you see what I mean +<$2> Mhm. <$1> +or
V6-22	and you can appropriate that body of theory do you see what I mean+ <$2> Mhm. <$2> +and
V6-29	nly really notice it for these particular reasons do you see what I mean + <$2> Yeah. <$1> + it's just a frame

The alignment between speech and gesture in the above example (V6-11) is achieved by the combination of the stroke phase and the post-stroke hold. As

is shown in Table 4.5 (presenting the temporal relationship between the TFSs and the gestures), the same phenomenon also applies to the remaining instances. The strokes have been colour-coded in red and post-stroke holds in orange. For black-and-white print, the strokes are in bold and the post-stroke holds are underlined.

With regards to those TFSs checking understanding co-occurring with post-stroke hold phases, the holding period tends to start slightly prior to the TFSs and continues to the end. The entire post-stroke hold phases act as prolonged parts of the preceding strokes that are co-expressive of the preceding noteworthy information. For this type of gesture, the instance V6-26 will be used for demonstration. As can be seen from the speech context below, the current TFS occurs at the end of a complex demonstration, the comprehension of which is confirmed by the student's head nodding. This indicates the function of checking understanding in this TFS.

Speech episode (V6-26):

> *<$1> I think either really either it's just a way of erm just a way of understanding it I mean cos what one of the things you're saying about modernist work is that space as you said becomes a way of understanding character and characters kind of meld with the space. Now it strikes me that's that's a general theoretical idea* ***do you see what I mean*** *(<$2>: head nods) that that is how space can be used that space is not something just neutral that you put an individual in+*
> *<$2> Mm.*

The gesture accompanying this TFS can be seen in the figure below (Figure 4.15). As she starts saying "that's a" (a-b), she slowly raises her elbow and hands until the palms face downwards and the index fingers and, presumably, the thumbs touching each other. During the stroke phase co-occurring with

"a general theoretic(cal)" (c), she quickly drops both hands downwards and inwards to a lower position until they are touching each other. The two hands during this post-stroke hold phase form a bounded space between the hands as if holding something. This hand shape lasts from the utterance slightly before the TFS, "(theoreti)cal idea", throughout the TFS (d-f), and slightly afterwards (g). The hands are then separated during a 0.2-second pause and "tha(t)" (h), preparing for the next stroke. This bounded-space gesture form in the stroke seems to be iconic of the image of holding "a general theoretical idea" by the hands. Therefore, the space formed by the hands represents the conceptual idea under demonstration, which is held and foregrounded throughout the entire TFS. As the TFS finishes, the gesture changes.

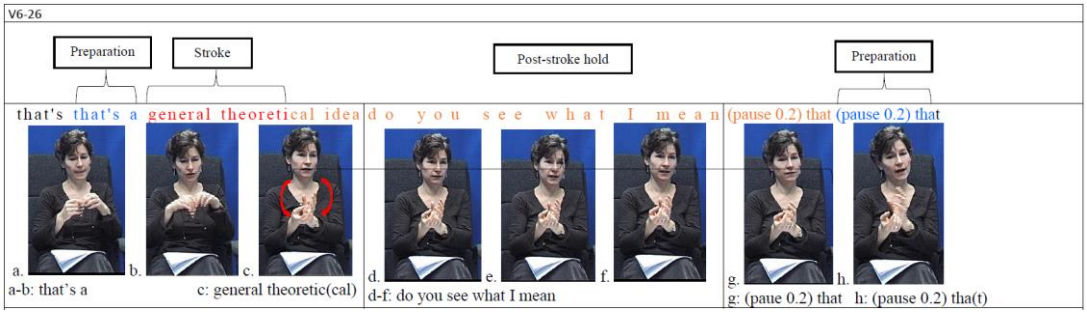


Figure 4.15: A TFS as checking understanding co-occurring with a post-stroke hold phase.

According to the table below (Table 4.6), the assessment of the temporal coordination between the TFSs checking understanding and the post-stroke hold phases in alignment with them reveals that the holding phases tend to start before the TFSs and last around the end of them. Again, the strokes are in red and in bold, and the post-stroke holds are in orange and underlined.

A vital insight gained from the previous analysis is the high occurrences of post-stroke hold phases (12 instances) co-occurring with the TFSs including those following the beat/beat-like strokes (5) and those post-stroke hold phases attached to the preceding strokes (7). What can be inferred from

this is the strong association between the meaning of checking understanding and holding. To confirm this assumption, a further pilot study was conducted, which will be discussed later in 4.5.2.1.

Table 4.6: Temporal relationship between TFSs checking understanding and post-stroke hold phases.

Items	Speech
V6-26	that's that's a **general theoreti**cal idea do you see what I mean (pause 0.2) that (pause 0.2) that
V6-27	in which the idea was developed to **explain** do you see what I mean+ <$2> Right yeah. <$1> So= so if you say take those ideas about the hotel lobby
V6-33	space th**eory is not** influencing the novel+ <$2> Yeah. <$1> + do you see what I mean cos space theory that that's where you've got be really precise
V6-36	mind is that this change ac**tually has ha**ppened do you know what I mean + <$2> Yeah. <$1> + and that when a reader is is is looking at
V6-44	got to think o**f it like f**eminist theory do you see what I mean+ <$2> Yeah. <$1> or Marxist theory or structuralism
V6-51	chapter to be wri**tten which actually links** Bowen do you know what I mean in a funny way+
V6-61	<$2> Mhm. <$1> Now that **strike**s me as a thesis (pause 0.3) do you see what I mean+ <$2> Yeah. <$1> +and and you've got

In the preceding part of this section, the main gesture phases in line with the TFSs have been demonstrated, particularly beat/beat-like gestures and holding. Compared to the top three types of gestures, the others are minor with 2 instances of circular gestures (V6-2, V6-9), 1 instance each of the retraction phase (V6-1) and the preparation phase (V6-53). Whereas it seems unnecessary to provide instances of the other two, one of the circular

gestures, V6-9, is presented in Figure 4.16 below as a particular recurrent gesture in daily life. The speech context (see the speech episode below) shows the occurrence of the TFS at end of a detailed demonstration regarding the contents at the beginning of the chapter, which is immediately followed by the student's response of agreement. This context, again, seems to point to the function of this TFS as checking understanding.

Speech episode (V6-9):

> *<$1> +so y= you can kind of introduce almost like the big themes of your thesis do you see what I mean just just in the first ch= first I'm talking about the first hundred words or so so that you can then move on to that and say well what do we actually need theore= or what do you need theoretically to actually allow you to investigate* ***do you know what I mean****. What approaches+*
>
> *<$2> Okay.*
>
> *<$1> +do you need in order to understand why the hotel has been useful to these particular writers and why its particular architectural space+*
>
> *<$2> Mhm.*

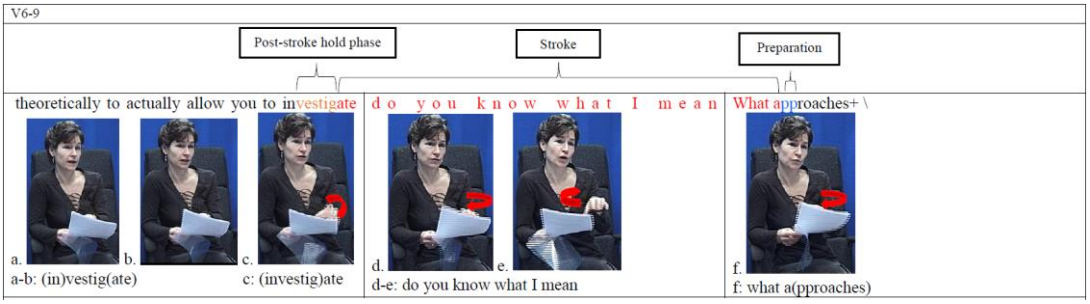


Figure 4.16: A TFS as checking understanding co-occurring with a circular gesture.

As for the circular gestures co-occurring with this TFS, the hand is held steady above the papers in a post-stroke hold phase (a-b) as the supervisor is

saying “(in)vestig(ate). Immediately before the beginning of the TFS, she begins to rotate the hand outwards horizontally with the hand open and the palm facing down (c). The hand rotates almost twice during the TFS (d-e), and stops immediately after it (f). As can be observed from the example above and the table below (Table 4.7), the temporal relationship between the TFSs checking understanding and the two circular gestures also demonstrate a relatively close coordination between speech and gesture.

Table 4.7: Temporal relationship between the TFSs checking understanding and the circular gestures.

Items	Speech-gesture alignment
V6-2	get into the very beginning of this chapter do you **know what I mean it's what ki**nd of question do you start off with
V6-9	theoretically to actually allow you to investig**ate do you know what I mean What a**pproaches+\

4.3.2.2. TFSs as a combination of checking understanding and discourse marker and embedded in stroke phases

The preceding sub-section reveals the fact that TFSs checking understanding tend to co-occur with three types of movement patterns, which entail beat/beat-like gestures, beat/beat-like gestures with long post-stroke hold phases, and post-stroke holds. The tendency of TFSs dominantly co-occurring with certain gesture phases has also been identified in TFSs as a combination of checking understanding and discourse marker. As is shown in the following table (Table 4.8), more than one third are coordinated with bodily movements occurring in the midst of large-sized stroke phases (7/19). Four co-occur with preparation phases, 2 with post-stroke hold phases and 2 with beat/beat-like strokes. With a much lower number of independent strokes associated with

TFSs of this function, the overall gesture types are likely to be less salient than those coordinated with TFSs checking understanding.

Table 4.8: Gesture phases co-occurring with the TFSs as a combination of checking understanding and discourse marker.

Embedded in stroke phase	**7**
Preparation phase	4
Post-stroke hold phase	2
Beat/beat-like stroke	2
Circular gesture	1
Retraction	1
No gesture	2
Total	19

The remainder of this sub-section will provide two examples of the two types of gesture phases, embedded in stroke phase (V6-10) and preparation phase (V6-20), which tend to co-occur more than the others with TFS as a combination of checking understanding and discourse marker.

Below is an example of TFS as a combination of checking understanding and discourse marker co-occurring with an embedded in stroke phase. Based on the discourse analysis of the interactions, the supervisor is attempting to explain to the student that she needs to explore why the hotel afforded opportunities to the writers in her research that would have not been available to earlier writers in the nineteenth century. The TFS is inserted in the middle of a lengthy utterance where the meaning has not been completed. This position of the TFS, therefore, seems to indicate the function of a discourse marker, or a filler. However, immediately after finishing the whole sentence, the student replies “Yeah”, which forms part of a

speaker-checking-and-listener-confirming sequence. This then suggests the function of this TFS as a means of checking understanding.

Speech episode (V6-10):

<$1> ... What approaches+

<$2> Okay.

<$1> +do you need in order to understand why the hotel has been useful to these particular writers and why its particular architectural space+

<$2> Mhm.

<$1> +affords writers possibilities that weren't afforded ***do you know what I mean*** *to writers early in the nineteenth century.*

<$2> Yeah.

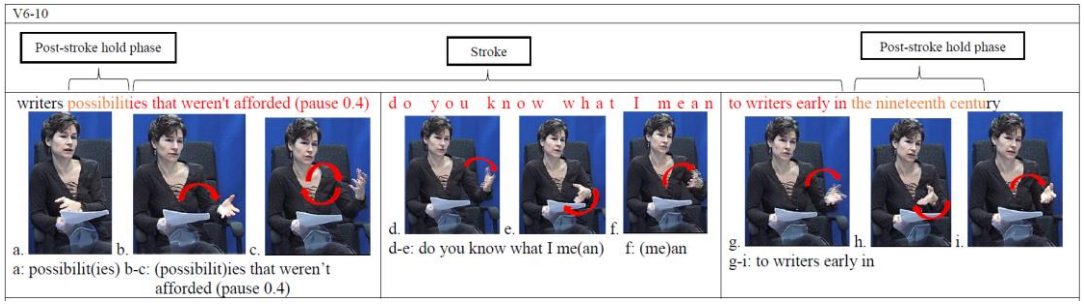


Figure 4.17: A TFS as a combination of checking understanding and discourse marker co-occurring with an embedded in stroke phase.

The gesture aligned with this TFS is presented in the figure above (Figure 4.17). Before the onset of this circular gesture, the hand is in a post-stroke hold position as she is saying “possibilit(ies)” (a), after which she starts to rotate her hand for a few times from before the TFS (b-c), throughout it (d-f) and after (g-i). The entire circular gesture lasts for a relatively long time, accompanying a lengthy utterance. Hence, as the speech part of the TFS is inserted in the middle of a long sentence, the hand-and-arm movements also occur in the middle of a long circular gesture.

Due to the limited number of instances, no concrete conclusion is attempted here concerning the functional association between embedded in stroke phase gestures and the current TFS function. However, the temporal gesture-speech relationship coordinates well with the functional relationship, and the same phenomenon can be observed in the remaining instances (see Table 4.9 below). As the strokes are in red and in bold, it can be seen that both the speech and gesture of the TFSs as a combination of checking understanding and discourse marker tend to be embedded in the middle of lengthy sentences and strokes.

Table 4.9: Temporal relationship between the TFSs as a combination of checking understanding and discourse marker and embedded in stroke phases.

Item	Speech
10	fords writers possibilit**ies that weren't afforded (pause 0.4) do you know what I mean to writers early in** the nineteenth century
14	it it's how you**'re going to do you see what I mean how you're going to integrate those two approaches** in one chapter
24	's a way of using a theoretical **tool do you see what I mean to interpret the significance of the histo**rical evidence
41	you keep trying to **worry away at that do you see wha**t I mean in this particular chapter
55	his particular way + <$2> Yeah. <$1> +**to explore do you see what I mean** aspects of aspects of psychology
62	a question mark you're **like (pause 0.2) trotting off+ <$2> <$E>Laughs</$E> <$1> + do you know what I mean down** a different track
64	's been raised here but try and r**esist (pause 0.1) do you know what I mean resist** the temptation of thinking this is interesting too

In addition to the 7 embedded in stroke phases, there are TFSs as a combination of checking understanding and discourse marker co-occurring with four preparation phases. Using V6-3 as an example, the speech episode of this instance shows that the supervisor is pointing out a problematic issue as the student used an excessive amount words to identify a theoretical gap in the chapter, this being the lack of research adopting space theory in the analysis of the twentieth century novel.

The speech episode (V6-3) below demonstrates that the TFS is embedded in the middle of this statement where the sentence has not yet been completed. The student's confirmation of understanding also occurs after the end of the entire sentence. Similar to the previous example, the TFSs in this type of discourse context suggest a dual function of the TFS combining the functions of checking understanding and discourse marker.

Speech episode (V6-3):

> *<$1> +it's what kind of question do you start off with cos it struck me there's a there's a slight tension at the moment between your argument that space theory hasn't been applied to the twentieth century novel* ***do you know what I mean*** *the chapter somehow becomes about space theory+*
> *<$2> Mhm.*

The gesture of this instance is presented in the figure below (Figure 4.18), which is a preparation phase preceding a circular stroke. The hand is held still in a post-stroke hold phase with the palm facing downwards and the fingers extended when she is saying "(cen)tury novel" (a-b). The moment she starts to say "do you know what I mean", she gradually raises her hand (c), slightly moves down (d-e) and then lifts the hand up to a position almost parallel to her neck after the TFS (f). The palm rotates from facing downwards to inwards as

she is lifting the hand. This is the preparation phase of the following stroke, which co-occurs with "(chap)ter somehow becomes about sp(ace)" (g-h). The hand rotates up and down about twice in the stroke.

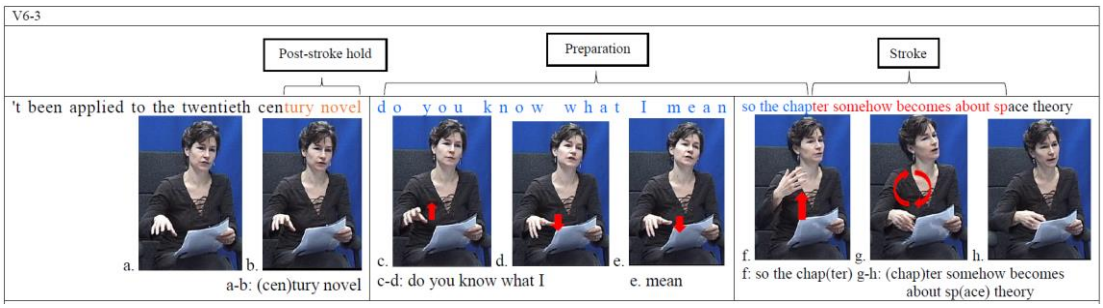


Figure 4.18: A TFS as a combination of checking understanding and discourse marker co-occurring with a preparation phase.

Table 4.10: Temporal relationship between the TFSs as a combination of checking understanding and discourse marker and preparation phases.

Items	Speech
V6-3	't been applied to the twentieth century novel **do you know what I mean so the chap**ter somehow becomes about space theory
V6-45	of got rid of the modernism hang up do you see **what I mean**+ <$2> Yeah. <$1> +**you can** move much more seamlessly
V6-46	it's a transition which is taking pla**ce do you know what I mean** over a long longish period of time
V6-63	do this is the aim of this chapter **do you know what I** mean set out in the introduction

Examining the temporal relationship between speech and gesture of this instance, the preparation phase generally starts at the beginning of the TFSs and ends after the end. Such a temporal coordination can also be identified in other instances in Table 4.10 above, though half of V6-45 is accompanied by a preparation phase (the preparation phase is in blue and in bold). Due to the

limited number of examples, no firm link is established between this function of TFSs and the preparation phases, and further research is required on this matter.

4.3.2.3. TFSs as discourse markers with no gesture

Whereas the previous sections illustrate how the TFSs checking understanding and the combination of checking understanding and discourse markers are likely to be synchronised with certain types of gestures, as is shown in the table below (Table 4.11), no gesture type is particularly foregrounded for TFSs as discourse markers. All gesture types only occur once, including the beat-like gesture, preparation phase, retraction phase, post-stroke hold phase + retraction phase and post-stroke hold phase + preparation phase. However, this may also be due to the minimal number of instances in the current case study and, therefore, further exploration of this matter with a much larger corpus may be illuminating.

Table 4.11: Gesture types co-occurring with TFSs as discourse markers.

Beat-like stroke	1
Preparation phase	1
Retraction phase	1
Post-stroke hold phase + retraction phase	1
Post-stroke hold phase + preparation phase	1
No gesture	**6**
Total	11

However, what can be inferred from the current figures is that more than half of the TFSs (6/11) used as discourse markers are employed by the speaker without any accompanying gesture. Furthermore, taking into account all the

instances in the entire video, this means that 6 out of 15 TFSs without gestures belong to those TFSs as discourse markers. These observations, therefore, suggest that, when TFSs are adopted by the speaker as discourse markers with little intention of checking understanding, they tend not to be accompanied by any gesture.

Below is an example of a TFS as a discourse marker without being accompanied by any gesture. The speech episode shows that this TFS occurs immediately after a question raised by the supervisor, yet the question has not been fully constructed. At this point, there is little intention on the part of the supervisor to check the mutual ground between her and the student. The TFS is more of a discourse marker that foresees the coming of a question and draws the attention of the student to this. The student does not respond directly after the TFS, but after the entire question has been clearly explained and articulated.

Speech episode (V6-16):

> *<$1> Yeah I mean tha= that works but I'm still mmm I'm still think= where <$E>coughs</$E> where do you want to put the historical stuff in then* ***do you know what I mean*** *cos one argument might be if this space theory is very very specific to the early twentieth century stuff why do you need the nineteenth century lead in?*
> *<$2> Mhm.*

The state of the hand and arm is presented in the figure below (Figure 4.19) where both hands are holding the papers. One counter-argument would be that, when the hands are holding something, this might influence the availability and mobility of the hands to perform gestures. However, other instances show that it is not uncommon for the supervisor to perform gestures with the existence of the papers. She either put them on her lap or uses one hand to

execute gestures or even performs a gesture while holding them. Hence, the need to hold the papers cannot fully account for the lack of gesture; the discourse context may have more impact in this regard.

a.  b. 

Figure 4.19: A TFS as a discourse marker co-occurring with no gesture.

For those TFSs as discourse markers co-occurring with certain forms of gesture, it is noticeable that the degree of temporal alignment between gesture and speech can vary (see Table 4.12), some are more congruent that others. The gesture phases co-occurring with the TFSs are in the second column and all the target gesture phases are colour-coded and in bold.

A more congruent gesture-speech alignment was found in the instance V6-4 (a beat gesture), V6-48 (a retraction phase) and V6-49 (a post-stroke hold + retraction phase). The extended preparation phase in V6-30 is caused by the dysfluent speech at the beginning of the utterance. In this case, the TFS is mostly likely adopted as a cohesive device or filler as the speaker is processing speech. Also, a TFS accompanied by two different gesture phases, such as the instance V6-50, can also result in a less congruent speech-gesture alignment.

Table 4.12: Temporal relationship between TFSs as discourse markers and gesture phases.

Items	Gesture phases	Speech
V6-4	beat-like gesture	chapter being about hotels (0.2) + <$2> Right. <$1> + **do you see what I mean/and why are h**otels an important issue
V6-30	preparation phase	one by aristocratic men. <$2> Mhm. <$2> **So it's do you know what I mean you you could** just describe that very briefly
V6-48	retraction phase	a a historical chapter which just goes back **do you know what I mean us**ing basic
V6-49	post-stroke hold phase + retraction phase	a historical answer to that so you could do **do you see what I mean (pause 0.4)** that kind of precise erm piece of analysis
V6-50	post-stroke hold phase + preparation phase	chapter that is ac**tually a chapter on class do you see what I mean and th=** and the servant and the detective is really

A further important observation is that, although instances V6-49 and V6-50 also partly co-occur with a post-stroke hold phase, the two post-stroke hold phases here differ from those co-occurring with the TFSs as checking understanding in that they are shorter and cut off by other gesture phases. Due to the shorter period of time in the two post-stroke holds, they appear less salient than their longer counterparts. As will be demonstrated later (in 4.5.2.1.), the length of post-stroke hold matters as the longer post-stroke hold phases are more likely to be associated with the function of checking understanding than shorter ones.

The analysis in the sub-sections of 4.3.2.1., 4.3.2.2., and 4.3.2.3. reveals that there seems to be a fundamental difference between the distinctive TFS functions regarding the types of gestures associated with them. TFSs checking understanding are closely tied to beat/beat-like gestures, beat-like gestures with long post-stroke holds and post-stroke hold phases. TFSs as a combination of both functions may relate to motions inserted in lengthy strokes, whereas TFSs as discourse markers tend to be used without any gesture. Nonetheless, some gesture types can emerge in all the three functions, including beat/beat-like stroke, preparation phase, retraction phase. Also, all the three functions contain instances of TFSs that are not accompanied by any gesture.

The occurrence of overlapping gestures is worthy of particular attention as the gesture can affect the meaning of the multimodal unit. That is, among TFSs with the same function, the meaning of the whole multimodal unit can be strengthed or diminished by the form of gesture co-occurring with them. The meaning of those TFSs checking understanding co-occurring with independent strokes (e.g. beat/beat-like gestures) seems to be much more salient than those accompanied by a preparation phase or a retraction phase. Likewise, the meaning of a TFS as a discourse marker can be strengthed by co-occurring with a beat/beat-like gesture or a post-stroke hold. Below is an instance of such, in which a TFS as a discourse marker co-occurs with a beat/beat-like gesture and therefore the meaning of the entire multimodal TFS moves slightly towards the meaning of checking understanding.

The speech episide below shows that the supervisor is demonstrating her concern that the chapter which is supposed to be about hotels has excessively focused on space theory. Immediately before the TFS in Line 9, the student has already replied twice in Line 5 and 6, confirming her understanding of the point made by the supervisor. The TFS emerges at the beginning of a turn, after which the student does not provide any instant response. Therefore, the

position of the TFS and the feature of interaction between the interlocutors suggests that it is functioning as a discourse marker with a very low intention of checking understanding.

Speech episode (V6-4):

<$1> +it's what kind of question do you start off with cos it struck me there's a there's a slight tension at the moment between your argument that space theory hasn't been applied to the twentieth century novel do you know what I mean the chapter somehow becomes about space theory+

<$2> Mhm.

<$1> +and utility of space theory for a literary critic as opposed to the chapter being about hotels+

<$2> Right.

<$1> + ***do you see what I mean*** and why are hotels an important issue cos I think you've you've got to make a decision are you using space theory as a mechanism to help you understand what's going in the hotel+

<$2> Mhm.

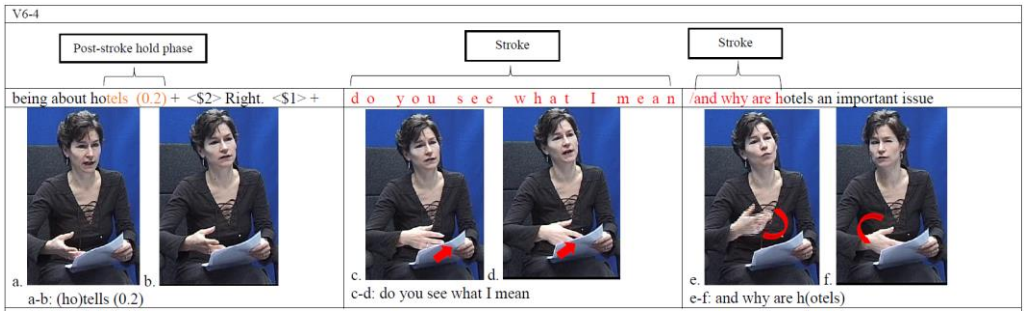


Figure 4.20: A TFS as discourse marker co-occurring with a beat-like stroke/gesture.

In terms of the beat/beat-like gesture that aligns with this TFS (Figure 4.20), before the TFS, the right hand remains still in a post-stroke hold phase with the

palm open and facing upwards (a-b). Instantly as she begins to say “do you see what I mean”, the hand rapidly moves inwards towards the speaker’s body (c-d). This is the stroke phase of this gesture. The moment she starts to say “and why the h(otels)” (e-f), a new stroke begins, where she rotates the hand inwards and outwards twice.

In brief, although the speech component of this TFS seems to mostly demonstrate the function of a discourse marker, the beat/beat-like gesture encodes the intention of checking understanding. Hence, the intention of checking understanding delivered by the whole multimodal unit seemingly becomes stronger due to the performance of beat/beat-like gesture.

4.4. Results of the extended data analysis

Although the main remit of the first study is to use this specific speaker to explore an approach that can establish a speech-gesture profile for the TFSs, further analysis of the same TFSs in other speakers can be valuable in terms of discovering the extent the patterns identified in the case study can also be applied to others. The results represent some similar characteristics of use among different speakers, however, they also indicate the requirement for further study using a much larger corpus involving many more instances and participants.

The search for the TFS was conducted in the remaining 12 supervision meetings (8 supervisors and 11 students) except for Video 6. In total, only 12 instances were found in the other three speakers, all supervisors, which is much less than expected (Table 4.13). Most of the 12 examples are contributed by Supervisor 3 (7 instances from two videos) and Supervisor 4 (4 instances from two videos); one from Supervisor 2.

Table 4.13: Instances of TFSs from other speakers

Items	Speaker	Video	TFSs
1	Supervisor 2	V2-1	you know what I mean
2	Supervisor 3	V3-1	Do you see what I mean
3	Supervisor 3	V3-2	do you see what I mean
4	Supervisor 3	V3-3	you know what I mean
5	Supervisor 3	V3-4	do you see what I mean
6	Supervisor 3	V3-5	you see what I mean
7	Supervisor 4	V4-1	you know what I mean
8	Supervisor 3	V11-1	you know what I mean
9	Supervisor 3	V11-2	you know what I mean
10	Supervisor 4	V12-1	do you see what I mean
11	Supervisor 4	V12-2	do you see what I mean
12	Supervisor 4	V12-3	do you see what I mean

In view of the small number of instances of the TFSs found in other speakers, no general claims are attempted here. As emphasised above, the goal is to simply compare different speakers with Supervisor 6 in order to examine the extent to which the findings of the case study can be applied to a new set of data. For future research, many more videos involving different speakers are required to generate a significantly larger number of instances of this TFS.

4.4.1. Temporal relationship between TFS and gesture

In the case study, the results of the temporal relationship (see sub-section 4.3.1.) between the TFSs and the gestures reveals a high percentage (45%) of instances co-occurring with independent gestures/strokes. This is confirmed in the analysis of other speakers, i.e. a strong tendency of co-occurrence has been found between the TFSs and the stroke phases. As is shown in the figure

below (Figure 4.21), 11 out of 12 TFSs are accompanied by gesture phases, among which 10 are independent gestures composed of at least one stroke phase. Other categories are minor with only one TFS co-occurring with a post-stroke hold phase and one without any gesture.

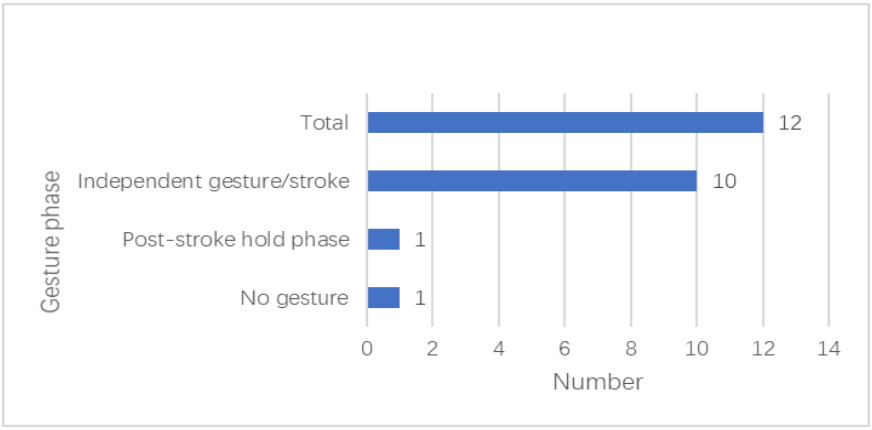


Figure 4.21: Distribution of temporal relationship between TFSs and gestures in other speakers.

4.4.2. Identifying gesture patterns based on functional variations

The results of the temporal relationship between the TFS and gestures found in other speakers largely resemble those found in the case study, i.e. the TFSs tend to co-occur with independent strokes/gestures. Nevertheless, the findings of the functional coordination between speech and gesture are much more complex. This, again, perhaps results from the small number of instances found in the NMMC.

Three speech functions have been identified in the case study, with only two of these, checking understanding and discourse marker, have been found in other speakers in the extended analysis. Moreover, a new dimension of function emerges in two instances, where the speakers are claiming mutual understanding between the interlocutors without any intention of seeking confirmation from the other speaker. The following sub-sections will focus on

reporting the gestures co-occurring with the three functions of TFSs found in the extended analysis.

4.4.2.1. TFSs checking understanding and their co-occurring gestures

In the 5 instances of TFSs checking understanding co-occurring with independent strokes/gestures (Table 4.14), two are beat/beat-like gestures and one is a post-stroke hold phase. Both have been identified as the most frequently occurring gesture phases synchronised with the TFSs checking understanding in the case study. A new gesture type, Palm Up Open Hand stroke/gesture, emerges; however, only one speaker (Supervisor 3) has used this. One example from each type will be presented as illustrations, following the sequence of V12-1 (Beat-like stroke), V3-4 (Palm Up Open Hand gesture) and V12-2 (Post-stroke hold phase).

Table 4.14: Gesture types co-occur with TFSs as checking understanding in other speakers

Items	Speakers	Videos	Gesture types	Speech function
2	Supervisor 3	V3-1	Beat-like stroke	Checking understanding
5	Supervisor 3	V3-4	Palm Up Open Hand stroke	Checking understanding
6	Supervisor 3	V3-5	Palm Up Open Hand stroke	Checking understanding
10	Supervisor 4	V12-1	Beat-like stroke	Checking understanding
11	Supervisor 4	V12-2	Post-stroke hold phase	Checking understanding

The speech episode below demonstrates the discourse context of this TFS. It occurs at the end of a lengthy demonstration from the supervisor concerning the approach that he took for his own PhD project, which was to establish a new framework drawn from discourse analysis which could be adopted by other researchers to analyse poems. He uses his own experience to illustrate

for the student that she might take a similar approach for her PhD as she is also building up a new analytical framework from different strands of research that could be used by future researchers. This TFS emerges at the end of the supervisor's demonstration and is instantly responded to by the student. This discourse context indicates the function of this TFS as a means of checking understanding.

Speech Episode (V12-1):

> *<$1> And in the end I went from writing about fifty poems to writing about six and that all I did because what you're doing was building the framework rather than interpreting the poems or+*
>
> *<$2> Yeah.*
>
> *<$1> D=* ***do you see what I mean.***
>
> *<$2> Yeah.*

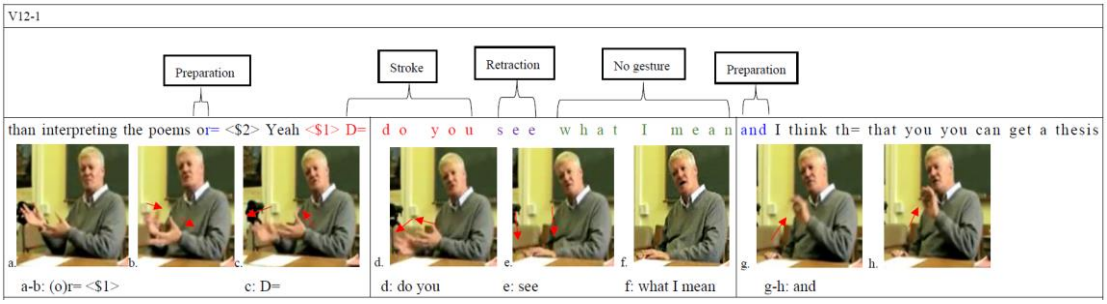


Figure 4.22: A TFS checking understanding co-occurring with a beat-like gesture.

The gesture (Figure 4.22) co-occurring with the TFS is a beat-like stroke. Both hands are facing upwards before the preparation phase, during which he quickly but gently moves towards his body and to the left (a-b). Immediately before the TFS, he rapidly moves both hands to the right side of the body, with the right hand moving slightly downwards and the left upwards. This fast-moving beat-like motion is the stroke of the gesture that co-occurs with

the beginning of the TFS “do you” (d), which is immediately followed by a retraction phase completed during “see” (e). Both hands relax and retreat to a rest position as he is saying “what I mean”.

A further beat-like gesture is performed by Supervisor 3 (V3-1) (see Figure 4.23 below), where he moves the hand downwards and leftwards as he is saying “do you see what I” (refer to File 4.3 in the CD-ROM data disk for details).

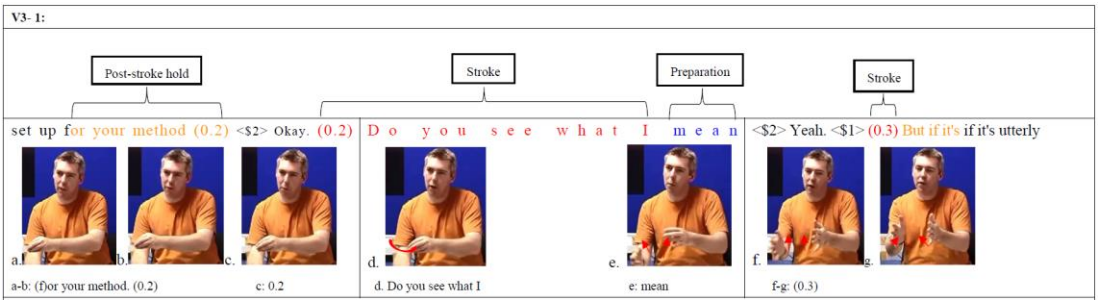


Figure 4.23: A TFS checking understanding co-occurring with a beat-like gesture.

The next instance (V3-4) is accompanied by a Palm Up Open Hand gesture performed by Supervisor 3. Examining the discourse context below, the TFS also occurs at the end of an enquiry, which requires the student’s justification of why she has chosen to conduct a contrastive analysis between doctor and patient rather than merely looking at one party. However, there seems to be an apparent lack of mutual ground between the speakers as the student replies “why?” rather than providing an answer to the supervisor’s question. The supervisor then makes the question more explicit, after which the student eventually provides her answers to his enquiry. In view of this context, this TFS is clearly adopted by the supervisor as an important technique to assure the mutual understanding.

Speech episode (V3-4):

<$1> That's that's why I'm wondering whether the two looking at the two types it suggests it's a contrastive study. Here's how doctors talk about stuff here's how patients talk about stuff now that's an obvious thing if you're going to look at those two sets. If you're not going to do that why are you looking at both doctor discourse on the one hand and patient discourse on the other. ***Do you see what I mean?*** *Wh=why+*

<$2> Why?

<$1> +Why not make your life easier and just look at the doctors' ones?

<$2> Just choose one. Erm well I was only looking at them at the moment cos erm I thought they were both interesting.

<$1> Right yeah. But is that I mean just as a sort of preliminary hunch is do do they do they look very different?

<$2> Yes they do look quite different. Erm although you quite often find patients bringing in things that they've already been told by+

<$1> Oh right yeah.

<$2> +medical practitioners anyway to explain+

<$1> Yeah.

<$2> +what they have been going through. So that's+

<$1> Yeah.

<$2> +sort of the voice of the practitioner through the+

<$1> Yeah yeah. That's internet stuff isn't it where people look up the condition first and come in+

<$2> Yeah.

<$1> +sort of challenge the doctor on "oh I read this stuff from" you know yeah. Whereas presumably that never used to happen.

<$2> Yeah.

The gesture co-occurring with the TFS is presented in the following figure (Figure 4.24). The hand form shown in picture a. maintains throughout the post-stroke hold phase attached to the last gesture as the supervisor is saying "patient discourse on the other" followed by a one-second pause (a-b). In the following 0.2-second pause (b), he rapidly opens the hand turning its palm from facing downwards to upwards, and then extends the hand and arm slightly forwards in the direction of the listener (c). This Palm Up Open Hand stroke starts from a 0.2-second pause directly before the TFS and is temporally synchronous with the former half of the TFS "do you see". During "what I mean" (d), the hand goes to scratch the head and continues after while he is saying "wh=why" (e).

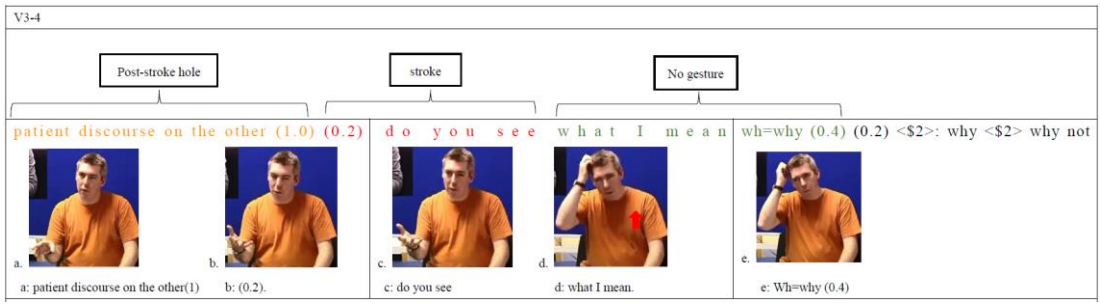


Figure 4.24: A TFS checking understanding co-occurring with an PUOH gesture.

A further PUOH gesture co-occurring with a TFS checking understanding is performed by the same speaker (see Figure 4.25 below), in which the hand is raised and open with the palm facing up during "you know what I mean" (refer to File 4.3 in the CD-ROM data disk for details).

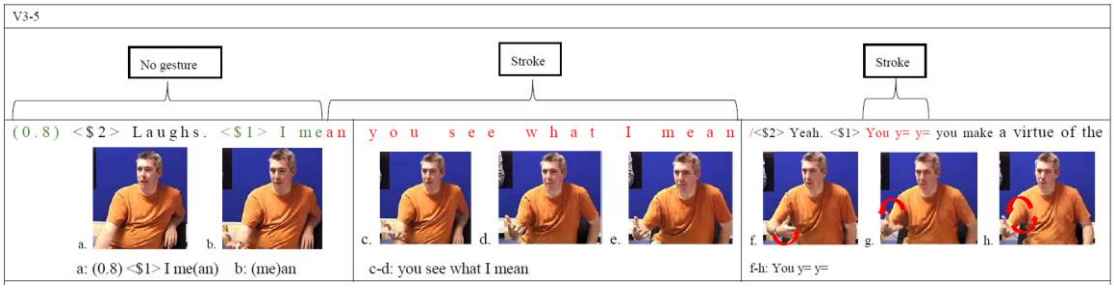


Figure 4.25: A TFS checking understanding co-occurring with an PUOH gesture.

The last example of a TFS checking understanding in other speakers is the one co-occurring with a post-stroke hold phase performed by Supervisor 4. There is a high co-occurrence between TFSs checking understanding and post-stroke holds in the case study, although it only occurs here once. The speech episode provided below shows that the supervisor is asking the student what she thinks her expertise is in the domain of applied linguistics. As the student is still at an early stage of her PhD, this is slightly difficult for her to answer. Probably because of this, the supervisor provides a lengthy explanation and uses this TFS to check the mutual ground between him and the student. Following the TFS is a confirmation of understanding from the student as she finally provides her answer by saying that she regards herself as a sociolinguist.

Speech episode (V12-2):

> *<$1> well you know Lauren's in a an applied linguistics department or at least in a working in an applied linguistics domain erm you know what can we ask Lauren about what's her expertise in linguistically what would you say it was.*
>
> *<$2> Erm.*
>
> *<$1> That's not a catch question+*
>
> *<$2> No.*
>
> *<$1> +and maybe there isn't an answer to it maybe you don't know yet*

because you're still analysing the data <$=> and nothing has. <\$=>

But you know+

<$2> H=

<$1> +if someone asked James Taylor he'd say oh vocabulary.

<$2> Yeah.

<$1> Do you see what I mean?

<$2> I positio+ I suppose I position myself as a sociolinguist and I'm+

<$1> Yes.

<$2> +looking at that.

<$1> Yeah.

<$2> How+

<$1> Okay.

<$2> +how language functions.

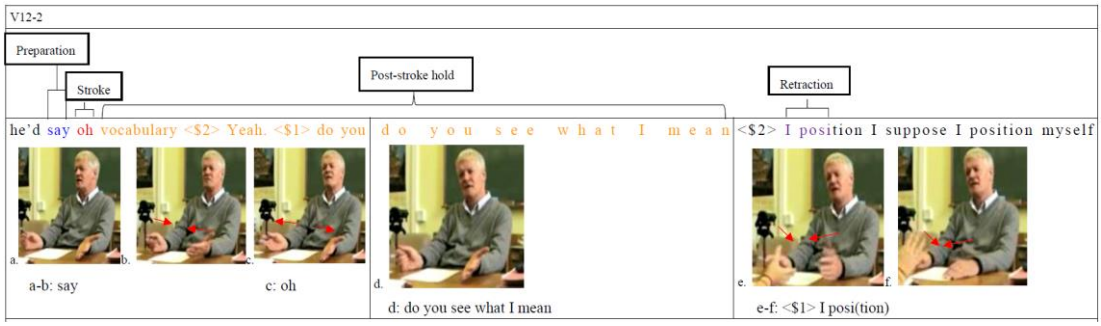


Figure 4.26: A TFS checking understanding co-occurring with a post-stroke hold phase.

In terms of the post-stroke hold phase co-occurring with the TFS (see Figure 4.26 above), both hands rest on the desk at the onset of the preparation phase, during which time he quickly and gently rotates both palms inwards till they are almost facing each other in a vertical position (a-b). As he is saying "oh", he rapidly opens both hands (c), which is the stroke phase followed by a long post-stroke hold phase with both hands held in that form and position (d). The post-stroke hold phase starts slightly before the TFS from "vocabulary <$1>

do you", throughout the TFS "do you see what I mean" (d), and finishes together with the TFS.

Table 4.15 below shows the temporal relationship between all the gesture phases co-occurring with the TFSs checking understanding in the extended analysis. The type of gesture phases is specified in the third column, and the target phases are colour-coded and in bold (the strokes are in red and the post-stroke hold in orange).

Table 4.15: Temporal relationship between the TFSs checking understanding and gestures in other speakers.

Items	Speakers	Videos/gesture phases	Speech-gesture coordination
2	Supervisor 3	V3-1 stroke	set up for your method (0.2) <$2> Okay. **(0.2) Do you see what I** mean<$2> Yeah. <$1> (0.3) But if it's if it's utterly
5	Supervisor 3	V3-4 stroke	patient discourse on the other (1.0) **(0.2) do you see** what I mean wh=why (0.4) (0.2) <$2>: why <$2> why not
6	Supervisor 3	V3-5 stroke	(0.8) <$2> Laughs. <$1> I me**an you see what I mean**/<$2> Yeah. <$1> **You y= y=** you make a virtue of the
10	Supervisor 4	V12-1 stroke	than interpreting the poems or= <$2> Yeah <$1> **D= do you** see what I mean and I think th= that you you can get a thesis
11	Supervisor 4	V12-2 post-stroke hold	he'd say oh **vocabulary <$2> Yeah. <$1> do you do you see what I mean**<$2> I position I suppose I position myself

The table reveals that most stroke phases are in line with the greater part of the TFSs, especially V3-1 and V3-5. The post-hold phase that in line with the TFS (V12-2) starts slightly before the TFS and lasts until the end of it. Consistent with the results from the case study, the gesture phases and the TFSs checking understanding share a very close temporal relationship.

4.4.2.2. TFSs as discourse markers and the co-occurring gestures

The results of the case study show that those TFSs as discourse markers tend to occur without any gestures. When they do co-occur with some gestures, the types tend to vary and no pattern of gesture use emerges. The results from case study is confirmed in other speakers (Table 4.16). They co-occur with different types of gesture phases except for the post-stroke hold phase, which was found in the case study to be closely associated with the function of checking understanding in interaction. Again, this may be the result of the extremely limited number of instances. Moreover, the only TFS without any accompanying gesture was, again, found to be related to the speech function of discourse marker, which also confirmed the results of the case study.

Table 4.16: Gesture types co-occur with TFSs as discourse markers in other speakers.

Items	Speakers	Videos	Gesture types	Speech function
3	Supervisor 3	V3-2	Palm Up Open Hand stroke	Discourse marker
4	Supervisor 3	V3-3	Bounded-space stroke	Discourse marker
7	Supervisor 4	V4-1	Beat stroke	Discourse marker
9	Supervisor 3	V11-2	Palm Up Open Hand stroke	Discourse marker
12	Supervisor 4	V12-3	No gesture	Discourse marker

Additionally, there are two types of gesture phases which were not identified in the case study, i.e. two PUOH stroke/gesture and one bounded-space stroke/gesture, all of which came from Supervisor 3. As there are also two PUOH gestures from the same speaker co-occurring with the TFSs checking understanding (refer back to 4.4.2.1.), no firm conclusions are made here in view of the limited number of instances. Further studies are required to examine the relationship between the functions of TFSs and the PUOH gesture. A more detailed discussion concerning this will be provided later in the discussion section.

One example of the bounded-space gesture (V4-1) will be provided and demonstrated in detail here (refer to File 4.3 in the CD-ROM data disk for other instances).

Speech episode (V3-3):

<$1> Well again it's it's it's the selection of data thing y= you've gotta chop the data down using one set of criteria rather than think I'm gonna chop the data down by going and finding the good examples.

<$2><$E> head nods<$E> <$E> laughs <\$E>+.

<$1> +(0.8) ***you know what I mean*** *you've got to do that you've got to do it first sort of+*

<$2> It's quite hard to <$H> know <\$H>+

<$1> +I'm just gonna look at all the ones dated X or you know only where people are talking about w= this particular subject. In other words have a have a criteria+

<$2> Some kind of principle.

<$1> +for for slimming it down that isn't the criteria for the thing you're going to look for.

<$2> Yeah.

In the speech episode provided above, the supervisor is trying to explain the different approaches for narrowing down the size of the data for analysis. The student shows her understanding by nodding her head instantly after this. This TFS occurs immediately after the student's confirmation of understanding at the onset of some further explanation by the supervisor. There is no further response from the student immediately after the TFS. The discourse context of the TFS then suggests the role of a discourse marker at the beginning of an utterance rather than as checking understanding.

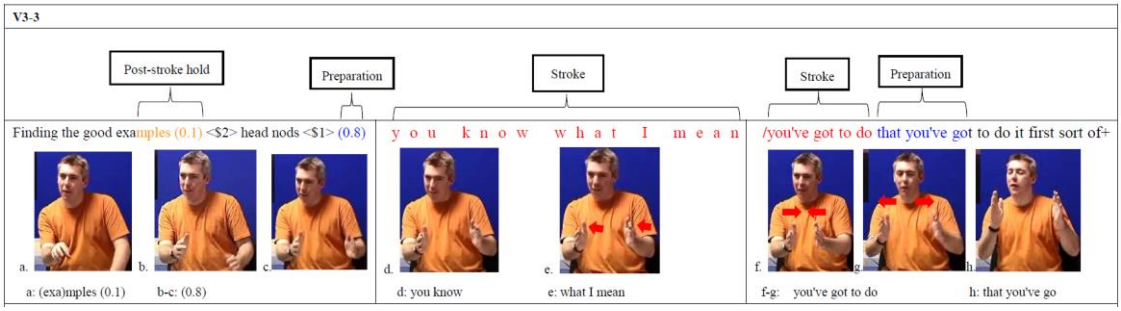


Figure 4.27: A TFS as a discourse marker co-occurring with a bounded-space gesture.

A detailed analysis of the bounded-space gesture that co-occurs with this TFS is presented in the figure above (Figure 4.27). The right hand (a) is held with the palm facing downwards and the fingers bent in front of his chest from the end of the utterance "examples" and a 0.1-second pause. During the student's head nods and a 0.8-second pause (b-c), he opens the right hand and lifts the left hand from the rest position, in order to form a bounded-space with the two hands facing each other. This is the preparation phase preceding the following stroke, which coincides with "you know what I mean". During the stroke, both hands are continuously and simultaneously moved upward and rightwards to the very centre in front of the supervisor's upper body (d-e). When the following speech "you've got to do" begins, he performs another gesture by

rapidly moving both hands horizontally towards (f) and then away from each other (g).

Examining all the TFSs as discourse markers and the co-occurring gestures from other speakers (Table 4.17), it shows that there is a close temporal coordination between speech and gesture as the gesture mostly starts at the beginning of the TFSs and ends as they finish. Except for the final one in the table, the other four instances are all stroke phases, which are coded in red and in bold in the texts.

Table 4.17: Temporal relationship between TFSs as discourse markers and gestures in other speakers.

Items	Speakers	Videos/Gesture phases	Speech-gesture coordination
3	Supervisor 3	V3-2 stroke	then you're alright. <$2> right (0.2) **do you see what I mean** (0.2) cos cos y= you're not there's no interfering
4	Supervisor 3	V3-3 stroke	the good examples (0.1) <$2> head nods <$1> **you know what I mean**/you've got to do that you've got to do it first sort
7	Supervisor 4	V4-1 stroke	a study that lasts and lasts and lasts do you know yo**u know what I** mean there's quite a lot of PhD thesis that people write
9	Supervisor 3	V11-2 stroke	That's a contradiction isn't it but y= (0.09) **you know what I mean** It's it's so not= noticeably well even when it's.
12	Supervisor 4	V12-3 no gesture	<$1> +your final chapter is going to be the most interesting one <$O76> do you see what I mean<\$O76> <$E> Laughs <\$E>

4.4.2.3. TFS as claiming mutual understanding and the co-occurring gestures

In the 12 instances in the extended analysis, there is a new speech function which has not been identified in the case study, that of claiming mutual understanding. Two speakers, Supervisors 2 and 3, have used the TFS for this purpose, both of which will be demonstrated in detail in this sub-section. Both TFSs of this function are closely related to the termination of self-repair of some trivial information on the part of the speakers. As is shown in the speech episode of the first example below, the supervisor is using the metaphor of graffiti to describe the online platform under discussion as a public place where people can post all kinds of comments. The long pause before "verbal kind of graffiti" in line 6 suggests his uncertainty of the usage, or the part of speech, of "graffiti". Although the student replies yes, he immediately corrects himself regarding this and says "well I suppose graffiti is verbal. You know what I mean". The student then confirms his understanding once again. It is clear that the student understands what he meant by graffiti, and the current use of graffiti is of little significance with regards to the main point that he is trying to demonstrate here. Such a discourse context, therefore, suggests the function of the TFS as a way of claiming mutual understanding on the part of the speaker after his attempt to self-repair his own words rather than checking the understanding of the interlocutor.

Speech episode (V2-1):

<$1> Erm <$E> pause <\$E> yeah so <$E> pause <\$E> as a forum or the venue for <$E> pause <\$E> putting that kind of rubbish+

<$2> Mm.

<$1> +or y= or you know just playful stuff into a zone.

<$2> Mm.

<$1> Almost like <$E> pause <\$E> verbal kind of graffiti+

<$2> Yeah yes.

<$1> +you know <$E> pause <\$E> well I suppose graffiti is verbal.

You know what I mean.

<$2> Yeah.

<$1> Erm <$E> pause <\$E> kind of a place where they can write something+

<$2> Mm. Mm.

<$1> +kind of publicly or <$E> pause <\$E> you know+

<$2> Mm.

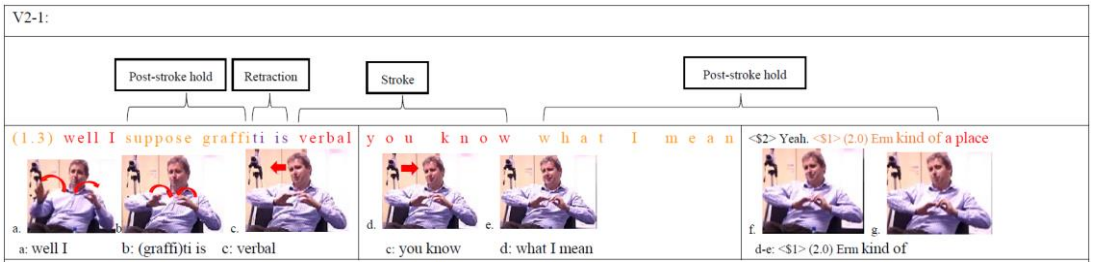


Figure 4.28: A TFS claiming mutual understanding co-occurring with a fling gesture.

The gesture (Figure 4.28) co-occurring with this TFS is performed by the head as the third hand when both hands are unavailable. During the 1.3-second pause before the stroke co-occurring with “well”, both hands have been held steady similar to the form in picture b, i.e. both hands open and the palms facing downwards and in line with each other. As he is saying “well I”, he abruptly opens both hands outwards (a) till the palms are held vertically and in line with the shoulder (a). This gesture form is maintained during “suppose graffi(ti)”, after which he again retracts the hands back to the original position during “(graffi)ti is”(b). Immediately before the TFS, when he says “verbal”, he suddenly throws his head to the right (c) and then bounces back during

"you know" (d), which is the stroke of this head fling gesture. The head and the hands are then held in that position for the rest of the TFS "what I mean" (e) and the following utterance (f-g).

A similar instance has been identified in V11-1 performed by Supervisor 3. As is shown in the speech transcript below, the supervisor is having trouble recalling the name of a museum and attempting to correct this. The TFS occurs at the end of this self-correcting effort in the same turn, which indicates no intention of checking understanding but conveys his confidence in mutual understanding. The student seems to have the same interpretation of the intention of the TFS and does not reply until the supervisor finishes his turn.

Speech episode (V11-1)

> *<$1> It's really odd cos Ru= Rufus Wainwright's sort of <$G2> say a well known gay man <$E> <$2> Laughs <\$E> or bisexual man but it's it's a prominent fact of his personality sort of <$O43> public persona <\$O43>. Erm and he sings this song erm about a school girl visiting the New York Metropolitan Museum of Metropolitan Art no Metropolitan Museum of Modern Art* ***you know what I mean*** *erm and seeing a Turner a painting and it's their art teacher who takes them to to th=see the Turner exhibition and she's in love with the art teacher. Erm <$=>and then it's a sort <\$=> told as a flashback in first person from her point of view with all inside her head and then later on as as the wife of a successful business executive and sort of a leading what she plainly sees as a boring life. Erm the last sort of line is that she never saw the art teacher again but in her house she has a Turner cos she was able to buy one+*

The fling gesture in line with this TFS is performed by the hand rather than the head. As is presented in the figure below (Figure 4.29), the hands are at rest (a)

before the preparation phase, where the speaker raises the hand as extending the fingers and increasing the tension of the hand (b-c) until it is held vertically and facing his body. The moment he starts to say "you know what I mean", he rapidly and abruptly throws the hand outwards and rightwards (d-e). The movement is so fast and intense that the hand bounces back slightly towards the body (f). Directly after the TFS, he releases the strength in the hand and folds the fingers (g-i) as the hand retracts.

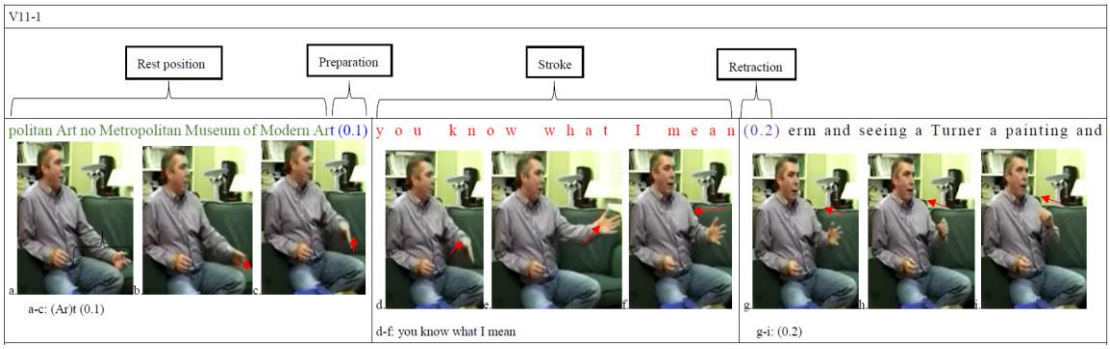


Figure 4.29: A TFS as claiming mutual understanding co-occurring with a fling gesture.

Table 4.18: Gestures co-occurring with TFSs claiming mutual understanding with other speakers

Items	Speakers	Videos	Gesture types	Speech function
1	Supervisor 2	V2-1	Head fling	Claim mutual understanding
8	Supervisor 3	V11-1	Fling stroke	Claim mutual understanding

The finding of this new function of the TFS (see Table 4.18 above) is of great significance as it not only points towards a more comprehensive and nuanced description of the meaning/function of this group of TFSs, but also further confirms the functional coordination between gesture and speech.

4.5. Discussion

This section will first discuss issues concerning the alignment between gestures and the TFSs, and the implications of using gestures as a criterion for identifying formulaic sequences (4.5.1.). It will then demonstrate the emerging functional coordination between different types of gestures (i.e. beat/beat-like gestures, holding, Palm Up Open Hand gestures, fling gestures) and their functional associations based on the findings from the case study and the extended analysis (4.5.2.). Based on the emerging functional associations between the gesture types and the TFSs, a speech-gesture profile for "(do) you know/see what I mean" can be established, in support of understanding the continuum of meaning in the multimodal TFSs (4.5.3.). The final section will comprise a discussion of the five discourse functions of the gestures co-occurring with the TFSs, thus producing one of the most comprehensive frameworks for describing gesture functions in the field of gesture studies (4.5.4.).

4.5.1. Temporal relationship between TFS and gesture

The results of the overall temporal relationship between the TFSs and gestures show that, when this TFS is used by the speaker, there is a good chance that it will be accompanied by some type of hand-and-arm movements. In the case study, about two thirds (49 out of 64) are accompanied by gesture phases (e.g.: preparation, stroke, post-stroke hold, retraction, etc.), and, among the 49 gesture phases accompanying the TFSs, nearly half (22 instances) are independent strokes/gestures. The same conclusion can be drawn from the extended analysis of other speakers, where 11 out of 12 instances are accompanied by gestures and the greatest proportion of them (10 out of 11) co-occur with independent strokes/gestures.

What is also worth noticing is that, when TFSs are used as a means of checking understanding, the incidence of them being accompanied by a gesture is much higher than those used as discourse markers. This might be the result of their different degree of importance in the interaction, i.e. TFSs checking understanding tend to have a greater influence on reaching the mutual understanding between the speakers and enhancing the inter-subjectivity between the interlocutors. Such a result points to the possibility that, for the same multifunctional expressions, when they have more impact on communication, the speaker is more likely to use both speech and gesture to foreground the meaning. This is in some ways consistent with previous findings that gestures (e.g. beat gestures) tend to co-occur with noteworthy information in speech (McNeill, 1992, 2005).

In terms of the specific gesture phases co-occurring with the TFSs, in the case study, 7 categories have been identified, i.e. independent gesture/stroke, post-stroke hold phase, embedded in stroke phase, preparation phase, retraction phase, post-stroke hold phase + preparation phase and post-stroke hold phase + retraction phase. In the extended analysis, probably because of the limited number of instances, only two gesture phases have been identified co-occurring with the TFSs, among which 10 are independent strokes/gestures with one instance of the post-stroke hold phase. Except for the TFSs co-occurring with embedded in stroke phases, other gesture phases tend to align with the beginning and end of the TFSs. Moreover, a more congruent speech-gesture temporal coordination has been found between those TFSs synchronised with independent strokes/gestures.

The fact that the TFSs can constitute a close temporal coordination with gesture phases other than stroke phases is valuable as previous research usually only focuses on the stroke phase, the phase in which the meaning of the gesture lies. However, the current analysis indicates that, depending on the function of formulaic sequences, they could also co-occur with other gesture

phases and form a congruent temporal relationship. This then furthers our knowledge of the temporal alignment between speech and gesture in that alignment can not only be found out between strokes and their associated speech, but also in other gesture phases.

The tendency of the TFSs co-occurring with gestures, especially independent strokes, suggests the multimodal nature of the processing of multiword expressions in spoken language. Furthermore, the temporal congruency between TFSs and different types of gesture phases confirms the cognitive hypothesis regarding speech, gesture and mind (Kendon, 2004; McNeill, 1992, 2005). That is, gesture and speech share the same psychological processing system and, depending on the discourse context, the speakers have the ability to design the best combination of the speech-gesture unit to deliver the intended meaning in the flow of online communication.

In addition, the speech-gesture alignment in the TFSs also indicates the function of gestures as a means of segmenting and determining formulaic sequences. Different criteria have been put forward in the past for the ways in which the boundaries of the units of formulaic sequences could be decided (refer back to Chapter 2, 2.3 for details). To date, there is no consensus regarding the best approach to do so and using mixed criteria is recommended by some researchers (Read & Nation, 2004; Wray, 2002, 2008). Wray (2008) has identified several criteria including corpus-based frequency lists, phonology, form idiosyncrasies, spelling, intuition, published lists, etc. However, little research has been conducted in terms of whether gesture could be a means of determining, or confirming, boundaries of formulaic sequences. Taking the TFS "(do) you know/see what I mean" as an example, their co-occurrences with the beginning and end of gesture phases seem to reveal or confirm their status as single multiword units that are likely prestored and processed as a whole.

4.5.2. Functional coordination between TFS and gesture

This section will address the issue of the emerging functional associations between gestures and the co-occurring TFSs. As textual analysis indicates the different frequencies of meaning variations of the same expression (Sinclair, 1991), we have little knowledge of the gesture patterning in relation to those variations. In other words, whether and what gestures tend to co-occur with each variation of meaning. The multimodal corpus-based approach adopted in the current study could be a robust method for exploring this matter, and the following sub-sections will report the three types of speech-gesture functional coordination found in the first study. They include:

a) TFSs checking understanding with beat/beat-like strokes, beat/beat-like strokes with long post-stroke holds and post-stroke hold phases;
b) TFS as checking understanding and as a discourse marker with Palm Up Open Hand gesture;
c) TFS claiming mutual understanding with the fling gestures.

The latter two gestures occur in other speakers in the extended study rather than the case study. Although they only occurred a few times, they were chosen here as two frequently occurring gestures in the gesture literature and in real life. It is hoped that the current discussion can shed some new light on their meanings and provide directions for further research.

4.5.2.1. TFSs checking understanding with beat/beat-like strokes, beat/beat-like strokes with long post-stroke holds and post-stroke hold phases

Compared with TFSs of other functions, those TFSs checking understanding not only co-occur with the most salient and significant gesture types, but also formulate a much closer temporal and functional coordination with gestures.

The case study reveals that TFSs checking understanding tend to co-occur with beat/beat-like gestures and post-stroke holds. Both types of gesture phases have also been identified in the five instances of TFSs checking understanding used by other speakers.

As a well-documented gesture, beats are metaphoric gestures (McNeill, 1992) that usually synchronise with the tonic centre of an intonation unit and mark-up the relatively more salient, noteworthy component(s) in speech (Gullberg, 2006; McNeill, 2005). Although beat gestures do not have any referential meanings, they are pragmatically significant in terms of adding extra emphasis to the important information in discourse (Gullberg, 1998). Such a pragmatic role played by the beat gestures are well coordinated with the meaning of checking understanding of the TFSs.

For those TFSs checking understanding coordinated with long post-stroke hold phases, it is indicated from previous studies that holding serves the purpose of sustaining the communicative intention of the speaker and eliciting responses from the listener. This function is also largely equivalent to the meaning of checking understanding in the TFSs. For instance, Kendon (1995) found out that after performing a conventionalised *purse hand gesture* (which has the meaning of questioning) throughout the entire question, the hand remains unmoved, extending the intention of questioning and seeking for an answer from the interlocutor. In a similar vein, Brookes' (2005) examination of the *Money gesture* demonstrates that, as the speaker was using this gesture while asking a question about money, the gesture was held after the enquiry and did not withdraw until an answer was provided.

To further confirm the hypothesis that the function of post-stroke hold phases as confirming mutual understanding and/or eliciting responses, further extra piloting analysis has been conducted using the same video, i.e. Video 6. All the gestures in a randomly chosen five-minute episode (from 18:00 to 23:00) were segmented and all the post-stroke hold phases were analysed. In

total, there are 54 post-stroke hold phases identified with a length longer than 0.1 second. The threshold of 0.1 was decided based on the observation that the normal gap between two moves is about 0.06 to 0.08 seconds, and for a holding to be a post-stroke hold phase, it needs to last longer. It appears that most post-stroke holds do last at least 0.1 second.

Of those 54 instances, half co-occur in a context where the supervisor triggers some response from the student, which is temporally synchronised with the post-stroke hold phases. What differentiates those post-stroke holds from the other half post-stroke holds is, for one thing, their positions tend to be adjacent to the end of an utterance or are accompanying the speech as it is approaching the end; for another, their average length (0.962 second) is nearly twice as long (0.507).

These results of the position and the duration of the post-stroke holds as seeking for responses well justify why they tend to co-occur with TFSs checking understanding at the end of utterances rather than TFSs as discourse markers at the beginning. They also account for the longer post-stroke holds in line with the TFSs checking understanding in the case study and the shorter ones with the two TFSs as discourse markers in other speakers. That is, compared to the shorter post-stroke holds, the longer ones tend to be the more likely indicate the meaning of seeking a response.

Moreover, in terms of the speech co-occurring with the 54 post-stroke holds, only two instances literally express the intention of checking understanding, with one instance of “do you know/see what I mean” and one “does that make sense”. This means that holding might be playing a more important role than speech regarding triggering responses from the interlocutors.

Although a much larger data-set is needed to verify the results, all the aforementioned results suggest that, regardless of the speech elements, post-stroke holds, especially those long ones positioned close to the end of an

utterance, are closely related to the context of triggering responses and/or seeking confirmation from the other party. The small pilot study then further confirms the functional coordination between post-stroke hold phases and TFSs checking understanding.

4.5.2.2. TFS as checking understanding and as discourse markers with the Palm Up Open Hand gesture

What worth noting is the recurrence of a relatively conventionalised gesture, the Palm Up Open Hand gesture (McNeill, 2005), though it was only used by one speaker (Supervisor 3). Previous research on the same gesture shows that it is a metaphoric gesture usually coordinated with the meaning of offering or presenting abstract ideas as well as the act of questioning (Calbris, 2011; Kendon, 2004; McNeill, 2005). This is more or less confirmed by the current study, although the results are not as straightforward. There are four instances of this gesture, two of which co-occur with TFSs checking understanding and two with those as a discourse marker. Therefore, it seems unclear which one of the two functions the PUOH gesture is more closely associated with. Due to the extremely limited number of the instances and the inexplicit functional coordination, no tentative conclusion could be made with regards to which function of the TFSs the PUOH gesture tends to be associated. Also, as the Palm Up Open Hand gesture only emerges in the data of one speaker, there is a question regarding whether this gesture is solely an individual preference of that particular speaker (Supervisor 3) or a type of gesture co-occurring with the TFS across different speakers. These two questions need to be explored further.

4.5.2.3. TFS claiming mutual understanding with a fling gesture

The fling gesture was not identified in the case study, but in two speakers in the extended dataset. The gesture was investigated as a recurrent gesture by previous researchers, namely, the throwing away gesture (Müller, Bressem, et al., 2013). It is claimed by the German scholars as embodying the meaning of abandoning, which seems to be embedded in the rapid outward motion in the stroke. The meaning of this gesture is reflected in the speech function of TFSs as a means of terminating self-repair and claiming mutual understanding, which further confirms the functional synchrony between the fling gesture and this meaning. However, it has to be clarified that the fling movement in supervisor 2 was performed by the head rather than the hand (refer to 4.4.2.3. for details). Although head movements are excluded from the analysis, in this particular instance, it represents a special use of the head as the third hand (McNeill, 2005) when both hands are not available (they held in parallel in the air).

Despite the considerably minimal number of instances, the finding of this new function of TFS is of great importance. It not only provides a more nuanced description of the meaning/function of the TFSs, but also further confirms that different speech functions tend to co-occur with distinctive types of gestures and gesture could play a role in distinguishing the meaning variations of multiword expressions. However, this new functional dimension was only found in two supervisors and in a particular form, i.e. “you know what I mean”. It is possible that this new dimension may only apply to “you know/see what I mean” without the question marker *do*, rather than “do you know/see what I mean”, in which the meaning of checking understanding is already indicated. Confirming this will require separate research on “do you know/see what I mean” and “you know/see what I mean”, drawing on a much larger spoken corpus.

4.5.3. Meaning on a continuum

Adopting a multimodal corpus-based approach, the detailed and systematic assessment of the temporal and functional relationship between the TFSs and gesture reveals a much more complex and dynamic picture of meaning construction compared to the findings of monomodal, text-focused studies. It indicates that the units of meaning composed by speech and gesture in the current TFSs are more appropriate to be viewed as on a continuum with the functions of checking understanding and discourse marker on each side. As is claimed by previous multimodal corpus-based research on head nods and response tokens (Adolphs & Carter, 2013; Knight, 2011), the multimodal units of meaning configured by different modes such as gesture and speech are more likely to be positioned on a spectrum. A function of the speech response token can be strengthened/weakened by the accompaniment of a strong/weak form of head nods, hence changing the function of the entire multimodal response tokens.

According to the dominant gesture types revealed in each TFS function from the case study, the figure below (Figure 4.30) represents a tentative speech-gesture profile for the current TFSs. This profile attempts to demonstrate ***the general tendency*** of speech-gesture coordination mostly in accordance with the findings from the case study. Admittedly, there are always exceptions in language use and the factors that lead to uncommon uses could be valuable; however, revealing the overall tendency could be more useful for wider applications such as dictionary and textbook editing (Sinclair, 1991).

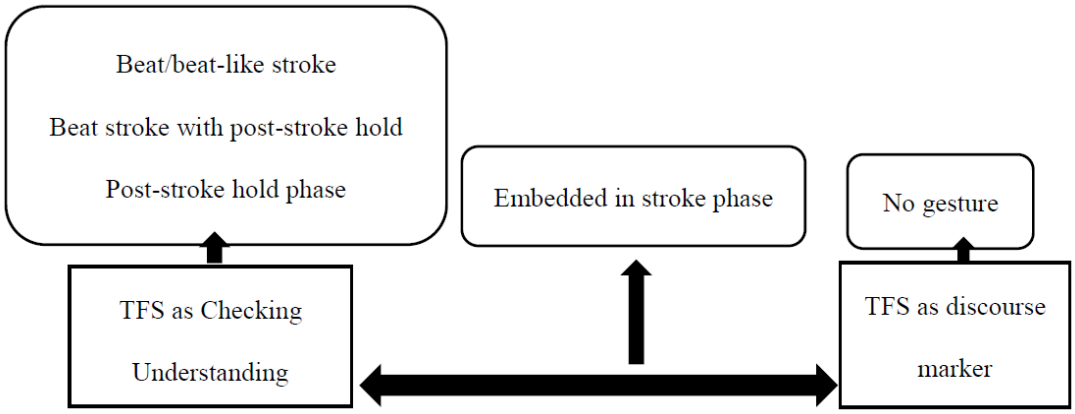


Figure 4.30: Speech-gesture functional profile for multimodal TFSs.

Some new implications from the extended analysis do exist; however, the number of instances is too small to affect the results of the case study. For instance, although the emergence of new gestures (the Palm Up Open Hand and fling gesture) and the new function of TFS as claiming mutual understanding are valuable and worthy of further exploration, the number is simply too small to be included in the profile.

The profile shows that, on the left side of the continuum, TFSs as checking understanding tend to be accompanied by beat/beat-like strokes, beat-like strokes with post-stroke holds, and post-stroke holds, all of which represent the most salient and meaningful gesture types. On the other side, those TFSs as discourse markers tend to occur without any gesture. Situated in the middle of the continuum are those TFSs that tend to perform both functions, with the majority of them co-occurring with motions embedded in stroke phases. In general, as the salience and significance in the speech functions and gesture phases decrease from the left to the right, so do the entire multimodal TFSs.

The functional relationship between the TFSs and gestures identified here has important implications. As is suggested by researchers on the functional coordination between head nods and response tokens (Knight, 2011), the associations between functional variations and gesture types indicate the

potential role of gesture in distinguishing speech functions. Similarly, the emerging patterns of TFS functions and gesture types also suggest gesture's role in disambiguating the functions of formulaic sequences. As corpus linguists usually rely on textual contexts to explore the meaning variations of formulaic language, gesture can be an alternative for them to do so.

Nevertheless, despite the dominant tendency in gesture-speech association for most TFSs, complications could be involved in positioning the function of those TFSs associated with those overlapping gesture types. For instance, the function of a TFS as checking understanding with no gesture moves slightly towards the other side of the continuum. On the contrary, for a TFS as a discourse marker that co-occurs with a beat-like gesture, its function swings, to a certain extent, to the pole of checking understanding. Similarly, the function of a TFS combining both functions slides to the left side when synchronised with a more salient gesture type such as a beat/beat-like strokes, whereas to the right while co-occurring with any weaker gesture phases such as a retraction/preparation phase, or even no gesture.

Due to the limited scope of the current case study, further investigations are required to verify and refine the functional profile for the current TFSs. However, the multimodal corpus approach proposed here does present great worth for unfolding the dynamic and complex nature of meaning in the multimodal units. Studies of a similar nature will shed significant light on the interplay of distinctive semiotic modes such as speech, gesture, facial expressions, etc., but also on wider applications such as language teaching and human-computer interaction (issues of applications will be addressed in Chapter 7).

4.5.4. The discourse functions of gestures

One of the research questions of the thesis is to explore the discourse functions of gestures in accordance with the co-occurring speech. Previous researchers (Calbris, 2011; Kendon, 2004; McNeill, 2005) have discussed the discourse functions of gestures and the most frequently mentioned are that they contribute to utterance meaning, segmenting utterances and underlining important information (refer back to Chapter 2, 2.6 for details of the functions of gestures). Nevertheless, the current thesis intends to explore other functions that have not been emphasised or mentioned in the existing literature.

Relating to the gestures co-occurring with the TFSs, all the functions that have been previously identified also emerge in this study, but also others. Hence, based on the current datasets, five discourse functions of the gestures can be proposed. They include 1) contributing to referential meaning, 2) segmenting utterances, 3) underlining noteworthy information, 4) maintaining coherence and 5) enhancing inter-subjectivity. Before illustrating further, it is worth mentioning that, as is claimed by Calbris (2011), most gestures tend to be multifunctional in nature as they can play at least two discourse functions simultaneously, i.e. discourse segmentation and one of the others. The current study takes a similar stance and admits that each gesture can play multiple functions simultaneously according to the discourse context.

Numerous researchers have demonstrated the fact that gestures make a referential contribution to the utterance meaning (Calbris, 2011; Kendon, 2004; McNeill, 2005). Kendon (2004) specifically discusses the six ways that gestures contribute to utterance meanings referentially, equivalent or complementary to the aligned speech components, as the hand depicts objects, enacts actions and makes references (deictics). Calbris' entire book (2011) is dedicated to a thorough discussion of the referential gestures, demonstrating the variations of form-meaning associations in gestures across discourse

contexts. Similar to Kendon, Calbris' definition of referential gestures seems to involve all the gesture types that contribute to the semantic aspect of utterance, including those gestures that McNeill (1992, 2005) designates as iconic (referring to concrete objects), metaphoric (referring to abstract objects/space in speech) and deictic/pointing gestures.

The gestures co-occurring with the TFSs also contribute to the meanings of utterances. Obvious examples would be those beat/beat-like gestures co-occurring with TFSs in checking understanding that add the meaning of emphasis to the utterances. Also, holdings that co-occur with TFSs as checking understanding that add the meaning of presenting/waiting for responses. Whereas the referential meanings of those gestures are valuable in delivering the intention of the speaker to reassure mutual understanding between the two parties, they seem particularly crucial when they accompany those TFSs acting as discourse markers. Instances of such include a few TFSs as discourse markers co-occurring with two beat-like gestures, two PUOH gestures and one holding. In those instances, the referential meanings of the gestures complement the spoken TFSs, adding more importance to the whole multimodal units.

The second function, demarcating utterances, is also revealed in the current study. As gestures generally are aligned with the speech components they are associated with, they are acknowledged to have the function of demarcating utterances into units of different sizes (Calbris, 2011; Kendon, 2004; McNeill, 2005). In relation to the TFSs, it seems apparent that the start and end points of speech and gesture tend to be synchronised. In this sense, the gestures segment the TFSs they co-occur with from the rest of speech.

As gestures provide listeners with visual access to meaning in communication, it could be argued that all the gestures performed by speakers, to a certain extent, serve the function of emphasising key information. This function applies to those iconic (e.g. depicting actual objects and action) and

pointing gestures (referring to an object) that contribute to the referential meaning of utterances, but also beat and metaphoric gestures that do not contain any meaning. Whereas the first function, contributing to referential meaning, emphasises the gesture's function of adding information, the third function refers to the gesture's role in marking-up utterance structure in the sense that they visually foreground the noteworthy components of speech. In gesture studies, the gestures' discourse function of emphasising certain parts of utterances has been discussed by many researchers in particular with relation to beat gestures (Kendon, 2004; McNeill, 2005). The beat gestures are usually performed in accordance with the stressed part(s) in speech, therefore, marking out crucial elements as the conversation unfolds. In addition to beat gestures, this role has also been found in some conventionalised gestures (Brookes, 2011; Kendon, 1992, 1995). For instance, in Kendon's (1995) research on the emblems in Southern Italy, he found that the Finger Bunch gesture marked out the topic of discourse as opposed to a comment, which he terms as a discourse structure marker.

The gesture's function of underlining noteworthy information is well reflected in those that co-occur with the TFSs, especially the beat gestures. However, the same effect can also be applied to other gestures that co-occur with the TFSs. That is, when the supervisor opens the hand (i.e. UPOH gesture), holds the hand (holding) or throwing away the hand (fling gesture) while saying "(do) you know/see what I mean", the TFSs are also underlined.

Gestures can co-occur not only with fluent speech, but also with pauses and non-fluent features. In those situations, previous research has barely discussed the potential role of such gestures in maintaining discourse coherence, but has mostly considered them as helping to retrieve lexical contents (Krauss, Chen, & Gottesman, 2000). For the gestures that co-occur with the TFSs, there are many instances where the speaker maintains the position of the hand at the end of an utterance while or after saying "(do) you

know/see what I mean", awaiting the response of the listener. Due to the particular transitional positions of the TFSs and the gestures, it seems that the gestures are playing the role of maintaining the flow of conversation between the two parties.

The one potential function of gestures that has long been ignored is its interpersonal role in discourse; however, this function is very much foregrounded in this study. As the gestures co-occurring with the TFSs serve the important function of checking the mutual ground and involve the speakers, they are significant in terms of maintaining a good relationship between interlocutors. This finding is consistent with the results of research on interactive gestures (Bavelas et al., 1992), which emphasises that some gestures are more closely associated with conversations than monologues as they tend to refer to the listeners rather than the contents. More discussion regarding the interactive nature of gestures will be provided in Chapter 6, 6.4.3.

4.6. Summary

This chapter demonstrated the approaches and procedures of constructing the speech-gesture profile of the TFSs, "(do) you know/see what I mean", as a means of exploring the multimodal units of meaning in speech and gesture. The overall processes included, first, coding the speech functions of the TFSs based on the discourse context and categorising the gesture types according to the gesture phases co-occurring with the TFSs, and then analysing the dominant gesture types in line with each function. During the processes, three research questions have been addressed including a) the temporal relationship, b) the functional coordination between the TFSs and the co-verbal gestures, and c) the discourse functions of the gestures in relation to speech. Supervisor

6 was used as the basis for the case study, with other speakers as the test bed for the validity of the major findings.

The results from both the case study and the extended analysis suggest a high likelihood of the TFSs accompanied by strokes, especially for TFSs checking understanding. As the strokes are usually more salient in form than other gesture phases (e.g. preparation, retraction), this means that TFSs checking understanding tend to be accompanied by more salient hand movements than the other two speech functions. The overall results also indicated a close temporal relationship between the TFSs and the different types of co-verbal gesture phases, with the highest congruency found in those TFSs in line with strokes. The findings, to some extent, suggest the integral cognitive relationship between speech and gesture, and the multimodal, rather than monomodal, processing of multiword expressions. They also seem to point towards the possibility of using gestures as a criterion for segmenting formulaic sequences. However, due to the limited number of instances involved in the current study, these results will need to be confirmed by further investigations drawing on much larger datasets.

In terms of the speech-gesture functional associations, three types of such relationships were identified in the case study and were mostly confirmed by the extended analysis, including a). TFSs acting as checking understanding with beat/beat-like strokes, beat/beat-like strokes with long post-stroke holds and post-stroke hold phases; b). TFSs as a combination of checking understanding and discourse marker with embedded in stroke phases; c). TFSs as discourse markers with no gesture. Based on those emerging associations, a speech-gesture profile for the TFSs was established. This profile took into account the influence of gestures on the units of meaning, allowing for a dynamic rather than static description of the speech-gesture relationship. It showed the meaning variations of the multimodal unit on a continuum rather than as discrete entries. In addition, the functional coordination between the

TFSs and certain gesture types indicates the possible role of gesture in disambiguating the functions of formulaic sequences.

Regarding the discourse functions of the gestures, this study drew on the dataset and proposed a comprehensive framework containing five categories, including 1) contributing to referential meaning, 2) segmenting utterances, 3) underlining noteworthy information, 4) maintaining coherence and 5) enhancing inter-subjectivity.

However, the instances in the extended analysis raised some new questions that require further investigations. These include whether the association between the two fling gestures and the new functional dimension of claiming mutual understanding will also emerge in other speakers, whether the Palm Up Open Hand gestures will be used by speakers other than the supervisor in the current study, whether the PUOH gestures tend to co-occur with TFSs checking understanding or as discourse markers, and how the functions of "you know/see what I mean" differ from "do you know/see what I mean".

The approaches and all the frameworks developed in this chapter can be adapted to future research. However, it has to be emphasised that the scope of the dataset used in this study is rather narrow, hence caution is needed when trying to generalise the findings to a larger population. Similar research on the same TFSs based on a much larger dataset is required to validate the results and refine the frameworks.

Chatper 5. Exploring the Unit of Meaning of the Circular Gesture

5.1. Introduction

The first empirical study of the thesis (Chapter 4) examines the multimodal unit of meaning by investigating gesture patterns in recurrent speech, i.e. formulaic sequences. The current study attempts to take an opposing approach, continuing to examine the speech-gesture relationship and identify the unit of meaning in a recurrent gesture (Ladewig, 2014a, 2014b), i.e. the circular gesture. To achieve this, Sinclair's (2004) classic unit of meaning model, i.e. *collocation, colligation, semantic preference* and *semantic prosody,* is applied to the analysis and description of the emerging patterns of speech co-occurring with them. Meanwhile, the study serves as a demonstration of how traditional textual-based corpus techniques (e.g. concordance, clusters/N-Grams, etc.) can be of great value to multimodal data analysis with particular reference to exploring speech-gesture relationship. For this matter, the free corpus toolkit AntConc will be used as an example.

The rationale behind the selection of this particular gesture, the circular gesture, as a demonstration of the approach explored here, is based on the results of the pilot study. Here it emerged as one of the most frequently occurring gestures that tend to co-occur with certain speech patterns (refer to 5.2.5. for details of identifying those patterns). Some commonly seen gestures, such as holding and beat gestures, tend not to co-occur with certain speech patterns, though they also have a very high frequency in the NMMC. In addition, according to previous research on gesture families (refer to Chapter 2, 2.5.2. for detail), it is likely that the open hand palm up/down

gestures might also co-occur with certain grammatical patterns; however, they do not occur as frequently as expected and such patterns do not emerge in the preliminary observations. Since identifying certain multiword expressions is crucial for establishing the speech-gesture profile (refer to 5.2.6. for detail), the circular gesture is more suitable for the purpose of the current study. Another important reason for choosing this gesture is that, despite its high occurrence, it is usually mentioned in passing in previous research, except for Ladewig's (2011) work. However, her focus of analysis is not on identifying the emerging speech patterns and her analytical framework tends to not capture some of the detail needed for the current study, which will be shown throughout the rest of this chapter.

The current study on circular gestures, to some extent, resembles and builds upon the research on the cyclic gesture family, initially proposed and systematically studied by Ladewig (2011, 2014b) (refer back to Chapter 2, 2.5.3. for a detailed description of the research). Ladewig presumes the meaning of continuity in this gesture, and claims that cyclic gestures add this meaning to the utterance even when the co-occurring verbal parts do not contain such a meaning. On this basis, one major category of her coding scheme was this referential function. The other two functions in her coding scheme were word/concept searching (i.e. those co-occurring proximal to pauses) and encouraging the interlocutor (to continue guessing in the assigned guessing task).

As insightful as Ladewig's research can be, there are some aspects of this particular gesture form, the rotational gesture, that requires further examination. To begin with, as Ladewig took a pragmatic, functional approach, we lack a fine-grained lexical and grammatical categorisation to describe the speech components that co-occur with the cyclic gestures, i.e. a detailed descriptive, linguistic framework that can code the lexical and grammatical patterns that align with the circular movements. As Ladewig

presumes the meaning of continuity in the cyclic gesture, the current study holds that more studies based on larger datasets are required to validate this. It is too early to presume the meaning of this gesture without further solid empirical evidence. Finally, since the examination of the cyclic gesture drew on a German corpus collected from a constrained research context (either the topic or the task was assigned in advance), using a more homogeneous and natural English-speaking corpus such as the NMMC might generate distinctive findings in relation to the uses of the rotational gestures.

Thus, the current research aims to contribute to these aforementioned aspects by conducting corpus-based research drawing on the NMMC. It intends to explore the unit of meaning of the circular gestures by conducting an empirical examination of the speech patterning aligned with them, developing a specific lexical and grammatical coding scheme. The framework will allow a new description of the speech-gesture relationship with particular relevance to circular gestures, and shed new light on the multimodal units of meaning. This study also affords an opportunity to understand the uses of circular gestures among English speakers in a spontaneous communicative context. This chapter will then report on the exploration processes of the following research questions:

1) What is the temporal relationship between the circular gestures and speech?
2) What is the unit of meaning of the circular gestures in accordance with their co-occurring speech?
3) What are the discourse functions served by the circular gestures?

The chapter will begin with detailing the approach adopted by the second study in 5.2., and then analyse and present the findings regarding the speech-gesture temporal relationship and the multimodal unit of meaning framework for the circular gestures (5.3.). The discussion (5.4.) will revolve around the role of textual corpus tools in multimodal research, the evaluation

of the multimodal unit of meaning framework and the discourse functions of the circular gestures. Section 5.5. concludes the whole chapter.

5.2. Approach

The general approach adopted by this study is, first, segmenting the strokes of the circular gestures and identifying the speech co-occurring with them, and then categorising the speech patterns based on Sinclair's (2004) unit of meaning framework, i.e. *collocation*, *colligation*, *semantic preference* and *semantic prosody*. This section will first describe the eight videos used in this study and demonstrate the standards for selecting and annotating the target circular gestures (5.2.1.). It will then illustrate how the forms of the circular gestures relate to the discourse context (5.2.2.). Sinclair's unit of meaning model will be explained in 5.2.3., which will be followed by a detailed description of how corpus tools can be applied to this study (5.2.4.). The lexical and grammatical framework based on the emerging speech patterns co-occurring with the circular gestures will be presented in 5.2.5. The important issue of how the unit of meaning can be adapted to categorising the emerging speech patterns will be explained in the final section (5.2.6.).

5.2.1. Defining and selecting the circular gesture

The research draws on the sub-corpus of the NMMC, the video recordings of supervision meetings, as the basis for analysis. After careful observation, Videos 1, 2, 3, 5, 6, 11, 12 and 13 were selected based on the quality of the videos and the amount of speech and gesture available for investigation.

Table 5.1: An overview of supervision meetings.

Video	Length	Word count
1	35:52:00	9,218
2	1:00:27	11,875
3	41:41:00	9,513
5	1:28:22	16,560
6	47:52:00	11,705
11	43.03	11,726
12	55:58:00	11,963
13	37:58:00	10,194

The total number of running words reaches 92,754 with an average number of 11,594 (see Table 5.1). Most videos have a similar length with a similar word count except for video 5, which contains 16,560 words as it lasts at least 30 minutes longer than the others. Four videos, including 7, 8, 9 and 10, were excluded from analysis for the reason that the supervisors and students were mainly working on a computer and the cameras did not capture the front image of the gestures. Also, this activity greatly reduced the amount of speech and gesture used in interactions, making these four videos inadequate for the investigation of the speech-gesture relationship.

The term circular gesture adopted here is based on the previous research of the cyclic gesture in view of the same core, rotational movement (Ladewig, 2011, 2014b). The current research aims for much more variations of movements for the purpose of capturing all types of speech patterns aligned with them. For cyclic gestures (Ladewig, 2011, 2014b), only those continuous, outward rotational gestures that remain in the same position were selected. No specific reason has been provided regarding why rotational gestures that move in other directions and/or change space were

excluded, except that the cyclic gesture was chosen as a frequently occurring gesture in daily life and in Ladewig and Bressem's corpus (2013). Such a sampling technique is most likely relevant to their aim of exploring recurrent form-meaning associations, i.e. how the different combinations of the core movement (i.e. the cyclic movement) and other gesture parameters (i.e. space and size), can lead to variations in meaning across different discourse contexts. Such an analysis therefore responds well to the questions concerning the degree of conventionalisation of recurrent gestures (Ladewig, 2014a, 2014b) and whether it is the gestural parameters that affect the overall meaning of the entire gesture (Calbris, 2011; McNeill, 1992, 2005). Thus, including other rotational gestures of different movement trajectories (anti-clockwise/outwards) and/or changing space (e.g. rotating forwards) means adding other variables to the analysis and further complicating the research question.



Figure 5.1: Examples of Cyclic Gestures from Ladewig (2014b).

Figure 5.1 above is a selection of examples presented in Ladewig's paper. As is shown in the pictures, all these gestures are executed by one hand with the hand rotating clockwise and outwards. There seem to be no clear instructions on the handedness of the cyclic gestures either, i.e. whether they are required to be performed by one hand or two. However, it can be inferred from her description that the cyclic gestures are all one-handed. The decision of only including those involving one hand is probably similar to the reason why she only included clockwise, outward-moving ones. That is, to keep the gesture forms as consistent as possible to control the analysis of gesture form-meaning mappings.

A further comprehensive discussion concerning circular gestures has been presented by Calbris (2011), though she did not report the exact number of instances in the given categories. Her analysis was similar to Ladewig's research in terms of focusing on the significance, or relevance, of gesture components in relation to the semantic variations of gestures depicting circles. The four elements analysed include the direction of the movement (clockwise or anticlockwise), the body part (e.g. finger, the edge of the hand), the repetition (>half circle, one circle, multiple circles) and symmetry (circles performed by one hand or two interlocking hands).

She concluded that the movement direction was of utmost significance while contributing to the meaning variations in the rotational gestures: a clockwise circle indicated progressive unfolding during the course of time, whereas an anticlockwise circle meant regressive unfolding. Plausible as this sounds, more corpus-based empirical research is required to confirm this finding, which, nonetheless, is not the focus of the current study. Regardless of progressive unfolding or regressive unfolding, both embody the meaning of continuous process.

Although the current study resembles Ladewig and Calbris' research with regards to the exploration of rotational movements, it uses a much more

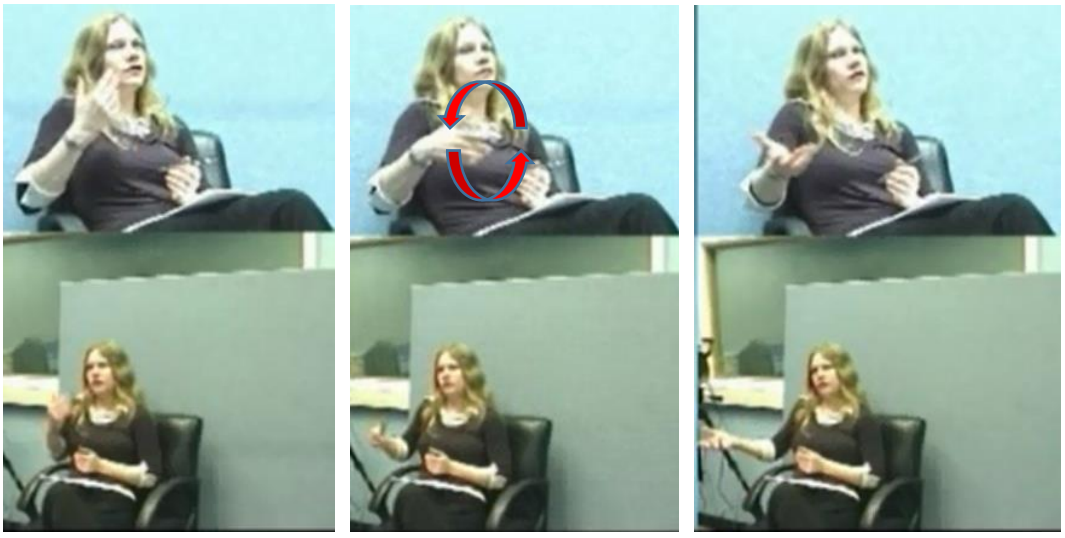
inclusive approach in data selection. It targets all the rotational movements regardless of other parameters. This means that, except for the essential rotational movements, there is no other restrictive criteria in terms of the physical forms of these gestures. The hand may be positioned in the same space while rotating; it can also move to a different space. Also, the hand does not have to move outwards, but can also rotate inwards. In addition, the circular gesture can be performed by one as well as two hands.

The decision of taking an inclusive approach was made considering the little knowledge we have regarding the actual meanings of the circular movements at this preliminary stage of investigation and the limited dataset of the NMMC. That is, we need to include as many instances as possible to identify the meaning variations of the rotational movements. In addition, it has been discovered in the data that many instances that have not been included in Ladewig's analysis also deliver the meaning of continuity as other cyclic gestures do. Moreover, as just stated, in Calbris research of the parameters of the circular gestures, she did not find a significant impact on the meaning of the gestures from the chosen parameters (e.g. number of the circles, body parts, and symmetry). Her research, therefore, indicates that the parameters of the circular gestures tend not to lead to meaning variations.

As is shown in the example (see Figure 5.2 below), the circular gesture co-occurs with the speech, 'to advance it' (all speech co-occurring with the circular gestures were annotated with 'CG# #'). The hand depicts two and a half circles as it moves outwards and forwards from a position close to the speaker to outer space. This instance would not be included in Ladewig's (2011, 2014b) analysis as the gesture changes its space, but can be categorised as the function of describing on-going activities (i.e. referential function) in her coding scheme. As the goal of the current study is to examine the distribution of variations of speech patterns aligned with the circular gestures, excluding instances like this means the risk of losing a

certain number of data and, therefore, affecting the results of pattern analysis.

really good feedback at PALA in terms of how CG#to advance it# further.



2.5 circles, the right hand rotating clockwise

Figure 5.2: An example of a circular gesture changing gesture space as the hand rotates.[6]

Nevertheless, the same as in Ladewig's (2011, 2014b) study, the feature of 'continuity' in the rotational motions maintains an essential criterion for selecting circular gestures. Thus, if a series of circles were interrupted by pauses, they were coded as separate circular gestures. This criterion was used because of the observation that the pauses between gestures tend to be synchronised with the boundaries of semantic units (Calbris, 2011; Kendon, 2004). Moreover, further observation of the current dataset shows that the feature of 'continuity' also indicates the similar velocity/intensity throughout the cyclic movements. This means, when the speed/strength of a certain part

[6] Each image contains two angles of the same participant of the three timestamps.

of the circle is particularly faster/stronger than the rest, it would be excluded from analysis.

A further complication concerns the resemblance between those one-circle circular gestures ending with an open-hand-palm-up hand and the well-known PUOH gestures. There are many instances of PUOH gestures, in which the hand is not fully rotated and/or the latter halves of which are performed at a higher speed. These instances were not coded as circular gestures. However, it seems not uncommon either for a circular gesture ending with a palm-up handshape after multiple rotations (e.g. the instance in Figure 5.2 above). From this perspective, it could be argued that those PUOH gestures with continuous, rotating motions can be regarded as one form of circular gestures.

Potentially, the circles could be depicted by different parts of the hand, primarily the finger and the palm against the rotation of the finger joints, the wrist, or the elbow. But a circular gesture could also involve an object, such as a pen held as an extended element of the hand. The instance below presents such a phenomenon (Figure 5.3), where the pen becomes the major part of rotation.

what outside of the novel was CG#going on# to influence the novel (pause 0.2)



1 circle, the pen rotating clockwise

Figure 5.3: An example of a circular gesture performed by a pen.

Possibly because of the nature of the NMMC, i.e. supervision meetings, it is quite common for supervisors and students to hold a pen and/or some papers. These physical objects then constitute an integral component of gestures, which reflect the dynamics and improvisation of real-life communication. Datasets collected from the real world may well require a more flexible coding scheme than those recorded in the laboratory.

It is crucial to admit that, despite the guidance of the coding scheme, manual coding of spontaneous human behaviour inevitably involves complicated and difficult decisions. This is especially true where the velocity and the trajectory of the gestures are concerned. Without accurate statistics, it is almost impossible to make precise, error-free judgements of the extent to which the velocity of the rotational movements remains constant. Neither is it possible to capture the exact trajectory of each circle, largely due to the shift of the hand movements from 3-dimensional to 2-dimensional in the video recordings. Indeed, it seems rather unlikely that the human hand would always depict perfect round circles without some slight changes of trajectory and/or speed due to our limited control over the hand. After all, humans are not machines. For these aforementioned reasons, it seems justifiable and more feasible as an approach to not let the minor variations in speed and/or trajectory impact the overall coding of the target gestures.

The current data capturing technology also limits the analysis and description of gesture space. As is presented in the figure below (Figure 5.4), due to the position of the camera, it does not always capture the exact front of the speaker. Even if the camera was positioned immediately in front of the subjects, the speakers may move their body and produce a small angle in the image. As the body of a speaker serves as the main reference for judging the position of the hand in relation to the body (refer back to Chapter 2, 2.4.3. for details of coding gesture space), this may affect the accuracy of coding.

Also, the fact that the hand can move further away from the body, i.e. the reference point, also presents challenges to the coding of gesture space.



Figure 5.4: A collection of images captured by the camera.

The coding of the size of the gesture is also challenging with data only captured by 2-D video cameras from one perspective. To date, there exists no systematic scheme to measure the size of gesture. Calbris' definition of the distinctive gesture sizes relies on an observation of the involved parts of the hand and arm in a specific gesture. A circular gesture with a stretched hand and/or with the forearm involved would be coded as large; one with a relaxed hand would be considered as medium in size, whereas one performed only by fingers as small. However, the actual coding may not be as straightforward as described, mainly for the reason that the parts of the hand and arm involved are not necessarily equivalent to the size of the circles depicted by the hand. For instance, it is possible that a stretched hand may depict a smaller circle than a relaxed hand rotating against the wrist. In addition, the extent to which the hand is stretched/ relaxed itself can be ambiguous, and the concepts of 'small', 'medium', or 'big' circles are also vague. Moreover, as the movements of the forearm, the hand and the fingers are all linked, it can be hard to distinguish which part(s) of the arm is/are actually involved.

Owing to the many issues of coding parameters mentioned above, the current study does not include these in the analysis, but only focuses on the

rotational movements. For future studies on circular gestures, relatively less intrusive data collection technology such as light-weighted data gloves could be used to capture more details of the gesture forms (size, space, velocity, trajectory, etc.), although the naturalness and authenticity of interaction could be compromised to some degree. A further alternative is, whereas the current corpus has only used one camera capturing the upper-front body of the participants, more cameras could be set up to better capture various viewpoints: upper front, side, top, etc. This may help researchers obtain a clearer view of the trajectory, size and space of gestures, allowing a more fine-grained analysis to be conducted (Calbris, 2011). To conduct a more detailed analysis of the parameters, regardless of the data collection methods, the corpus will need to be much larger in order to extract enough instances for each category.

5.2.2. The form of the circular gesture and discourse context

Although most circular gestures appear similar due to the common rotational movements, they can vary in various degrees in terms of their hand shape, space, size, trajectory and velocity. This could be the combined effect of multiple factors, ranging from the discourse context (i.e. the speech they co-occur with it), the kinesic environment (i.e. the gestures preceding and following it) and individual factors (i.e. differences in the biological make-up of the hand, gesture style, etc.). Although all those factors mentioned above are worthy of investigation, the influence of discourse context on gesture forms seems to be the most widely researched and acknowledged factor (Calbris, 2011; Kendon, 2004; McNeill, 2005).

Although no systematic analysis of the gesture components has been conducted for the restriction of the data, the remainder of the section will illustrate briefly the ways in which the discourse context can lead to slight

variations in the forms of the circular gestures, including the number of circles, the direction of rotations, the shape of the hand, the size of circles and the space of gesture. However, it has to be emphasised that the slight differences in the gesture forms do not necessarily change the category of speech patterns to which they belong.

5.2.2.1. Number of circles

As continuous movements of different circles are taken as a gestalt (i.e. the entire gesture including its components), the circular gestures can be made up of circles of various amounts, ranging from one to multiple. It seems that, in general, the number of circles is relevant to the length of speech co-occurring with them. For example, those instances containing more lexical items and/or last longer[7] are usually accompanied by more circles in the gesture.

you're you're CG#going to be constantly (pause 0.1)# (pause 0.5) i= (pause 0.4) reflecting (pause 0.2) …



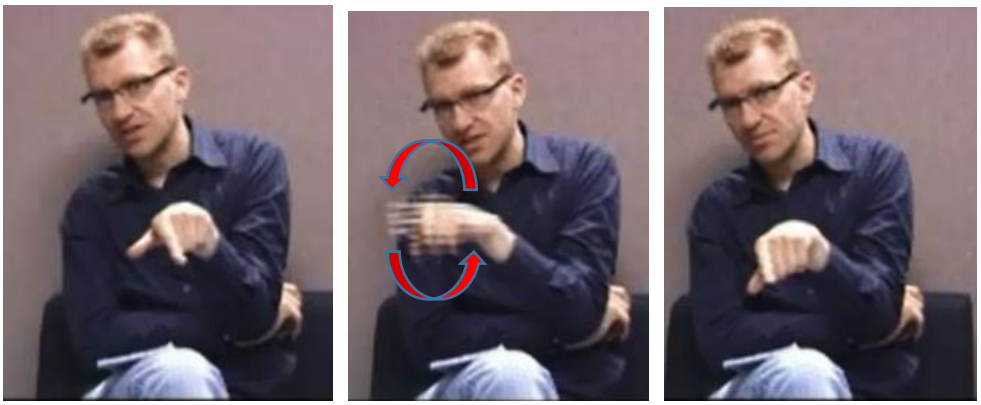
3 circles, both hands rotating inwards, anti-clockwise

Figure 5.5: An example of a circular gesture with 3 circles.

[7] Utterances can be longer due to long pauses.

The two instances in Figure 5.5. and Figure 5.6. demonstrate such a phenomenon. Whereas Figure 5.5 presents a circular gesture with three circles co-occurring with a long utterance (CG#going to be constantly (pause 0.1)#), the other circular gesture in Figure 5.6 has only 1 circle co-occurring with one word, CG#(pause 0.7) Der#by.

like sorry erm that that paper I mentioned at CG#(pause 0.7) Der#by



1 circle, the finger rotating anti-clockwise

Figure 5.6: An example of a circular gesture with 1 circle.

5.2.2.2. Direction of rotation

Although the majority of the circular gestures rotate clockwise, a small number of them move anti-clockwise. As proposed by Calbris (2011), the direction of the rotational movements constitutes a crucial element of meaning making as the clockwise movements reflects progressive unfolding as opposed to regressive, anti-clockwise circles. Her finding is only partially confirmed by the current dataset as not all the circular gestures are associated with the meaning of progression.

Using the same example in Figure 5.5, although it might be more justifiable for an expression such as *going to* to move outwards based on the

sense of future implied in the tense, this circular gesture rotates towards the speaker. This is most likely affected by the word ‘reflecting’ following it, which is also accompanied by a one-circle circular gesture (CG#reflec#ting) with almost the same gesture form. This example not only lends support to Calbris’ finding, but also to the fact that the form of a gesture may not only be influenced by the co-occurring speech, but also the wider discourse context, i.e. the speech occurring before or after. Another anti-clockwise circular gesture, however, may not reflect the meaning of regressive progression at all. The example in Figure 5.6 above demonstrates such a phenomenon as the gesture co-occurs with a new piece of information that has not been mentioned previously, the name of a location, ‘Derby’, where the speaker has presented his paper. Using the same pointing hand shape, the speaker points the index finger downwards while saying “that that paper” before ‘Derby’, anaphorically referring to the paper as an object. The pointing hand is then kept until he says “Derby”, where he depicts a circle with the finger, foregrounding the location of the presentation. Hence, it seems unlikely that the meaning of regressive progression is indicated in this circular gesture co-occurring with the word ‘Darby’.

5.2.2.3. Shape of hand

It seems that speakers, often, use a relaxed open hand to perform the circular gestures. However, the meaning of the speech can lead to various hand shapes. Using the previous example in Figure 5.6, the pointing hand shape co-occurring with *CG#(pause 0.7) Der#by* may well be the result of the deictic word ‘that’ prior to it, which is accompanied by a pointing gesture of the same hand shape.

Observing a few other instances in the category of deictics reveals that, when deictic pronouns such as *this/these*, *that* are used, the pointing hand

shape may be used to perform the circular gestures. In other words, when a circular gesture occurs in a context where the speaker intends to refer to a specific object, regardless of whether concrete or abstract, a pointing hand shape may be adopted. The example in Figure 5.7 represents one of those instances that co-occurs with the deictic expression *your*.

the clearer you get that at this stage then CG#your sort of me# CG#thodological frame#work



1 circle, the finger rotating anti-clockwise

Figure 5.7: An example of a circular gesture with a finger-pointing hand shape.

5.2.2.4. Size of circles

Most of the circular gestures seem to be medium in size, but some may be particularly small or large depending on relevant speech components. An example of such a type follows (Figure 5.8), where the speaker is talking about ‘a list of things’. This circular gesture covers a particularly large space by rotating the hand all the way from the bottom to the top twice and, eventually, returning to the bottom. Therefore, the outline portrayed by the hand is iconic of the layout of things listed in real life and the large size of this gesture emphasises the sense of broadness embedded in the phrase.

CG#gave me a list of things that they are intere#sted



2 circles, the left hand rotating clockwise

Figure 5.8: An example of a circular gesture with a particularly large size.

5.2.2.5. Space of gesture

The discourse surrounding a circular gesture can also have an impact on the gesture space where it is performed. Although most circular gestures seem to be positioned in the broad central space in front of the speaker's upper front body, sometimes they can be performed in a peripheral space influenced by speech meaning.

because of your interpreCG#tation# of (pause 0.5) + <$1> Mm. <$2> +of wh



2 circles, the finger rotating anti-clockwise

Figure 5.9: An example of a circular gesture performed in the upper peripheral space.

For instance, in Figure 5.9, the circular gesture co-occurs with ‘interpretation’, which mainly refers to the thinking/action of the brain. Due to this, the speaker performs this gesture in the upper peripheral space with the finger close to and pointing to the head, with the head simultaneously leaning slightly towards the finger. A similar example was also found in Calbris’ data (2011), which she described as a body-focused gesture as the hand rotates while pointing to the head with the head tilting towards the hand, signifying the cracking of the mind.

Moving forward, research can be conducted to examine the meaning variations of the circular gestures in different forms, those which are formulated by the core rotational movement combining one or more of the component(s) mentioned above. It could be the case that the more restrictive components are included in the sampling process, the less data would be involved and, possibly, the less meaning variations would be expected. However, it is also likely that the variations in gesture components in general may not lead to major meaning variations.

Nevertheless, as emphasised above, this study aims to select as many circular gestures as possible in order to get a broad picture of the lexical and grammatical patterns co-occurring with them, and to establish a comprehensive framework for the unit of meaning of the circular gestures. The following section will then specifically illustrate what is referred to by the unit of meaning model and how this model, with the assistance monomodal corpus techniques, can be used to describe gesture and speech relationship in multimodal research.

5.2.3. The unit of meaning model

As one of the pioneers in Corpus Linguistics, Sinclair (2004) correctly foresees that, since corpus analysis reveals the tendency of the co-occurrence

of certain words, the analysis of the unit of meaning is likely to be extended from single words to the phrasal level, or to multiword expressions. In accordance with his corpus-based findings, he established a model for analysing the unit of meaning, illustrating the relationship between the searched item, or the node, and the words co-occurring with it.

Sinclair's unit of meaning model contains four concepts that can be used for linguistic description, which include *collocation*, *colligation*, *semantic preference* and *semantic prosody* (Sinclair, 2004). The concept of collocation refers to those words frequently co-occurring with the searched item, underpinning which is the conception that words rarely exist independent of other words (Sinclair, 2004). The concept of colligation means the tendency of a certain lexical item co-occurring with a group of words that belong to a particular grammatical structure. Semantic preference is relevant to the co-occurrence of a specific item with a general notion (e.g. *visibility*) expressed by different but semantically related words (e.g. *see, detect, obvious*). Distinct from the other three concepts in the model, semantic prosody is associated with the function, the attitude, or the pragmatic aspect expressed by the discourse environment of the node (e.g., the attitude of reluctance associated with *true feelings*) (Sinclair, 2004).

Using the node *naked eye* as an example, Sinclair (2004) demonstrated that, in 95% of the concordance lines, occupying the first position to the left (N-1) of this node is the word *the*, which is thus the collocate of the *naked eye*. Moving to the left of N-1 (N-2), 90% of the instances in this position was taken up by prepositions including *with, to, from, upon,* etc. This then reveals the colligation of the prepositions of the *naked eye*. Observing the position of N-3 or nearby, those words, particularly *see* and *visible*, were identified, which are indicative of the semantic preference of visibility. So far, the pattern, *visibility + preposition + the + naked eye*, can be established. However, a closer look at the variety of words used to modify *visibility*,

Sinclair (2004) found out that they all embody the meaning of difficulty, hence suggesting the negative semantic prosody associated with the *naked eye*.

Understanding the collocation, colligation, semantic preference and semantic prosody of lexis is especially useful for lexicographers, language researchers and teachers in order to provide language learners with knowledge regarding the dominant uses of lexis of a certain language. As the only concept describing the attitudinal and functional aspects of lexis, semantic prosody, initially proposed by Louw (1993), can shed meaningful light on the attitudes of writers to a lexical unit. For instance, concordance analysis of the word *spinster(s)* in the BNC by Baker (2008) suggests two negative prosodies: sexual frustration (marked by collocates *frustrated, love-starved, lonely, repressed, sex-starved*) and unattractiveness (indicated by *atrocious, plain-Jane, terrible, dried-up, over-made-up, whey-faced*). What can be inferred from research as such is that examining the semantic preference and semantic prosody of key words can be particularly useful for conducting sociolinguistic research exploring attitudes and ideology in relation to issues such as gender, power, race, religion, etc. in different media platforms (e.g. Twitter, Facebook, and news, etc.) (Baker, 2010).

Despite the fact that this model has been, and is still being used, by corpus-based text analysts as a vital tool for linguistic and sociolinguistic analysis (Adolphs, Brown, Carter, Crawford, & Sahota, 2004; Atkins & Harvey, 2010; Biber et al., 1998; Kennedy, 1998; Romero-Trillo, 2008), it has not as yet been applied to describing the relationship between speech and gesture. The following sections will then focus on reporting the procedures of exploring the unit of meaning of the circular gestures in relation to the emerging speech patterns co-occurring with them. During this process, basic corpus techniques are creatively and strategically applied to analyse gestures,

which provides valuable insights into the ways in which software designed for textual analysis can also be important for multimodal research on gesture.

5.2.4. The use of corpus techniques in multimodal analysis

This section will focus on reporting the corpus techniques adopted in the study. As a multimodal corpus-based study, the purposes of the research require the use of both monomodal and multimodal toolkits. Whereas multimodal tools are crucial for gesture segmentation and annotation, monomodal software can be valuable for identifying speech patterns.

In respect of the gesture, after deciding the principles of selecting instances of the circular gestures (see 5.2.1 for details), both gestures and the aligned speech were annotated in ELAN. As the aim of the study only targets the meaning of the gesture, only the strokes of the circular gestures, i.e. the rotational movements were annotated, as the stroke is the gesture phase where meaning is embodied. The speech parts in line with the strokes were then transcribed verbatim together with details of dysfluent features (McCarthy & Carter, 1995) such as 'erm', pauses (more discussion of the length of pauses will be in 5.2.5.), stretching component of syllables (marked by '='), and incomplete speech. The detailed transcription of such dysfluent features is due to that this may be one of the emerging patterns. This speculation is based on the preliminary observation of the data and the considerable number of circular gestures co-occurring with, or adjacent to, pauses in Ladewig's (2011) research on the cyclic gestures. She concluded that these circular gestures perform crucial cognitive functions in terms of helping the speakers to retrieve word concepts.

Following the multimodal mark-up of gesture and speech in ELAN, the next step involved annotating the speech co-occurring with all the circular gestures in texts. Textual annotation is essential for monomodal corpus

investigation as it adds extra value to the corpus by allowing for automatic corpus search and in-depth analysis (Hunston, 2002). Although there are some popular mark-up systems available such as XML (refer back to Chapter 2, 2.2.2.3. for more mark-up schemes) (McEnery & Wilson, 1996), the current study chose a simpler and shorter symbol, 'CG# #', to obtain more speech context in each line. Many utterances co-occurring with the circular gestures can be lengthy so that the shorter the tag is, the better the chance that all the speech can be included and represented in the same concordance line.

All the speech components that co-occur with the circular gestures were annotated in the speech transcription with 'CG# #'. The two hashtags serve the purpose of marking-up the beginning and end of the speech aligned with the gestures. The texts were then readily available for basic automatic corpus research, the results of which can be exported to other software such as Excel and Word for further qualitative and quantitative analysis. Among the different software packages available, AntConc was chosen here as it provides adequate functions for the current investigation and is freely available online for download and usage.

5.2.4.1. Concordance

As one of the fundamental research methods in corpus linguistics, the concordance function generates a list of all the instances of utterances positioned to the left or right side of the searched item, or the node (Kennedy, 1998). In textual analysis, the concordance lists are extremely useful for discovering the most frequent uses and meaning variations of a word or a phrase, as well as the most frequent words/phrases that co-occur with them (Hunston, 2002). This method is crucial for many areas of application, especially dictionary compilation and language teaching, as corpus findings

usually reveal lexical and grammatical patterns that cannot always be captured by native-speaker intuitions (McEnery & Wilson, 1996; Sinclair, 1996).

Whereas the concordancing tool is offered by many textual analytical software packages including AntConc, it is not equipped for generating gesture concordance, let alone the co-occurring speech. This is because, to generate a concordance list, one needs to have one specific search item; however, most gestures are too variable for building up automatic searching and extracting programmes and, to date, most fine-grained gesture annotations can only be done manually. Multimodal alignment, annotation and observation tools are currently available, such as Anvil, Elan and DRS, among many others. These are mostly used for facilitating gesture observation and analysis (refer back to Chapter 2, 2.4.5. for more multimodal alignment and annotation tools).

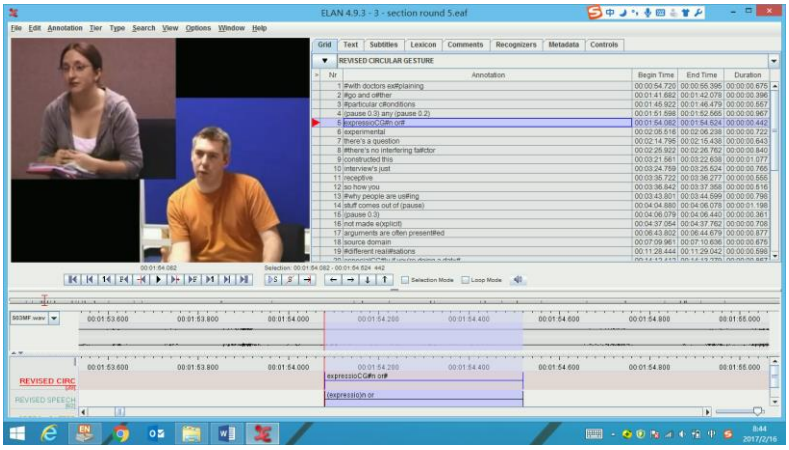


Figure 5.10: A screen shot of Elan.

As can be observed in the screen shot of the ELAN interface above (Figure 5.10), all the 570 instances of CGs were identified, segmented and annotated, along with the speech components co-occurring with them. The researcher can conveniently jump to any chosen instance for analysis, but also export

pictures of each gesture to a Word document for further examination (e.g. manually creating a concordance of gestures).

Compared to gesture analysis, the concordance list of all the speech that co-occurs with the circular gestures can be extracted much more efficiently with the aid of AntConc. As all the speech parts were annotated in the speech transcripts using the marker, 'CG# #', by running a search for the item 'CG' in AntConc, all the instances can be automatically listed and sorted. Below (Figure 5.11) is a screen shot of the concordance list of all the co-gestural speech in the transcripts of the eight videos with all the instances sorted alphabetically. This concordance accumulates instances of similar lexical and grammatical structures, and therefore significantly facilitates the identification and description of the emerging speech patterns.

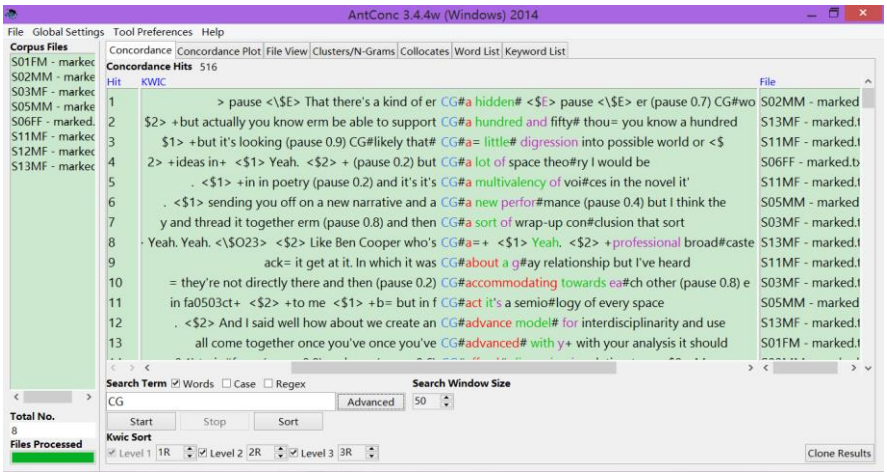


Figure 5.11: A screen shot of a concordance list obtained by searching 'CG'.

5.2.4.2. Frequency analysis

Concordance plot has been used as a function useful for examining the number and distribution of the searched item, or the hits (see Figure 5.12 below). The figure reveals that the largest and the smallest number of the circular gestures emerge in S02 and S03, respectively, with the larger one

almost triple the size of the smaller. A fairly large number of instances were also found in S06 (89), S12 (81), S01 (72) and S13 (70), followed by S05 (51) and S11 (52). Concordance plot also helps visualise the distribution of all the occurrences of the circular gestures in each video file, which seems to occur across most of the supervision meetings.

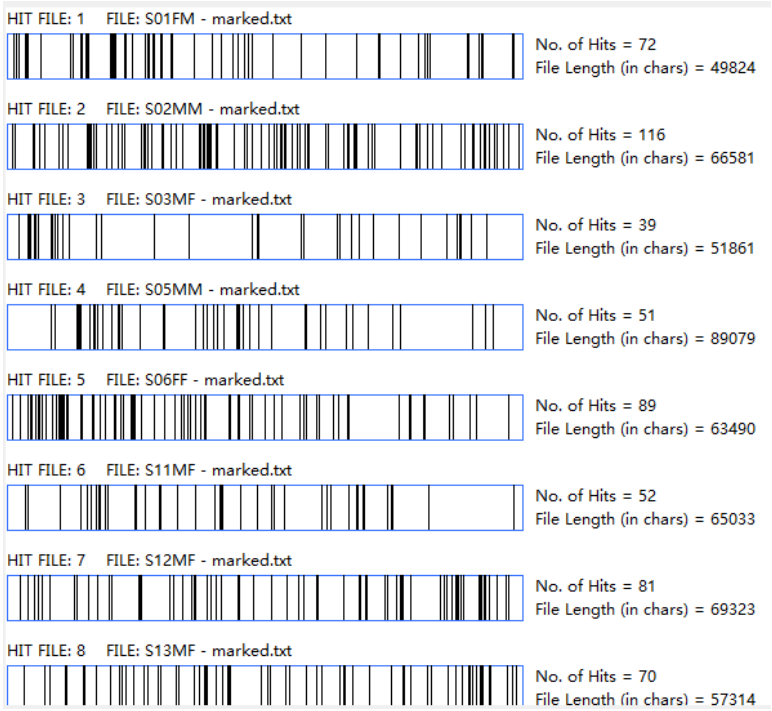


Figure 5.12: The concordance plot of circular gestures.

However, considering the variations of the length of the supervision meetings, a more valid analysis of the frequency of occurrence of circular gestures in each supervision meeting involves counting the average number of circular gestures per minute (see Table 5.2 below). In general, the circular gestures tend to occur frequently with an average of at least once per minute in almost all the videos except for video 4.

Table 5.2: The number of circular gestures per minute.

Item	Video	Minute	Occurrence	Occurrence/Min
1	1	35	72	2.06
2	2	60	116	1.93
3	3	41	39	0.95
4	5	88	51	0.58
5	6	47	89	1.89
6	11	43	52	1.21
7	12	55	81	1.47
8	13	37	70	1.89
	Average	51	71	1.50

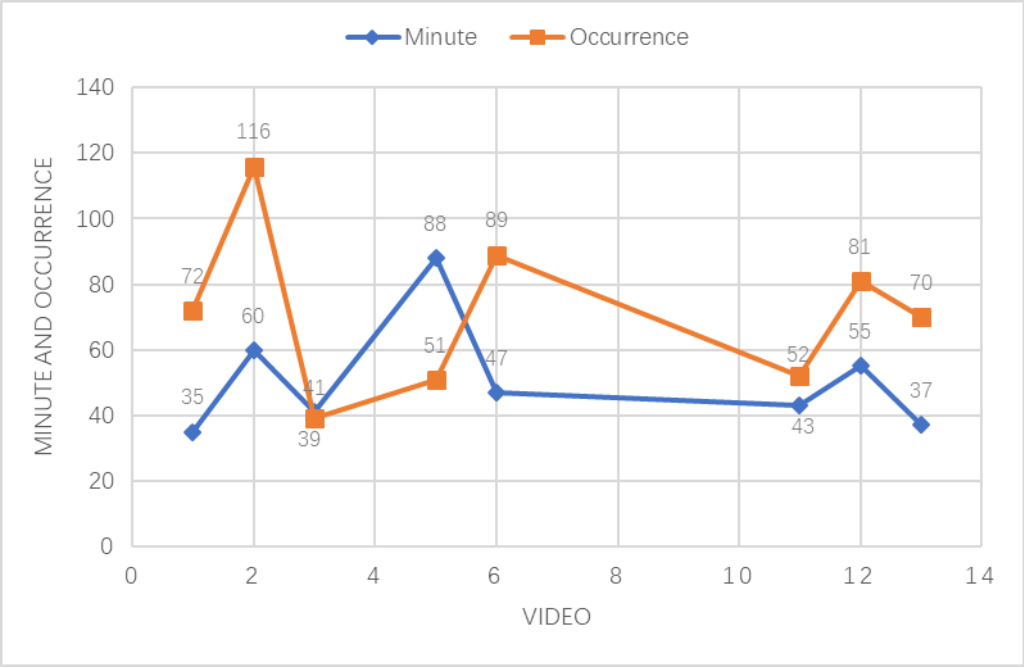


Figure 5.13: The correlation between the length of the video and the occurrence of the circular gesture.

Figure 5.13 above presents the correlation between the length of each video and the occurrence of the circular gestures, the results of which are not entirely straightforward. For most videos (e.g. 1, 2, 6, 12, 13), the relatively

longer the video, the larger the number of generated instances is; however items 3, 5 and 11 do not have the expected number of instances in accordance with their length. The correlation analysis seems to indicate that, although corpus size is generally crucial for generating a large number of instances for investigation, sometimes, the size of the data does not necessarily guarantee the generation of the target nonverbal features. This is probably owing to the occurrence of gestures in general, and the fact that the occurrences of different gesture types are closely related to a variety of factors, such as the nature of the task, the physical environment and individual styles, etc. Probably for this reason, previous gesture studies (Gullberg, 1998; Kita, 2009; McNeill, 2005) were inclined to design a controlled environment for data collection for the purpose of obtaining a sufficient number of instances of the target gestures.

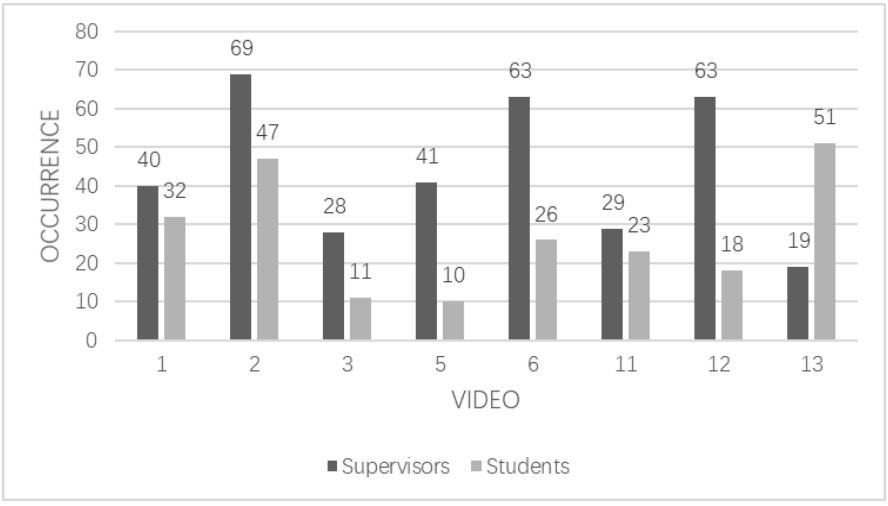


Figure 5.14: The distribution of circular gestures by supervisors and students.

Whereas Figure 5.13 focuses on comparing the frequencies of circular gestures across the videos using per minute frequencies, Figure 5.14 above shows the raw frequency of them contributed by each speaker. The raw frequency is used here mainly because, compared to per minute frequency, it can better demonstrate the origins of the instances, i.e. who performed those

circular gestures in the NMMC. Such information is essential to measure the representativeness of the datasets. What can be inferred from Figure 5.13 is that there seem to be relatively major differences in the number of instances contributed by some speakers. Whereas the largest number reaches 69 by Supervisor 2, Student 5 only performs 10 circular gestures. In view of the uneven distribution of the circular gestures among speakers, caution is needed when making generalisations and more research involving a much larger dataset is required to validate the findings of the current research.

In addition, another observation concerns the distribution of the circular gestures between supervisors and students. As can be observed from Figure 5.14, although both parties perform the circular gestures, supervisors overall contribute more instances than the students (except for video 13). This may be a result of the fact that supervisors normally speak more than students in supervision meetings, which can be a reflection of their more powerful position in comparison to the students. To confirm this, per minute and per word frequencies of both parties will have to be calculated and compared. However, such an analysis is not necessary here as the issue of power relationship between supervisors and students is not the concern of the current study.

Hence, as the major goal of the current research is to demonstrate a multimodal corpus-based approach to explore the unit of meaning of the circular gestures, the differences in the number of the instances is not a major issue, at least at this stage of research. Also, compared to previous gesture studies, the size of the data involved here is already reasonably large, probably one of the largest in the area of hand gesture at this level of granularity. The small size of the majority of existing gesture research is mainly caused by the extremely painstaking and time-consuming nature of manual gesture segmentation and annotation.

5.2.4.3. Clusters/N-Grams

Whereas the concordance list of the circular gestures enables the discovery of similar linguistic patterns in alphabetical order, a search for the clusters/N-grams helps the researcher foreground the most frequently occurring lexical items to the right of the node ‘CG’. The ultimate purpose of this analysis is not to pinpoint clusters as in most previous studies, but to identify the top recurrent lexical items that co-occur with the circular gestures and the primary positions they occur in discourse.

Table 5.3: The first 10 most frequently clusters/N-Grams to the right of ‘cg’.

#Total No. of Cluster Types: 310 (Cluster range = 2)				
#Total No. of Cluster Tokens: 570				
Rank	Frequency	Range	Cluster	Percentage
1	52	8	cg#(pause	9.12%
2	21	8	cg#and	3.68%
3	20	7	cg#you	3.51%
4	15	6	cg#the	2.63%
5	12	7	cg#i	2.11%
6	10	6	cg#so	1.75%
7	9	7	cg#a	1.58%
8	9	5	cg#that	1.58%
9	8	6	cg#it	1.40%
10	7	5	cg#to	1.23%

Table 5.3 above shows the result of such an analysis, revealing a significant number of instances containing pauses, *and*, *you*, *the* and *I*. Based on this, further discourse analysis of these frequently occurring items in conversation

was then conducted. This method was proven to be very useful for identifying and describing the temporal relationship between the circular gestures and the speech co-occurring with them (see section 5.3.1.2. for a detailed discussion of these results).

In addition to the basic cluster analysis, there are potentially more complicated analyses that can be conducted. For instance, the concordance in most corpus tools allows for a search of a node in relation to certain words in a certain span. This can then be used to explore any item of interest co-occurring with, or adjacent to, the circular gestures. Additionally, the use of wild cards can also be beneficial. For instance, a search for the cluster 'cg#and'[8] in AntConc would show all the instances co-occurring with the circular gestures with the word 'and' in the second position.

5.2.5. Coding lexical and grammatical patterning

As illustrated above, concordance lines can be used to explore the lexico-grammatical patterns of speech. However, when dealing with an extremely long list of instances, the human eye can only manage to observe a limited number of instances. This number might be approximately 100 instances for observing general patterns and 30 for exploring details (Hunston, 2002). For this reason, sampling techniques, such as a random selection of 30 instances (Sinclair, 1999), have been proposed in textual analysis. That is, randomly select 30 concordance lines and analyse patterns, then select another 30 for any new patterns and so on, until no new patterns can be identified.

[8] In AntConc, the wildcard symbol '#' stands for any word located between 'cg' and 'and' rather than a tag delimiter as it is used in the current thesis. The more the '#' is added between 'CG' and 'and', the more words occurring between them.

Similarly, it is impossible to explore speech patterns in all the 570 concordance lines without adopting any sampling technique. Therefore, a sampling approach was specifically designed for the current study in order to build a descriptive framework for categorising speech components that co-occur with all the circular gestures. The whole process began with a pilot analysis of the first 5 instances from the eight videos regardless of whether they were contributed by supervisors or students (40 instances for each round). Following this, the following 5 instances from each video were then selected to identify new patterns and adjust the old ones. After repeating the same process three times, the initial framework was established. Taking into consideration the potential differences among the videos and the speakers, the purpose of selecting 5 examples from each video is mainly to include an equal number of instances across all the supervision meetings, which may lead to a more inclusive framework.

This initial framework was then tested and adopted to categorise all the instances in the concordance list, during which process small revisions were then made in order to attempt to make it as inclusive as possible. Ultimately, a lexico-grammatical coding scheme of 15 categories was able to be generated, which includes 1) on-going action, 2) noun phrases, 3) modification, 4) deixis, 5) dysfluent speech, 6) intensification, 7) the beginning of clauses introduced by pronouns, *th* (e.g. *this* etc.) and *wh* words (e.g. *what* etc.), 8) compare and contrast, 9) preposition phrases, 10) vague language, 11) negation, 12) monitoring mutual ground, 13) discourse markers, 14) conjunction, 15) confirmation.

Below is a table of the descriptive speech patterns and their definitions (Table 5.4). A more fine-grained illustration of each category will be provided in 5.3.2.

Table 5.4: Descriptive coding scheme for speech co-occurring with circular gestures.

Speech patterns	Definition
on-going action	Speech that indicates the meaning of continuous progress or events, such as those verb frames marked by–*ing* forms.
noun phrase	Single nouns and noun phrases.
modification	Expressions that are used to qualify or quantify someone or something, such as adjective phrases and *of* structure.
deixis	Those speech parts that mainly clarify the objects, time and locations referred by expressions such as *it*, *this/these*, *that/those*, *here/there*, etc.
dysfluent speech	Short utterances mainly comprise indicators of dysfluency such as *erm*, *er*, pauses (longer than 0.2 second), sound stretching (marked by =), repetitions, false starts, etc. In long utterances, at least two dysfluency features are required.
intensification	Utterances containing intensifiers, mostly adverb phrases, such as *particularly*, *really*, *very*, *whole*, *quite, just*, ect., which are usually situated at the end of the speech parts co-occurring with the CGs, i.e. the gesture stops after they have been uttered.
the beginning of clauses	Speech that mainly constitutes the beginning of clauses led by pronouns (e.g. *this* etc.) and *wh* words (e.g. *what* etc.).
compare and contrast	Speech that contains comparative and contrastive expressions such as *differen*t, __*er* and __ *than*.
preposition (phrase)	Speech that co-occurs with a preposition or preposition phrases such as *for* ___, *on* ___.
vagueness	Speech that primarily entails hedging such as *seem*, *maybe* and vague language such as *kind of*, *sort of*, *a little bit*.

negation	The part of an utterance that mainly expresses the meaning of negation marked by *no* or *not*.
monitoring mutual ground	Speech that is composed of the end of the turn and a pause, followed by certain response(s) from the listener.
discourse marker	Expressions such as *you know*, *I mean*, *for example* that occur at the beginning of an utterance.
conjunction	Linking expressions such as *and*, *or*, etc.
confirmation	Speech from the listener that confirms the speaker with *yes* or *exactly*, etc.

It has to be admitted that these categories are by no means entirely exclusive to each other. One of the major challenges during the processes of building this framework is due to the high frequency of some lexis such as pronouns, functional words and other prevailing spoken discourse features (e.g. dysfluent features, discourse markers, vague languages, etc.) (McCarthy & Carter, 1995). Specific examples include pronouns (e.g. *you*, *this*, *that*, *wh-clause*, etc.), prepositions (e.g. *into*, *on*, etc.), conjunctions (e.g. *and*), discourse markers (e.g. *so*, *and*, *you know*, *I mean*, etc.) and vague language (e.g. *seems*, *a little bit*, *sort of*, *kind of*, etc.), among many others.

The linguistic features mentioned above have led to overlaps and challenges in classification. The solution to this issue is that only those instances that dominantly co-occur with them were coded as such. For example, as pause is one of the most frequently occurring feature, an instance such as 'CG#(pause 0.7) Der#by' was coded as a noun phrase rather than a pause. Similarly, expressions such as *kind of* are so ubiquitous that an instance such as 'CG#there's no kind of# democratic (pause 0.4)', was assigned to the category of negation due to the principal meaning expressed in this utterance.

A further challenge worth mentioning relates to pauses. So far, there seems

to be a lack of distinction between normal and long pauses. Normal pauses tend to be necessary for marking the boundaries, demarcating the entire discourse of interactions into units of meaning of various sizes. However, long pauses can be seen as an indicator of dysfluent speech caused by difficulties in language processing and problems in communication. An investigation of the transcription conventions in some of the most well-established works in Conversational Analysis reveals that 0.2 second is the minimum length that was annotated (Atkinson & Heritage, 1984; Have, 2007; Jefferson, 2004), which also seems to be implicitly used as an indicator of dysfluency. This is similar to

the minimum threshold for a pause in fluency research, which lies at around 0.2-0.3 seconds (Dahlmann & Adolphs, 2007). For the sake of consistency, the current research adopted the same standard. All the pauses were annotated in ELAN with the aid of sound waves as visualisation of acoustic channel, helping the transcriber locate the start and end of a pause.

Due to the prevalence of pauses, it seems unjustifiable to code all the instances involving a pause longer than 0.2 second as a dysfluent utterance, especially those short utterances containing less than around 5 words. Therefore, another criterion of coding has been adopted, which fixates on the number of dysfluent features rather than the length. That is, for long utterances, in addition to pauses, at least one of the other dysfluency features (including pauses) needs to be present in order for the instance to be classified as dysfluent. For example, in a short utterance such as ‘CG#and have (pause 0.2) worked though#’ with only one 0.2-second pause was not coded as dysfluent speech, but as an on-going action in accordance with the core meaning of the utterance indicated in the expression ‘work through’.

5.2.6. The multimodal unit of meaning of the circular gesture

Before a detailed demonstration of how the unit of meaning model has been adapted, two fundamental differences between the application of the concepts in Sinclair's (2004) original model and the current study have to be clarified. One is that, whereas the unit of meaning model attempts to describe the relationship between the node and the frequently co-occurring words, here it aims to illustrate the relationship between the target gestures and the speech co-occurring with them. The other distinction is that, in textual concordance analysis, the position of the co-occurring words tends to move further away from the node from the concept of collocation to colligation to semantic preference to semantic prosody, such a tendency is not reflected in the current uses of the four concepts since they were all used to describe the speech in line with the circular gestures.

In addition to the aforementioned distinctions, a further vital point is that not all the four categories in the unit of meaning model were able to be adopted in the current analysis. As *collocation* refers to the dominant co-occurrence of specific words with the node, in this case the circular gesture, this concept does not allow such a kind of description, as in most instances the words co-occurring with the circular gestures vary considerably. For this reason, only three categories were applied here, namely, *colligation*, *semantic preference* and *semantic prosody*.

As colligation refers to the grammatical co-occurrence between the node and certain grammatical choices, the speech patterns that can be categorised into this concept include noun (phrase), modification, deixis, dysfluent speech, clauses, preposition (phrase), discourse marker, conjunction. These patterns represent the grammatical features that co-occur with the circular gestures.

As those instances in the speech pattern of on-going action and negation all share the same meaning, they were assigned under the concept of semantic preference. On the other hand, it seems that the speech belonging to the patterning of intensification, compare and contrast, vagueness, monitoring mutual ground, confirmation tends to serve important pragmatic functions, express attitudes and enhance inter-subjectivity. For this reason, they were designated as semantic prosody. Specifically, they can serve the purposes of emphasising noteworthy information (e.g. intensification), differentiating meaning (e.g. compare and contrast), softening speech (e.g. vagueness), reassuring listener understanding (e.g. monitoring mutual ground) and indicating agreement (e.g. confirmation).

Whereas the coding of most instances is quite straightforward, the categorisation of some patterns can be more challenging. This is expected as certain semantic preference (e.g. negation) and semantic prosody (e.g. confirmation) may inevitably involve similar grammatical features (*not* and *no* for negation; *yes* for confirmation). On the other hand, certain colligation (e.g. deixis) can project similar semantic preferences (the meaning of time and space relation).

5.3. Data analysis and results

This section will report on the results of the temporal analysis between the circular gestures and the co-occurring speech. It will also demonstrate how the unit of meaning of the circular gestures was generated based on the co-occurring speech patterns. During the processes, brief descriptions of the ways in which corpus techniques have been incorporated into the analysis will also be provided.

5.3.1. Temporal relationship between speech and the circular gesture

The concordance list and cluster analysis enabled the investigation of the temporal alignment between the circular gestures and the co-occurring speech. Two observations were made based on the emerging characteristics in the dataset. One was the co-occurrence of the circular gestures with multiword strings; the other was their tendency to occur at the beginning of an utterance. The remainder of this section will primarily focus on reporting on these two findings.

5.3.1.1. The length of the words

Although the strokes of gestures tend to be in temporal proximity of their co-expressive speech components, they do not necessarily start or end at the exact speech boundaries. An instance such as ‘CG#narrating their ex#perience of using these drug’ means that the rotating movements ends in the middle of the word ‘experience’. Therefore, the decision had to be made in terms of how to count the number of words in such instances. In the current study, any word that has two letters that co-occur with the CG was coded as one word. The threshold of two letters was decided on based on the potential error of one syllable while transcribing the co-gestural speech. That is, if a word has at least two syllables co-occurring with the gesture, there is a much better chance that it is actually aligned with the gesture.

Contracted expressions such as “that’s” was coded as two words; those instances that co-occur with pauses were coded as “0”. When co-occurring with incomplete speech, regardless of whether a word has been completely uttered or not, as long as there are two letters co-occurring with the gesture, this would be counted as one word. Finally, overlapping speech and/or

listener responses were not counted as words co-occurring with the circular gesture.

Table 5.5 and Figure 5.15 below illustrate the frequency of the instances with the distinctive number of words co-occurring with the circular gestures.

Table 5.5: Frequency of the number of words co-occurring with the circular gestures.

Words involved	Frequency	Percentage
0	9	1.58%
1	76	13.33%
2	119	20.88%
3	115	20.18%
4	100	17.54%
5	70	12.28%
6	30	5.26%
7	23	4.04%
8	11	1.93%
9	10	1.75%
10	2	0.35%
11	3	0.53%
13	1	0.18%
20	1	0.18%
Total	570	100.00%

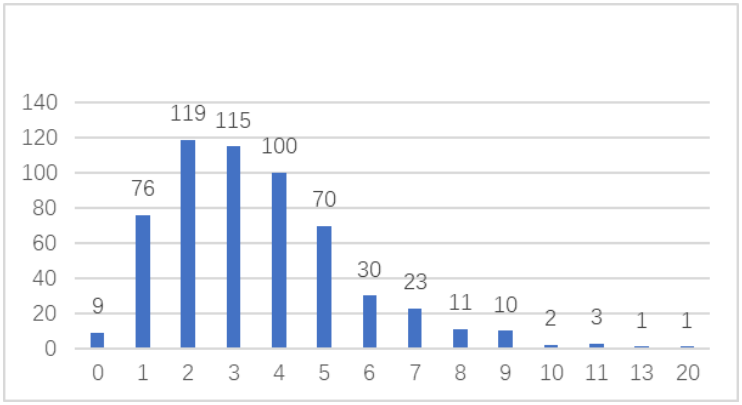


Figure 5.15: Frequency of number of words co-occurring with the circular gestures.

As can be seen, there are 8 instances that do not co-occur with any words, i.e. pauses. A large number of circular gestures tend to co-occur with word clusters that are composed of two, three and four words, with 119, 115 and 100 instances, respectively. They comprise approximately 60% of all the instances. Five-word strings also have a fairly large number, 70 instances, around 12%, followed by 6-word (30, 5%) and 7-word (23, 4%) expressions.

Whereas a more detailed description of patterning will be presented later, a general quantitative analysis as such suggests that it is likely that circular gestures tend to co-occur with multiword expressions that are composed of at least two words. This is especially true for those instances consisting of more than 4 words. When the number reaches and goes beyond 5 words, the chance of a circular gesture aligned with more than one formulaic expression increases. For those instances containing more than 10 words (see Table 5.6 below), they not only involve multiple formulaic sequences, but also clause(s).

Table 5.6: Examples of speech co-occurring with circular gestures containing more than 10 words.

Item	Left	Speech co-occurring with the circular gestures (marked by CG# #)	Word length
223	ledge can be can be reinvigorated and <$O64> also	CG#keep it at the cutting edge of it's it's sector and it's field and it's <\$O64> understanding#	20
150	the doctors who run the site erm (pause 1) s=	#gave me a list of things that they are intere#sted	10
352	= you know a hundred and fifty <$O51> projects to	CG#(pause 0.1) to do some fact finding <\$O51> and and some relationship buil#ding	10
111	know I like the idea of looking at the	CG#detective and the movement round the hotel+ <$1> Right. <$2> +space a= (pause 0.3) in oppo#sition	11
181	(pause 2) he's (pause 1) <$=> c= I think he <\$=>	CG#I saw him sort of smiling when Tim Jones was doing#	11
444	and <$O44> heritage and <\$O44> and	CG#that is a contraflow it <$O45> is a a kind of exchan#ge	11
183	of <$E> pause <\$E> jokey emails. Erm (pause 0.6)	CG#I think we discussed this before where we said that (pause 0.2) possibly there's the#	13
311	"I can't cope any02110more" and erm (pause 0.3)	CG#(pause 0.5) a lot of them er (pause 0.2) talk about relationships going wrong= and "I'll always love you=" and very dramatic er (pause 0.8)#	20

In terms of the duration of the circular gestures, the bar chart (Figure 5.16) and the table (Table 5.7) below reveal an average length of 0.8 second per gesture. This number tends to be shared by most speakers except for the

student in Video 2 and the supervisor in Video 3, both of whom, on average, spend more than 1 second performing a circular gesture. The similar amount of time in individual speakers seems to indicate a common pattern of use in terms of the duration of the circular gestures across this certain group of people, which might also be applied to other speakers.

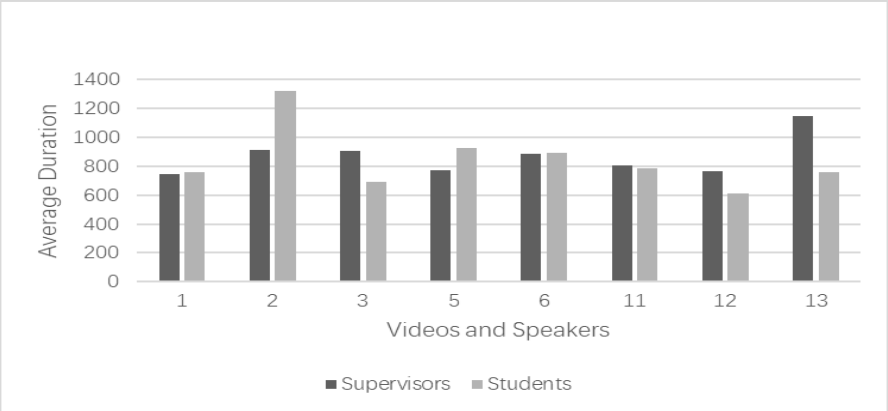


Figure 5.16: The average duration of circular gestures performed by each supervisor and student in each video.

Table 5.7: The average duration of circular gestures performed by each supervisor and student.

Videos	Supervisors	Students
1	745 m/s	756 m/s
2	914 m/s	1320 m/s
3	903 m/s	689 m/s
5	769 m/s	926 m/s
6	887 m/s	894 m/s
11	809 m/s	787 m/s
12	764 m/s	608 m/s
13	1146 m/s	758 m/s
Average	867 m/s	842 m/s

5.3.1.2. The position of circular gesture and linguistic patterning

A further investigation into the temporal relationship was based on a search for the most frequently occurring clusters to the right of the node 'CG' in AntConc, which foregrounds some distinctive features. As is demonstrated in Table 5.8, among the top 10 most frequently occurring, the one at the very top is pause. A closer inspection shows that 52 instances of the circular gestures (nearly 10% of all the instances) co-occur with speech including pauses of various lengths, most of which (42 out of 52) are located in the middle of the utterances. This then indicates the possibility of many dysfluent utterances co-occurring with this gesture.

Table 5.8: Top 10 most frequently occurring clusters to the right of the node 'CG'.

Rank	Frequency	Range	Cluster	Percentage
1	52	8	cg#(pause	9.12%
2	21	8	cg#and	3.68%
3	20	7	cg#you	3.51%
4	15	6	cg#the	2.63%
5	12	7	cg#i	2.11%
6	10	6	cg#so	1.75%
7	9	7	cg#a	1.58%
8	9	5	cg#that	1.58%
9	8	6	cg#it	1.40%
10	7	5	cg#to	1.23%

Apart from pauses, a further analysis of the other three clusters (*and*, *you*, *I*) among the top five items on the list reveals that numerous circular gestures

occur at the onset of an utterance. For instance, 21 instances of *and* are discourse markers positioned at the beginning of the utterances. They are usually followed by other discourse markers such as *you know*, *I mean/suppose*, and the beginning of *th-*/*wh*-clauses. Similarly, there were 20 instances of sentence stems that begin with "you" (e.g. "you know", "you can/could/have") and another 12 instances beginning with "I" (e.g. "I mean/think").

With regards to the third item on the frequency list, "the" (15 instances), textual analysis demonstrates the co-occurrence of some circular gestures with modifications. Two rather similar structures emerge, which are *the + noun/adjective phases* and *the __ of __* (e.g. 'the quality of nature', 'the power of control'). In view of the above discussion, it seems that the circular gestures tend to occur at the beginning or in the middle of utterances rather than at the end.

5.3.2. The multimodal unit of meaning of the circular gesture

Drawing on Sinclair's unit of meaning model and the speech patterns explored in the previous sessions, a profile for the unit of meaning of the circular gestures was constructed (see Figure 5.17 below) (refer to File 5.1 in the CD-ROM data disk for all the instances in each speech pattern, and File 5.2 to File 5.16 for the gesture transcripts for all the instances in each pattern from all the speakers).

The results show the circular gestures are more likely to co-occur with grammatical features such as noun (phrase) (64), modification (62), deixis (55), dysfluent speech (47) and clauses (35) than preposition phrase (17), discourse marker (8) and conjunction (7). Semantically, a considerable number of the circular gestures co-occur with on-going action (172 instances), indicating the meaning of continuity. Only 15 instances align with

the meaning of negation. With regards to the semantic prosody of the circular gestures, they are more likely to co-occur with the speech patterns that serve the functions of intensification (39), compare and contrast (20), and vagueness (16) than monitoring mutual ground (10) and confirmation (3).

The overall distribution of the 15 types of the semantic and grammatical patterns (see Table 5.9 and Figure 5.18 below) shows that a significant number of instances (172) have a semantic preference for describing on-going action (i.e. continuity), occupying more than 30% of the total number. This number almost equals the sum of the following three grammatical categories, i.e. noun phrases, modification and deixis, with about 60 instances (10%) for each. For the remainder of the instances, although dysfluent speech (47, 8.25%), intensification (39, 6.84%) and clauses (30, 5.36%) occupy a relatively larger number of instances, the number in other categories is small, especially in respect of the last four.

The remaining part of the section will report on some emerging recurrent linguistic features based on the colligation, semantic preference and semantic prosody of the circular gestures.

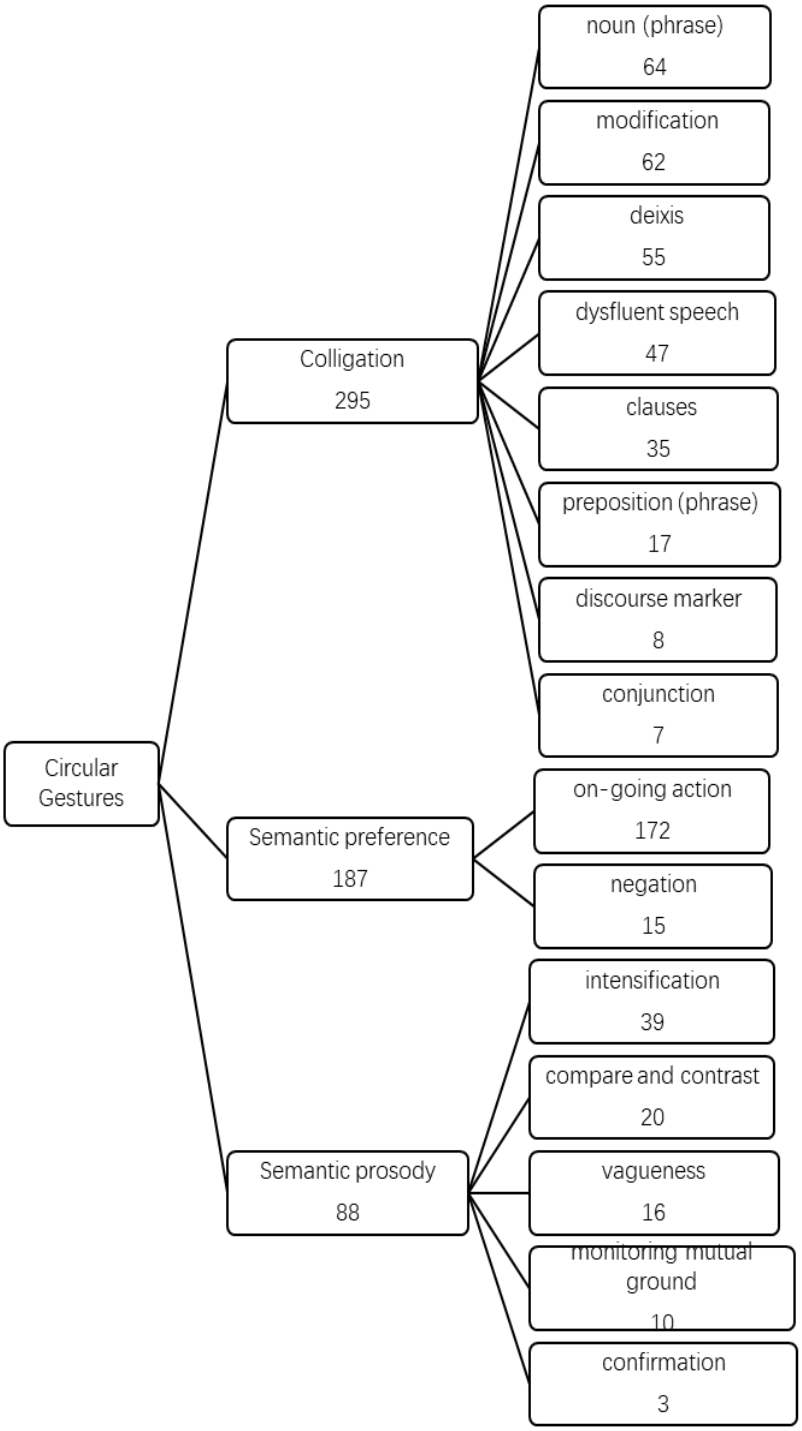


Figure 5.17: The unit of meaning for circular gestures in relation to co-occurring speech.

Table 5.9: Distribution of frequency of occurrence of speech patterns.

Speech pattern	Tokens	Percentage	The unit of meaning model
on-going action	172	30.18%	Semantic preference
noun (phrase)	64	11.23%	Colligation
modification	62	10.88%	Colligation
deixis	60	10.53%	Colligation
dysfluent speech	47	8.25%	Colligation
intensification	39	6.84%	Semantic prosody
clauses	30	5.26%	Colligation
compare and contrast	20	3.51%	Semantic prosody
preposition (phrase)	17	2.98%	Colligation
vagueness	16	2.81%	Semantic prosody
negation	15	2.63%	Semantic preference
monitoring mutual ground	10	1.75%	Semantic prosody
discourse marker	8	1.40%	Colligation
conjunction	7	1.23%	Colligation
confirmation	3	0.53%	Semantic prosody
Total	570	100.00%	

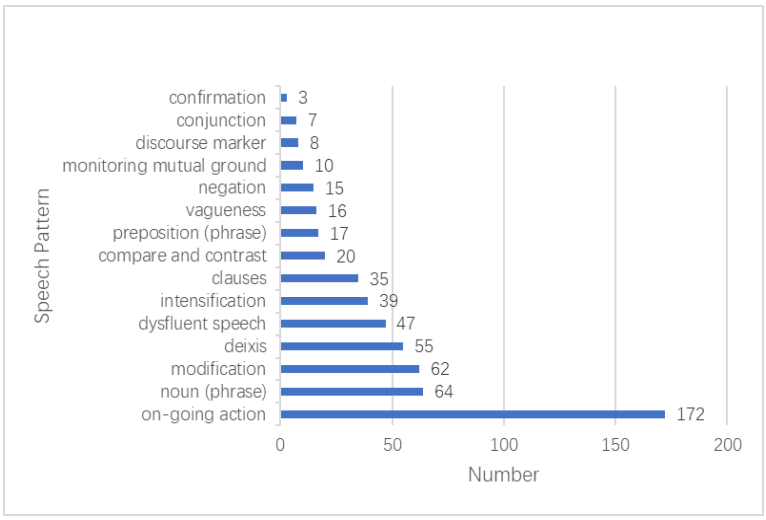


Figure 5.18: Distribution of frequency of occurrence of speech patterns.

5.3.2.1. Colligation

✧ Noun phrases

The number of instances that belong to the category of noun phrases is ranked second among all the emerging patterns. An observation of all the instances suggests that they mostly refer to certain objects, events, entities, names, theories, etc. Instances include *interview* (201), *occupation* (288), *prevailing judgements* (369), *corpus linguistics* (372), *relationships* (389), *Tim Johns* (489), among many others. Further analysis of the grammatical structure foregrounds the prevailing structure of *a/an/the + modifier + noun*, with slight variations in their combination, such as *the garage door* (144), *a new performance* (7), etc.

Another noteworthy feature in this category is the parallel structure linked by conjunctions such as *and* and *or*. Examples involve *attack and critiques* (61), *emails or language* (124), *corpora and corpus linguistics* (95), *addressivity and reception* (431).

✧ Modification

A vast majority of the speech as modification co-occurring with the circular gestures are pre-modifiers used to qualify or quantify a prior noun phrase. Many instances are composed of an adjective phrase, which is usually used to modify a noun phrase, such as *CG#architectural# space* (52), *CG#educational# purposes* (123), *CG#quite a few paper#s on health care* (375), *as a kind of CG#receptive# (0.3) theory*, etc.

Another recurring structure used as modification is the __ *of* pattern, and the circular movements are usually completed before the following noun phrases. Examples as such include *certain CG#forms of spoken# language*

(145), *CG#over a period of# time* (304), *a very CG#loose sense of# of democratisation* (240), *And eCG#ven that mode of# working* (507), etc.

✧ Deixis

Those instances of deixis mainly function as a means of clarifying time and space relationships in both physical/concrete and abstract senses with the help of prepositions and pronouns. There are many examples referring to abstract, rather than concrete, locations, possibly due to the academic nature of the NMMC. Examples include *it's CG#all part of the a#cademic experience* (20), *CG#in this chapter I will# be= or in this* (195), *CG#outside of the novel#was* (303), *a lot of the metaphor CG#stuff comes out of (pause)#* (434), to name just a few. However, there are a few instances referring to physical space, these are *the CG#individual within the hotel lo#bby and those is* (197), *the erm researcher CG#will just sit in# the class* (538), *last CG#year and I moved round different grou#ps* (546) and *CG#you inform everybody around you#* (552).

There are also many instances of speech accompanied by circular gestures in this category involving pronouns such as *this, that, these, those, all, something*, and, again, they mostly refer to abstract concepts rather than concrete objects. Examples as such include *getting CG#everything in#to a form where you can work* (128), *many so much of it written up already (pause 0.3) CG#it'll all start# falling into place* (211), *CG#put all your best ide#as like this into* (373), *but I'm sensing CG#reading this a little bit#* (381), *CG#what I was going to include in this one#* (522), *CG#things like table#s et cetera* (485), etc.

The instances with reference to time are the least frequent with only a few instances and the only indicator is the word *then* with a similar patterning. They are *CG#then they're# th= th= can you think of* (467),

CG#you could then t#alk about that (551), and *CG#You you can then# propose different types of* (567).

✧ Dysfluent speech

It seems apparent that those examples of dysfluent speech co-occurring with the circular gestures mostly contain pauses. Some circular gestures may only co-occur with pauses, such as *I I again I'm being careful not to CG#(pause 0.6)# (pause 0.4) just reproduce erm* (340). Some may have two pauses, such as *CG#(pause 0.4) an area that's (pause 0.4) just begi#nning* (313).

In addition to pauses, repetition is another frequently occurring dysfluent feature, such as *CG#and that that would a= again be in#no=* (37), *Erm s= CG#but I'm I'm kind of# (pause 0.2)* (75), *whCG#ich which again fi#ts with what (pause 0.1) you* (186), *well that CG#that's that's a# very familiar use of* (447), *well I mean CG#you can you can# schedule yourself I just want* (549), *CG#You y= y= (pause 0.1)# you make a virtue of* (566), to name just a few. As is presented in the examples, they tend to take place at the beginning of utterances, and the repeated parts aligned with the gestures are mostly the pronouns and the following verbs.

✧ Clauses

Those circular gestures co-occurring with clauses are temporally aligned with the beginning of main or subordinate clauses headed by pronouns (e.g. *you, this, that*), *there, which* and *wh*-clauses (e.g. *how, when, what, where, why*). The circular movements usually end one or two words after the sentence stems are uttered.

Among all the examples, there are 6 instances headed by the structure of *how* ___, such as *it's actually CG#how those# pronouns are used* (170), *no*

matter= (pause 0.2) CG#how frequently# they occur in materials (169) and *Or you CG#know how it's how the mechanics are a#* (229). With a similar structure to *how* ___, there are many instances headed by *wh-*, such as *what* (5 instances), *when* (3 instances), *where* (3 instances), and *why* (2 instances). Such instances include *CG#and what they#* (43), *CG#when I# (pause 0.1) lay on my couch* (528), *I can CG#see where the ell#ctions can be made* (406) and *CG#and why are ho#tels an important issue* (44).

In the clauses headed by *that*, all the five instances emerge with the pattern of *that's* ____, and, similar to *how*, the gestures tend to end instantly or one or two words after *that's*. A typical example is *language CG#that's re#levant to mental health* (445), and *Because CG#that's the ro#ute* (448). The similar pattern also applies to the three examples headed by *this*, where the structure of *this is* ___ emerges, such as *CG#and this is why# (pause 0.2) when you talk to* (42).

✧ Preposition (phrase)

The majority of the instances in this category share a similar patterning, which is constructed of a preposition followed by a noun phrase or a verb in a *–ing* form, such as *aCG#bout sa#lience* (70), *CG#by arti#fice* (77), *CG#for exploring# psychology* (141), *CG#(pause 0.1) to her#* (353) and *CG#with doctors ex#plaining (pause0.6) certain* (540), among many others. Many of the prepositions occur more than once. For instance, *about* ___ occurred five times; *for* ___, four times; *to* ___ and *with* ___, twice.

Other than single prepositions, there are also four examples of prepositional phrases with a similar pattern. They are two instances in the form of *on an/a* ___ *basis*, which are *CG#on an on#going basis* (294) and *CG#on a regular ba#sis* (293). The other two prepositions are *CG#in the same respe#ct* (194) and *CG#of and in the pro#cess of (pause)*.

✧ Discourse marker

Among the small number of instances of speech co-occurring with the circular gestures used as discourse markers, *I mean* is the most frequently occurring with three examples. They are *they don't have the f= CG#er I mean# the hotel is so different because* (126), *Sorry CG#I mean question#naire responses* (178), *but CG#I mean# you know it's worth chasing that* (179). Apart from *I mean*, there are two instances of *do you see what I mean* as discourse markers. Other expressions only occur once, which are *for example* (140), *I mean like I say* (177) and *you know* (555).

✧ Conjunction

Only a few instances co-occurring with the circular gestures are conjunctions, among which *and* is the largest. Examples include *CG#and concre#tisation* (28), *pla CG#ce and lo#cation* (80), *her CG#chapters and what# you went back to before* (83) and *argu CG#ment and# (pause 0.1) talk about some* (250). The other three instances contain *as* (*CG#as a#s you're constructed*, 55), *or* (*linguistic expression CG#n or# or*, 263) and *once* (*CG#once you have a# fixed place to finish*, 195).

5.3.2.2. Semantic preference

✧ On-going action

A considerably large amount of speech co-occurring with circular gestures describes on-going actions and communicates a sense of continuity. These actions can take place in the past, present or the future. In addition, certain verb phrases and grammatical structures tend to recur. For instance, the

suffix of *–ing* paired with various tenses and verbs occur repeatedly. This is rather expected as *–ing* inherently has the meaning of continuity. In particular, there are numerous instances involving the structures of *be going to do* __ or *be doing*, which tend to be frequently used to describe continuous past, present and future actions.

When there is no progressive marker *–ing*, the meaning of continuity is indicated in the verb phrases co-occurring with the circular gestures. These verbs usually describe events or actions that demand a certain period of time to accomplish. Examples include *work* (45), *analyse* (58), *present* (54), *construct* (91), *come* (109), *do* (122), *follow* (139), *go* (155), *write* (173), *discuss* (183), *formulate* (230), *learn* (232), *try* (267), *speak* (430), *explore* (492), *describe* (493), *reproduce* (390), among many others. Many occur more than once (a more detailed analysis will be provided later in this sub-section).

There are quite a number of instances expressing the similar meaning of change and development with verb phrases such as a*dvance* (16), *continue* (93), *change* (116), *progress* (370), *become* (470), *move on to* (260), *transform* (748), etc. A further observation worth mentioning is that a small number of instances suggest the working or action of the mind, such as *read* (i.e. interpret) (74), *indicate* (196), *inform* (237), *believe* (322), *want* (175), *think* (486), *reveal* (497), etc.

It can then be safely concluded that the meaning of on-going process is mainly embedded in different forms of verb phrases. Also, it is somewhat noticeable that certain lexical items can recur in different forms. In attempt to discover those recurrent verbs, all the verbs of different tenses and forms were transformed into their original form, and were alphabetically sorted in Excel. The following Table 5.10 presents the verb phrases that occur at least twice. As is shown in the results, different forms of *go*, particularly *going to*

as an indicator of future tense, and *do* tend to be more frequently occurring than others.

Table 5.10: Number of verb phrases that occur twice or more.

Verb phrases	Number
going to	8
do	7
give	5
become	4
describe	4
go	4
have	4
keep	4
present	4
construct	3
read	3
talk	3
work	3
write	3
advance	2
change	2
come	2
create	2
explore	2
follow	2
have got	2
get into	2
indicate	2

know	2
modify	2
move	2
perform	2
set	2
try	2
understand	2
use	2
want	2

✧ Negation

The speech in the category of negation tends to indicate the meaning using the forms of *(can/be/do)n't/not*, such as *you CG#can't ever ignore# that* (79), *CG#'cos it doesn't ex#ist* (96), *CG#hasn't been formula#ted yet* (165), *CG#it's not made e#xplicit (pause 0.4)* (215) and *CG#that aren't#* (443). Other forms include *there's no* ___ in the examples of *CG#there's no= forecasting there's no=#* (472), *CG#there's no interfering fa#ctor then* (475) and *CG#there's no kind of# democratic (pause 0.4)* (471).

The phrase *not just* occur twice with the examples of *CG#not just the# sub forums* (279) and *CG#So not just# interpretation (pause 0.9)* (419). In addition to *not*, there are other instances of negation that consist of prefix such as *–mis* (*CG#the misus#ing (pause 1) aspect* (457)) and *–in* (*CG#you just invali#date the whole thing* (553)). Also, there is one instance worth mentioning, where the negation is used as a response from the listener in *CG#No be#cause* (275).

5.3.2.3. Semantic prosody

✧ Intensification

The instances of intensification co-occurring with the circular gestures mainly serve the function of modifying an adjective phrase or the whole utterance with intensifiers. Many of the intensifiers repeatedly occur, especially *particular(ly)* and *quite*, both of which have occur five times, such as *CG#it's a particular# kind of location* (212) and *CG#I found it quite#* (174). *Very* and *obviously* occur four and three times respectively, such as *CG#is a very d#istinctive use of the particular* (202) and *CG#and I mean obviously# it's you're doing* (32).

Many are used twice, including *too* (*CG#become too# conscious of it,* 22), *only* (*CG#frameworks are only# ever provisional,* 148, used with very), *just* (*CG#and just# be you know you can you can*, 35), *completely* (*CG#completely# kept away from the class stuff*, 86), *always* (*you will (pause 0.7) CG#as alway#s with these things,* 56), and *all* (*as CG#saying all it# does is help us to understand,* 399). As can be observed from the examples, the circular gestures usually end instantly after the intensifiers.

✧ Compare and contrast

In those instances expressing the meaning of compare and contrast, the most frequently used word is *more*. Instances as such include *vCG#ery much in that more#= (pause 0.4) social* (127), *certain community forums CG#are more=# CG#successful# in= creating=* (53) and *CG#it's much more# repressive* (214).

Other expressions that emerge include *rather than* (e.g. *CG#rather than sort of continuous# prose,* 379), *in comparison to* (e.g. *in CG#comparison to*

the servant#, 85), *different* (e.g. *you have CG#different reali#sations of*, 112), *distinct* (e.g. *really CG#distinct+ <$1> Yeah. <$2> +plan#* 115, 209), *___er* (e.g. *if you liCG#ke higher le#vels of evidence*, 222), *rather* (e.g. *I would have thought CG#rather than sort of continuous# prose*, 379, 401)

✧ Vagueness

Among all the instances of vague language co-occurring with the CGs, *kind of* and *sort of* recur five and four times respectively. Examples of *kind of* include *Anna's CG#kind of# area* (226), *this new= CG#kind of (pause 0.3)#* (227), *it's CG#kind of a mixed mo#de* (225), *which is CG#kind of what those# adjectives are* (228) and *y= CG#you kind of basi#cally map out (pause 0.3) that* (554). As can be observed, the circular gestures tend to start and end in close proximity to *kind of*.

Sort of is another recurrent phrase adopted to express a sense of vagueness in examples such as *CG#a sort of wrap-up con#clusion* (8), *CG#if you could sort of# set it up and* (187), *CG#to read and sort of have a (pause 0.1)#* (494) and *CG#your sort of wor#king hypotheses* (570). Other expressions in this category only occur once, these are *a little* (4), *in general* (190), *likely* (235), *a little bit* (238), *maybe* (314), *seem to be* (407) *and some* (424).

✧ Monitoring mutual ground

Two patterns have been identified for the circular gestures emerging in the context of monitoring mutual ground on the part of the speaker. This is achieved in two ways, i.e. with and without speech. With speech, there are three instances using the sequence of *do you know what I mean* in the instances of 119 (*CG#do you know what I mean. What a#pproaches*), 323

(*possibilities that weren't afforded CG#(pause 0.4) do you know what I mean to writers early#*) and 520 (*do you know CG#what I mean+ <$2> Mm. <$1> +it's what k#ind*).

For the remaining instances, the circular movements mostly occur at the end of a turn, generally aligned with the last few words, followed by a pause and a response from the listener. The pause seems to indicate the intention of checking the understanding of the speakers in this particular discourse context and such a communicative goal is usually decoded and responded to by the listener. To illustrate, an episode of such a context is presented below, although, in this instance, the circular gesture occurs in the middle of an utterance, co-occurring with *CG#reflective (pause 0.3)#***.** Here, the supervisor is talking about the importance of being critical and reflective about one's own research at this stage of PhD study. The 0.3-second pause seems to suggest the supervisor's intention of checking the mutual ground between her and the student as the pause offers a chance for the other party to confirm understanding.

Speech transcript (S02MM):

<$1> You know y= <$E> pause <\$E> in some in some ways too much confidence <$E> pause <\$E> <$=> is a <\$=> is a bad thing here.

<$2> Yes.

<$1> Because you know that kind of <$E> pause <\$E> level of well I'm sorted. It's sorted.+

<$2> Mm.

<$1> +I know exactly where everything is where we're going" <$E> pause <\$E>+

<$2> Mm.

<$1> +wouldn't er suggest to me that you're being critical <$E> pause <\$E>+

<$2> Mm.

<$1> +and and ***CG#reflective (pause 0.3)#****+*

<$2> Mm.

<$1> +about what you+

<$2> Mm.

<$1> +your work.

<$2> Sure.

Instances that belong to this speech pattern include 60 (*CG#assumptions+ <$2> Mm. (pause 0.1)#*), 247 (*CG#makes the space (pause 0.5)# + <$1> Well maybe+*), 282 (*participaCG#nt observation (pause 0.2)# (pause 0.1) or*), 345 (*CG#(pause 0.1) said (pause 0.2). <$F> So# coming*), 388 (*CG#reflective (pause 0.3) + <$2> Mm#. <$1> +about*), 433 (*CG#structure (pause 0.1)+ <$1> Yes. <$2> +and# I*) and 502 (*um pseCG#udotitles (pause 0.2). <$F> Yeah (pause 0.5) yeah#*).

✧ Confirmation

Only three instances emerged in this category, where the circular gestures co-occur with expressions that indicate the meaning of confirmation. They are 130 (*<$M> Yeah. CG#exactly so# that's still knocking around*), 547 (*= erm <$E> Pause <\$E> the photograph (pause 0.2) CG#yes camp# (pause 0.1) I suppose has become erm*) and 548 (*CG#Yes# I've got it <\$O113> with me*).

5.4. Discussion

This discussion will first demonstrate the utility of textual corpus analytical toolkits provided for exploring speech-gesture relationship (5.4.1.), followed

by a description of the multimodal unit of meaning of the circular gestures established based on the current dataset (5.4.2.). The final sub-section 5.4.3. will continue to use the framework developed in the first study to describe the discourse functions of the circular gestures.

5.4.1. Corpus approach and gesture-speech relationship

This study is among the few novel studies (Knight, 2011; Tsuchiya, 2013) which demonstrate how monomodal corpus-based approaches and tools can be invaluable for multimodal research. As the major part of this study examines the temporal and lexico-grammatical relationship between speech and gesture, the concordance lines of the searched item, i.e. the speech components in line with the circular gestures, became a crucial and efficient means of identifying the frequently occurring locations of the circular gestures in discourse and for observing frequent lexical and grammatical structures in each category.

One of the most commonly acknowledged advantages of the corpus method is foregrounding patterns that would not otherwise be captured by mere observation (Adolphs et al., 2004; Atkins & Harvey, 2010; Baker, 2010; Harvey et al., 2008). This is also reflected in the current study. The analysis of the top five words on the concordance list reveals the occurrence of the circular gestures at the beginning or in the middle of utterances. In addition, manual calculation of all the gestures accompanying utterances in the concordance lines suggest that the circular gestures tend to align with multiword expressions, mostly ranging from 2 to 4 words; however, they can also co-occur with very long clauses containing more than 10 words. The average duration of the circular gestures performed by most speakers is 0.8 second, which indicates a universal pattern of use for this particular gesture across speakers.

Concordance analysis also gathers recurrent speech and grammatical patterning, which is of great significance for building a lexico-grammatical analytical framework. Without the assistance of corpus analytical tools, it would be impossible to generate these findings for all the 570 instances. The majority of the 15 categories emerged and coded here have not been discovered by previous researchers, and only two share some similarities to Ladewig's findings (2014b). One is the function of describing on-going events, which also occupies the largest proportion of all instances here. This finding then points to a potential cross-linguistic similarity of speech-gesture association between circular gestures and the meaning of continuity. The other similar pattern shared by the two studies is dysfluent speech identified here, corresponding to the cognitive function (i.e. assisting lexical retrieval) in Ladewig's framework. Nevertheless, the current study has used a much wider variety of non-fluency features in order to code dysfluency (e.g. repetitions, incomplete sentences, etc.), whereas Ladewig only used pauses. It could be argued that using pauses as the only indicator of dysfluency might be an oversimplification. The only function demonstrated in Ladewig's study but has yet to be identified here is encouraging the interlocutor to continue the floor. This function may be a result of the particular task assigned in her research, this being a guessing game.

To summarise, to a large extent, corpus approaches and techniques not only help develop the most comprehensive speech coding scheme for the circular gestures, but also assist further examination of recurrent linguistic features in each category. The findings of the second study also lend support to the conclusions of the first study on the speech-gesture relationship (Fricke, 2013; Ladewig, 2014a; Müller, Bressem, et al., 2013), indicating recurrent speech-gesture associations in multimodal units of meaning. A better understanding of those patterns is crucial for various applications,

especially English language teaching (Steen & Turner, 2013) and artificial intellegence (Bickmore & Cassell, 2005).

5.4.2. The multimodal unit of meaning of the circular gesture

The current study adapted Sinclair's multimodal unit of meaning model to describe all the emerging lexical and grammatical patterns co-occurring with the circular gestures in the concordance lines. In monomodal corpus linguistic studies, the four aspects of the meaning of a lexical item, i.e. collocation, colligation, semantic preference and prosody, is of vital importance for understanding not only the frequent semantic and grammatical patterns of the item, but also its pragmatic functions (Sinclair, 2004). Although differing from their original use, three of the four concepts, with the exception of collocation, have been applied to describe the co-occurring speech patterns of the 570 circular gestures. Despite some challenges regarding assigning some instances, current analysis demonstrates the feasibility of adopting Sinclair's multimodal unit of meaning model in examining recurrent speech-gesture associations.

As a method emphasising the norm of language use, adopting a corpus-based approach requires discovering the relatively frequent and less frequent occurring patterns because not all the meaning variations own the same degree of frequency and significance in real language use (Sinclair, 1991). In other words, characteristic of the corpus method is the clarification of the typical or relatively frequent uses of lexis based on quantitative evidence (Hunston, 2002; McEnery & Wilson, 1996). Relevant to this point, Sinclair (1991) criticises traditional dictionaries and language teaching materials for not providing information of the most frequently used meaning variations of lexical items. Native speaker intuitions can be valuable to some

degree and should not be disregarded, but they are indeed too subjective to draw reliable conclusions.

These concerns of making unreliable assumptions regarding the common uses of lexis can also be applied to the study of the recurrent gesture-speech patterns. Few studies have adopted a corpus-based approach to generalise objective, empirical-based findings of the meaning variations of recurrent gestures, and new evidence may well emerge if we do so. Using the current investigation of the circular gestures as an example, the results reveal that the meaning of this gesture is much more complex than that discovered by previous researchers (Calbris, 2011; Ladewig, 2011, 2014a, 2014b). Corpus results of the 15 speech patterns not only suggest a much more fine-grained profile of meaning variations of the circular gestures, but also the distinctive frequency of occurrence of each variation. That is, which speech patterns tend to co-occur with the circular gestures more often than others.

For instance, in terms of colligation, the circular gestures co-occur with certain grammatical features, with a larger number of noun (phrase) (64), modification (62), deixis (55), dysfluent speech (47), fewer instances of clauses (35) and preposition phrase (17), and only a few examples of discourse marker (8) and conjunction (7). Semantically, the circular gestures have a clear preference for on-going action with a significant number of 172 instances, indicating the meaning of continuity. Only 15 instances align with the meaning of negation. With regards to the semantic prosody of the circular gestures, they are more likely to align with speech patterns that serve pragmatic functions of emphasising and providing noteworthy information such as intensification (39) and compare and contrast (20). Other less frequently occurring patterns in semantic prosody are vagueness (16), monitoring mutual ground (10) and confirmation (3).

5.4.3. The functions of the circular gesture

Previous studies on cyclic gestures (Ladewig, 2011, 2014b) have limited their description of gesture functions to semantics (i.e. describing continuous activities), cognition (i.e. retrieving lexical items) and pragmatics (i.e. encouraging the interlocutor to take the floor). All these functions can be identified in the instances of the current study. For example, a considerable number of instances in the categories of on-going action and deixis made an integral aspect of referential meaning in a multimodal unit, as the circular gestures depict a continuous event or a referred target by deictics. Also, those circular gestures occurring in the dysfluent speech suggest a role of facilitating lexical retrieval (cognitive function) and those in confirming mutual understanding seem to elicit listener response (i.e. pragmatic function).

Nevertheless, this thesis intends to explore and establish a comprehensive framework for describing the discourse functions of the gestures in relation to their co-occurring speech. Based on the datasets of the first study, five discourse functions of the gestures accompanying "(do) you know/see what I mean" were proposed. These are 1) contributing to referential meaning, 2) segmenting utterances, 3) underlining noteworthy information, 4) maintaining coherence and 5) enhancing inter-subjectivity. A more detailed illustration of the definitions has been provided in Chapter 4 (refer back to 4.5.4. for details) and it has to be emphasised that these functions are not exclusive as gestures usually perform more than one function simultaneously. Drawing on this framework, the remaining sub-sections will describe these functions of the circular gestures in sequence.

5.4.3.1. Referential function

It is self-evident that gestures contribute to the semantic aspect of utterances and have a referential function, this has been discussed by many previous researchers (Calbris, 2011; Kendon, 2004; McNeill, 2005). According to the current dataset, a large number of circular gestures also perform referential functions in discourse, which are not necessarily restricted to a certain category of speech pattern. A significant number of circular gestures contribute to the referential aspects of the multimodal utterances, especially those instances aligned with the speech patterns of on-going action, deixis (both concrete and abstract) and modification. Many circular gestures in these categories are iconic of the speech components with which they are associated, expressing an equivalent or complementary sense by the hand. The example below (Figure 5.19) is a deictic gesture aligned with the meaning of 'be a part of' and the area depicted and visualised by the hand is then iconic with the part belonging to that specific group.

Speech transcript:

<$2> And anybody browsing <$H> an <\$H> internet is free to see it so you could+

<$1> Oh right.

<$2> +er potentially just (pause 0.6) ***CG#be a part of that#*** *group as+*

<$1> Yes.

<$2> +just a reader you wouldn't have to+

<$1> Yes.

<$2> +contribute or join.+

<$1> Okay.

CG#be a part of that# group as+ <$1> Yes. <$2> +



1.5 circles, both hands rotating up and down with the right hand moving clockwise and the left anti-clockwise

Figure 5.19: A deictic gesture with a referential function.

5.4.3.2. Segmenting utterances

The function of gestures' segmenting utterances into units is also widely recognised as they predominantly co-occur with the segments of speech with which they are associated (Calbris, 2011; Kendon, 2004; McNeill, 2005). This discourse function is also reflected in circular gestures as they demarcate the sentences into both short and lengthy units. This function can be exemplified by those circular gestures co-occurring with discourse markers at the beginning of a clause, such as *I mean* (instances 126, 178 and 179), *for example* (140), *I mean like I say* (177) and *you know* (555). As is illustrated in the following instance (Figure 5.20), the circular gesture begins and ends exactly as the speaker says 'er I mean', a discourse marker signifying the beginning of a rephrased utterance.

Speech transcript:

<$2> Erm and just the kind of alienation of the individual especially cos there's a whole thing about lobby theory+

<$1> Yes yeah. <$E>Coughs</$E>

<$2> +so especially about the hotel and the character within the lobby+

<$1> Mhm.

<$2> +and then they compare the hotel and especially the lobby to churches and museums and say how that+

<$1> Yeah I remember that stuff from your other work.

*<$2> +churches and museums are so different because they don't have the f= **CG#er I mean#** the hotel is so different because it doesn't have the function that a museum or+*

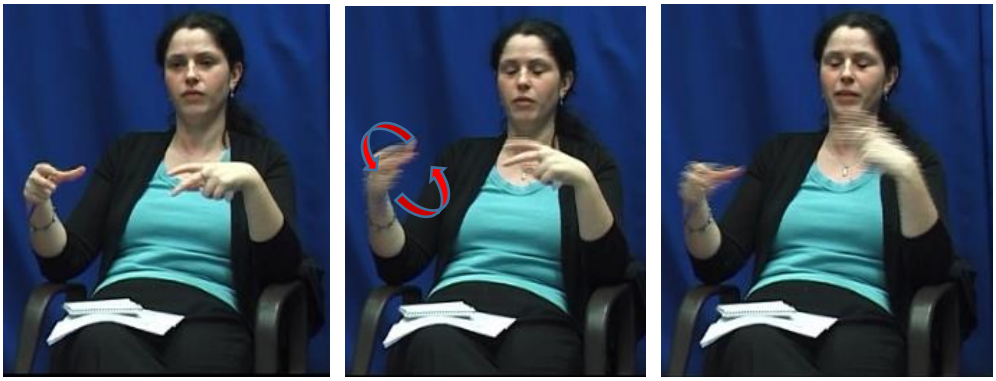
<$1> Right.

<$2> +a church does or a purpose for the individual+

<$1> No.

<$2> +as much so that's how I was kind of applying it to modernism.

CG#er I mean# the hotel is so different because



1 circle, the left hand rotating clockwise

Figure 5.20: A circular gesture with discourse marking function.

5.4.3.3. Underlining noteworthy information

In general, all gestures seem to add a visual aspect to the utterance and play a role in underlining important information (Kendon, 2004; McNeill, 2005). The current study also discovered a similar role in the circular gestures, especially when they co-occur with speech that is unlikely to be referentially

equivalent. Examples might include a large number of instances associated with noun (phrase), intensification, compare and contrast, preposition phrases, and vagueness.

As is demonstrated in the examples below (Figure 5.21 and Figure 5.22), where the speaker performs the circular gestures during the two noun phrases, “city scapes” and “hotel buildings” (see speech transcript below), which specify the concept of modernist space as new and noteworthy information.

Speech transcript:

<$1> Does does the does the material on space theory actually talk about earlier spaces I mean is is that that theoretical body of work simply grown out of modernist spaces do you know what I mean is it very specific to explaining modernist+

<$2> Erm.

*<$1> +spa**CG#ces and you know city scapes#** and **CG#hotel building#**s and.*

<$2> It's really interested in it's written after modernism so it's not+

<$1> Right.

The two gestures (see Figure 5.21 and Figure 5.22 below) as such seem unlikely to formulate any semantic relationship with the two noun phrases. Instead, the role of the two gestures largely resemble the beat gestures in terms of metaphorically ‘circling’ the specific speech components, which are ‘cityscapes’ and ‘hotel buildings’ in this case.

<$2> Erm. <$1> +spa CG#ces and you know cityscapes# and



2, the left hand rotating clockwise

Figure 5.21: A circular gesture co-occurring with a noun phrase.

CG#hotel building#s and.



1, the left hand rotating clockwise

Figure 5.22: A circular gesture co-occurring with a noun phrase.

5.4.3.4. Maintaining discourse coherence

Previously, when gestures co-occur with dysfluent speech such as pauses, they would be regarded as serving the cognitive function of assisting lexical retrieval (Krauss et al., 2000). On cyclic gestures in particular, they have been claimed by Ladewig as assisting lexical searching and performing a crucial cognitive function, as the gestures co-occur with pauses. However,

when interpreted from another point of view, those gestures that occur during pauses can also be seen as part of the multimodal utterances and, therefore, help maintain the fluency of multimodal discourse. That is, as the speech gets interrupted, the gesture continues the flow of conversation.

In the current dataset, circular gestures have also been found co-occurring with pauses, together with many other non-fluency features such as repetition, false starts, etc., which point to the function of keeping the fluency of interaction by adopting the gestural mode of communication. In addition to circular gestures co-occurring during pauses, there are many that align with conjunctions and the onset of main/subordinate clauses led by pronouns such as *you, this, that, there, which* and *wh*-clauses (e.g. *how, when, what, where, why*). Gestures occurring in such a context seem to be encouraging the speaker to continue the utterance and oiling the wheels for their own speech. This function is likely to be vital for maintaining the flow of discourse, particularly in online interactive communication.

The speech transcripts and the figures below (Figure 5.23 and Figure 5.24) demonstrate two instances of this function. One of the circular gestures is mainly aligned with a *it*-clause (*CG#you know it's do with world construction psycho#logisation*) and the other with a conjunction *and* (*CG#and concre#tisation*). The two gestures closely merge with only a fleeting pause between them.

Speech transcript:

> *<$1> <$=> But I think it is worth to <\$=> yeah I think y= how it's received and how it's constructed in in on a reception side. Clearly you have got it he's going to be a poet. Clearly you have got a there's is a linguistic analysis to do I mean it's a cognitive linguistic analysis really* ***CG#you know it's do with world construction psycho#****logisation+*

<$2> Mm.

*<$1> + CG**#and concre#**tisation so but an=and there is a textual basis to that. Erm but having established those you can go and explore which bits trigger which things.*

<$2> Yeah.

The first circular gesture (see Figure 5.23 below) lasts for a relatively long time and the hand rotates for four times, accompanying most of the clause. There seems to be no semantic relationship between the form of this circular gesture and the aligned speech components, but it is suggesting the function of encouraging and maintaining the flow of utterance on the part of the speaker.

CG#you know it's do with world construction psycho#logisation



4 circles, the left hand rotating clockwise

Figure 5.23: A circular gesture co-occurring with an it-clause.

The second circular gesture (see Figure 5.24 below) mainly aligns with a linking word, ‘and’, and the gesture ends instantly at the beginning of the noun, which seems to indicate a similar function.

<$1> +CG#and concre#tisation so but an=and there is



2 circles, the left hand rotating anti-clockwise horizontally

Figure 5.24: A circular gesture co-occurring with conjunction.

5.4.3.5. Enhancing inter-subjectivity

The circular gestures seem also to be serving an interpersonal role in terms of involving the listeners and enhancing inter-subjectivity. The exploration of this function in speech and gesture is the main focus of the first study as the target formulaic sequences "(do) you know/see what I mean" function as a way of checking mutual understanding. The dominant types of gesture phases emerging in the first study are beat/beat-like strokes with(out) long post-stroke holds, and post-stroke holds.

However, there were two circular gestures identified in both studies also co-occurring with the same sequence "do you know what I mean". This suggests the function of the circular gestures as enhancing inter-subjectivity. In addition, the second study discovered more circular gestures performing such a function. This is indicated in the 8 instances where the circular gestures align with the end of utterances with pauses. The circular gestures do not finish until the listener provides a response as confirmation of common ground.

The following example (Figure 5.25) demonstrates such a use as the circular gesture co-occurs with the last word of the speaker's turn, 'assumptions', and the gesture lasts until the interlocutor confirms his understanding by uttering a minimal response token, 'Mm'.

Speech transcript:

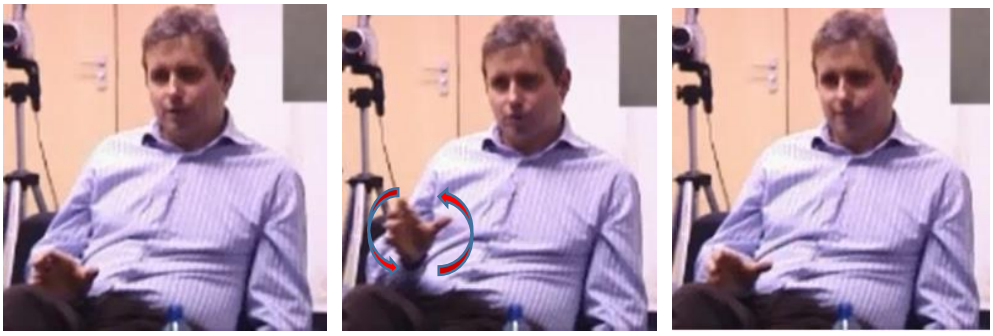
<$1> +a= approach really is bracketing out your own preconceptions and <$E> pause <\$E> your own <$E> pause <\$E> er (pause 1) CG#*assumptions+*

<$2> Mm. (pause 0.1) # (pause 0.4)Mm.

<$1> +prevailing judgments that you've got.

<$2> Sure.

CG#assumptions+ <$2> Mm. (pause 0.1)# (pause 0.4) Mm



2 circles, the right hand rotating clockwise

Figure 5.25: A circular gesture co-occurring with checking understanding.

5.5. Summary

This chapter has demonstrated the methods and development of the multimodal unit of meaning of the circular gestures, depicting a much more complex lexical and grammatical speech patterns associated with this recurrent gesture than previous research. It has also shown how monomodal

corpus approaches and toolkits could be effective in assisting research on the interplay between speech and gesture. This study drew on 570 instances of circular gestures in 8 videos of supervision meetings in the NMMC (100,000 running words). Many instances of circular gestures were identified in each speaker, with a similar length of 0.8 seconds for each gesture. These figures suggested the common use of this gesture across speakers.

All the circular gestures were segmented in ELAN and annotated in plain text with 'CG# #'. Using CG as the search item, the pain text files were processed in AntConc for concordance, clusters and distribution analysis, all of which were effective in detecting the temporal relationship between speech and gesture, and the emerging lexical and grammatical patterns. Three research questions have been addressed, namely 1). What is the temporal relationship between the circular gestures and speech?, 2). What is the unit of meaning of the circular gestures in accordance with their co-occurring speech?, and 3). What are the discourse functions served by the circular gestures?.

The analysis of the temporal relationship between the circular gestures and the co-occurring speech in the concordance list showed the tendency of the gestures to be synchronised with multiword expressions, particularly those of 2 to 4 words. They also often co-occurred with lengthy utterances that contain over 10 words. A further finding of the temporal relationship is that the circular gestures tended to occur at the beginning or the middle of an utterance rather than the end. In addition, there were many instances co-occurring with dysfluent speech.

This chapter has also described the processes of using Sinclair's unit of meaning model in order to establish a speech profile for the circular gestures. Three concepts have been applied here, including colligation, semantic preference, and semantic prosody, which have been used to describe the relationship between the circular gestures and the 15 emerging speech

patterns. In accordance with the frequency analysis of all the patterns, the circular gestures tended to be associated with certain speech patterns more often than others. Moreover, in each pattern, some recurrent lexis and grammatical structures were identified, which further demonstrated the patterned uses of speech with circular gestures.

In terms of the discourse functions of circular gestures, the framework developed in the first study was also applied to the descriptions of the functions of the circular gestures. They include 1) contributing to referential meaning, 2) segmenting utterances, 3) underlining noteworthy information, 4) maintaining coherence and 5) enhancing inter-subjectivity.

The beginning of the chapter also discussed briefly how discourse contexts can have some impact on the forms of circular gestures (e.g. the number of circles, space, etc.). For further studies, with the support of more advanced data capturing, recording and analysing digital technologies, a more fine-grained investigation of the relationship between sub-categories of circular gestures and the associated linguistic patterns can be pursued. That is, to identify the distinctive features of circular gestures in each lexical and grammatical pattern in terms of their circles, size, space and velocity. In order to generate enough instances in each category, a much larger dataset will be necessary for such an investigation.

A final note regards the limited implication of the second study for the general population due to the uneven distribution of the instances across speakers. For instance, one speaker only contributes 10 instances. The study is also limited by the special context of the NMMC. Some emerging lexical and grammatical patterns might be the results of academic speech, which might not emerge in other communicative settings.

Chapter 6. Towards a new understanding of speech-gesture relationship: temporal, meaning and discursive functional perspectives

6.1. Introduction

The goal of this chapter is to use the findings of the two chapters of analysis to demonstrate some of the important implications for the fields of gesture studies. Drawing on a multimodal corpus-based approach (Adolphs & Carter, 2013), the thesis attempts to work towards the notion of the multimodal units of meaning by establishing a speech-gesture profile for a recurrent multiword expression and a recurrent gesture. The two empirical chapters of the current thesis, i.e. Chapter 4 and 5, documented the theories and research paradigms relating to the building of each speech-gesture profile as well as the specific approaches, findings and implications. Both draw on the Nottingham Multimodal Corpus (NMMC), Chapter 4 focused on establishing a multimodal profile for a group of target formulaic sequences (TFSs), "(do) you know/see what I mean", as an example of recurrent speech, and Chapter 5 concentrated on the profile for a recurrent gesture, the circular gesture.

Both studies addressed three research questions:

1). What is the temporal relationship between speech and gesture?

2). What is the coordination in meaning between speech and gesture?

3). What are the discourse functions of the gestures in a discourse context?

As the analysis and discussion in Chapter 4 and 5 consistently revolved around the multimodal units of meaning of the TFSs and the circular gestures, other important implications for previous theories and debates within the field

of gesture studies await discussion. This is, therefore, the primary intention of the current chapter. The discussion of those implications will be structured in parallel with the research questions listed above.

Relating to the first research question, sub-section 6.2 will offer some new insights on the temporal relationship between speech and gesture. Regarding this issue, Chapter 4 focused the discussion on the patterns of gesture phases (e.g. preparation phases, stroke phases, etc.) co-occurring with the TFSs, and the temporal alignment between them. The general conclusion was that the TFSs and the co-occurring gesture phases are mostly aligned, and the salient pattern is the temporal coordination between TFSs checking understanding and the stroke phases. Chapter 5 focused on the length of words co-occurring with the circular gestures, which led to the conclusion that the circular gestures tend to co-occur with multiword expressions of various lexical-grammatical patterns. Additionally, the circular gestures may sometimes co-occur with lengthy sentences. However, the analyses of both studies also fill the void in the area of exploring the temporal relationship between non-iconic gestures (i.e. metaphoric and beat/beat-like gestures) and the speech co-occurring with them. That is, whereas previous gesture research drew on iconic gestures (i.e. gestures describing concrete objects and actions) and their corresponding speech to explore the temporal relationship between speech and gesture (Beattie & Aboudan, 1994; Chui, 2005a; Church, Kelly, & Holcombe, 2013; Leonard & Cummins, 2011), the metaphoric and beat/beat-like gestures emerging here can shed some meaningful light on the alignment between non-iconic gestures and speech. This is a somewhat under-researched area in gesture studies and the new methodological and theoretical insights offered here might expand our knowledge in this matter.

Following 6.2., 6.3. will present a discussion of the meaning coordination between speech and gesture, which constitutes the second research question of the thesis. Chapter 4 concentrated on illustrating the co-occurrence between

the meaning variations of the TFSs and gesture forms (e.g. TFSs checking understanding co-occurring with beat/beat-like gestures and post-stroke hold phases), which formed the basis for establishing a speech-gesture profile for the TFSs. Similarly, the discussion in Chapter 5 mainly concerned the co-occurrence between the circular gestures and the emerging lexical and grammatical patterns (e.g. some circular gestures co-occurring with speech indicating the meaning of on-going actions). Such a focus, again, serves the purpose of constructing a speech-gesture profile for circular gestures. Nevertheless, as most of the emerging gesture types in the two studies are metaphoric in nature, their association with certain speech patterns/meanings also builds our understanding of some important issues relating to gesture, metaphor and cognition (Chui, 2012; Cienki, 2013; Cienki & Müller, 2008b). These implications were not been examined in detail in Chapter 4 or 5, and will be discussed in 6.3.

6.4. will focus on the implications for the functions of gesture. Based on the datasets of both studies, several discourse functions of gestures were discussed separately in Chapter 4 and 5, including 1) contributing to referential meaning, 2) segmenting utterances, 3) underlying noteworthy information, 4) maintaining coherence and 5) enhancing inter-subjectivity. However, the findings of the current thesis also contribute to the debate concerning two of the major functions of gestures in the existing literature: are gestures mainly performed for speech production or communication? In addition, the current thesis also adds new knowledge to the research on a special category named ‘interactive gestures’ (Bavelas et al., 1992). A detailed demonstration of these contributions will be provided in 6.4., which will be followed, in 6.5, by a summary of the main contents of this chapter.

6.2. Temporal relationship between speech and gesture

This section discusses some new implications for one of the most widely investigated issues in the field of gesture studies, relating to the temporal relationship between speech and gesture. It begins with sub-section 6.2.1, which demonstrates some new methodological insights and empirical findings regarding where the stroke phases of non-iconic gestures tend to start and end in relation to their associated speech. These insights then expand our understanding of the same issue drawing on iconic gestures and their associated speech. The second sub-section (6.2.2) discusses some new discoveries regarding the position of some recurrent gestures in utterances in line with their meanings/functions. The discussion will be based mostly on two emerging recurrent movements, holding and the circular gestures, with the former seemingly occurring towards the end of an utterance and the latter at the beginning and in the middle.

6.2.1. TFSs and non-iconic gestures

Both studies of the current thesis began by examining the temporal relationship between gesture and speech, which is an essential analytic procedure for any multimodal study involving the two modalities, though with different levels of granularity. It has been generally accepted that the strokes of the gesture, i.e. the most laborious part of a gesture, tend to occur in proximity with the segments of speech that bear the closest semantic relationship with them; the preparation phases come preceding the associated speech and the retraction phases afterwards (Kendon, 2004; McNeill, 2005). Schegloff (1984) targeted the exact word(s) to which a gesture is most closely related in meaning and termed it/them the lexical affiliate of the gesture. Meanwhile, he observed that the gestures tended to precede their lexical

affiliates; however, it was not clear to which part of the gestures, preparation phases or strokes, he was referring.

As pointed out in the literature review, previous claims of the temporal relationship between speech and gesture was mainly made based on observations of iconic gestures and their co-occurring strokes. As a type of gesture initially proposed by McNeill (1992), iconic gestures refer to those hand movements that depict objects, outline shapes and enact actions. Compared to metaphoric, beat and deictic gestures (refer back to Chapter 2, 2.4.2. for definitions), the lexical affiliates of the iconic gestures, i.e. gestures describing concrete actions and objects, tend to be more obvious. Despite the general agreement that gestures and their lexical affiliates tend to be close to each other temporally, the exact alignment of their start and end points remains a controversial issue and lacks empirical evidence.

Only a few researchers (Chui, 2005a; Ferré, 2010) provided detailed and thorough data regarding the exact temporal coordination between the onsets and/or ends of the strokes and their lexical affiliates. Based on single referential words, Chui's (2005a) study of 303 iconic gestures from Chinese speakers found out that a substantial number of the strokes (60.1%) were synchronised with their related speech without onsets. Whereas 35.6% of strokes preceded the speech, the number of the strokes following the associated speech components was only slight with 13 instances (4.3%). Drawing on a French corpus and iconic gestures, Ferré (2010) conducted a specific analysis of both the onset and end of 244 strokes in relation to their lexical affiliates with different results being generated: 72% of the gesture strokes began before their associated speech and 28% of them after and ending slightly after them. Ferré (2010) also found out that the strokes tended to finish after the lexical affiliates (87%) rather than before.

It is noticeable that there was no instance reported in Ferré's (2010) research of the stroke phase co-occurring with the lexical affiliate. This is in

great contrast with Chui's results (2005a), who found out 60% of the strokes co-occurring with their lexical affiliates. Such a difference may be a result of the distinctive language under investigation, with Chui's research on Chinese and Ferré's on French. Nevertheless, it is also likely that this difference is a result of the different gesture annotation approaches adopted. Ferré (2010) admitted that he annotated the start and end points of the strokes based on the frame-by-frame movements in the software rather than pinpointing exactly where the strokes start and end. This imprecise annotation of the gestures can lead to misleading results.

Another alternative explanation may relate to the segmentation of speech. As speech is made up of a string of closely linked sound, at times it is difficult to segment the syllables. Hence, different ways of segmenting the syllables may lead to differing analysis results in terms of where speech and gesture co-occur. For instance, when segmenting speech in English, once the stroke of a 'throwing' gesture in a sentence such as "He threw that out" has been annotated, the researcher will need to listen to the speech and decide the segment of speech that co-occurs with the stroke. As the sound of [i:] in "He" is usually closely linked to the sound of [θ] in "throw", one researcher may decide the speech starts from the vowel [i:] in "He", which precedes the lexical affiliate, whereas the other may code it as starting from "throw", which is precisely the start of the lexical affiliate. Hence, it is likely that a stroke that could be coded as synchronised with the lexical affiliate by one researcher might be coded as preceding the lexical affiliates by others. Similarly, one stroke coded as preceding its lexical affiliate by one person might be coded differently by others. However, regardless of the issues in analysis and the future need for more precise tools for speech segmentation, it can be safely concluded from the existing research that the starting points of the strokes tend to either precede or co-occur with their lexical affiliates rather than afterwards.

In addition, the temporal exploration concerning where the stroke ends in relation to speech is somewhat under-researched.

As mentioned earlier, previous research on the temporal relationship between speech and gesture predominantly drew on iconic gestures and their lexical affiliates owing to the more salient semantic relationship between the movements and the word(s) delivering the same meaning in composition with non-iconic gestures such as beats, metaphoric and pointing gestures (Chui, 2005a; McNeill, 2005). For beat gestures, they maintain the same movement of beating in various directions (e.g. up-down, left-right, in-out) and do not have a substantial associated referential meaning. Therefore, strictly speaking, the beat gestures do not have a lexical affiliate; however, there is a whole area of research that focuses on the temporal relationship between beat gestures and stressed syllables (Leonard & Cummins, 2011). Similar to the beat gestures, deictic gestures are also relatively stable in form, being mostly composed of a pointing movement. The speech co-occurring with them might contain some deictic expressions such as 'this', 'that' and 'there', etc., but also a variety of other lexis. It seems possible to identify the lexical affiliates of the deictic gestures, but, until the present, few corpus-based investigations of the temporal relationship between the deictics and the co-occurring speech has been conducted.

Differently from beat and deictic gestures, metaphoric gestures are also iconic in essence, but they are iconic of something abstract (McNeill, 1992) and their lexical affiliates can be difficult to identify. This is the reason provided by Ferré (2010) in his study, explaining why he and many other researchers chose iconic gestures as their target. He provided an example of a Palm Up Open Hand metaphoric gesture embodying the meaning of presenting/offering co-occurring with an utterance, "we hadn't spoken about it again", and it seems impossible to decide the word(s) with which this gesture is associated.

Similarly, Harrison's (2009, 2013) research on the temporal relationship

between the Palm Down Open Hand gesture (with the core meaning of negation) and the grammatical expressions of negation generated a rather complex hierarchical model. His research shows that the Palm Down Open Hand gestures did not necessarily co-occur with *no* or *not*, but with other words in the utterances depending on the goal and context of communication. In addition, a brief glance at Ladewig's (2014b) examples of the cyclic gestures (i.e. rotational movements) and the second study on the circular gestures in this thesis reveals that they co-occur with a wide range of expressions and it is really difficult to decide on their lexical affiliates.

As indicated in the above discussion, there is still a void in terms of the temporal relationship between different types of gestures and their associated speech, and the first study of the current thesis can provide some new methodological insight in terms of examining where metaphoric gestures co-occur with their lexical affiliates. As the first study selected a group of formulaic sequences (Kuiper, 2004; Wray, 2008) with an abstract meaning, "(do) you know/see what I mean", that can be seen as a stable lexical affiliate, all the gestures that have been identified co-occurring with them are metaphoric. They are beat/beat-like gestures, Palm Up Open Hand gestures, bounded-space gestures, circular gestures and throwing away gestures.

Although drawing on a very limited number of instances (31 examples), the analysis of the temporal relationship between the TFSs and the non-iconic gestures is largely consistent with the previous findings based on iconic gestures (Chui, 2005b; McNeill, 1992, 2005). Firstly, echoing Kendon (2004) and McNeill (1992), etc., the strokes of the 31 non-iconic gestures and the TFSs, "(do) you know/see what I mean", are predominantly synchronised. Second, in terms of the start points of the strokes, similar to previous findings (Chui, 2005a; Ferré, 2010; Kendon, 2004; McNeill, 2005), the strokes of the non-iconic gestures in the first study tend to emerge either preceding the speech or during rather than following. Almost half of the strokes (14 out of 31)

have their onset beginning slightly before the TFSs and the other half starting simultaneously with the onset of the TFSs.

In the majority of previous research concentrating on the temporal coordination between the starting point of stroke and speech, there is a general lack of attention as to where the strokes end in relation to speech. The first study here also has some implications regarding this issue. The data show that the end of the stroke phases tends to occur either slightly before the end of the TFSs or at the same time. However, the non-iconic gestures do not end after the TFSs. This result slightly differs from Ferré's (2010) study on iconic gestures where it was found that the end of the strokes tends to continue slightly after the lexical affiliates. The difference might result from the different types of gestures under study, i.e. iconic versus non-iconic gesture, or to the specific relationship between this particular group of TFSs and their co-occurring gestures. An alternative explanation would be the different annotation approaches taken by Ferré's (2010) study and the current PhD. That is, as mentioned earlier, Ferré annotated the end points based on the frame-by-frame changes of the hand, whereas this study observed and annotated the actual end points of the strokes. As each second contains 24 frames, the end point based on frames may not be the actual end point of a stroke. To confirm this, more research based on larger datasets is undoubtedly required.

The research on the temporal relationship between the TFSs and gestures indicates that, although determining the lexical affiliates of the metaphoric gestures can be difficult, it is methodologically possible to do so. To identifying metaphoric gestures, functional formulaic expressions such as "(do) you know/see what I mean" can be a good starting point. The same method might be applied to other formulaic expressions such as "you know", "I mean" and "sort of".

6.2.2. The position of recurrent gestures

As previous gesture studies have mainly focused on the alignment between recurrent gestures and their co-occurring meanings at the lexical and sentence level (Ladewig, 2014a), there is a lack of understanding in terms of the position of those gestures at the discourse level according to the meanings/functions. That is, it is possible that different recurrent gestures might tend to occur at the beginning, middle or the end of utterances, which might relate to their distinctive meanings/functions. This is important because it can further our knowledge of the uses of those gestures at the discourse level.

Concerning this issue, Calbris (2011) briefly demonstrated how the gestures placed at the beginning of utterances could predict the attitude of the speaker to the following utterance. For example, before a speaker provides a negative response, the gesture of waving the hand or shaking the head can already anticipate the negative prosody in the following utterance. Streeck (2009a) describes how a hand gesture at the beginning of a turn could perform a speech act such as requesting or rejecting, which he terms forward-gesturing. Calbris and Streeck's viewpoint seem consistent with some casual observations of communication in real life, where head gestures such as head nodding/shaking are likely to occur at the beginning or middle of utterances, predicting a positive/negative response. It could also be assumed that, based on the function of beat gestures as underlining noteworthy information, they might be more likely to occur in the middle rather than at the very beginning or end of utterances. However, empirical evidence for these observations is still required.

Those aforementioned observations indeed raise the question of whether certain types of gestures tend to emerge in certain places in the utterance according to the meanings/functions. The current study can probably

contribute to this. For instance, the analysis of the post-stroke hold phases co-occurring with "(do) you see/know what I mean" at the end of the utterances indicates that holding might tend to occur at the end of a turn, primarily serving the function of checking the understanding of the listener. Also, an extra analysis of a five-minute episode of all the post-stroke holding leads to the result that relatively lengthy holdings with an average duration of 0.962 tend to occur approaching the end of an utterance, and they always trigger some response from the speakers (refer to Chapter 4, 4.5.2.1. for details).

Although the data could be much larger, the empirical evidence here suggests that holding tends to occur at the end of utterances with the main function of eliciting a response. This is consistent with previous research on conventionalised quotable gestures where the speakers held their hand to prolong their intention until responses were provided by their interlocutors (Brookes, 2005; Kendon, 1995). Hence, instead of regarding all the holdings as dependent post-stroke hold phases that are used to help the stroke phases to achieve gesture-speech alignment (Kendon, 2004), they might be playing a more important role in communication in terms of enhancing interpersonal communication, especially for long holdings. More research based on larger datasets is required in order to examine the threshold for the duration of the holdings for them to perform such a function.

Whereas holding seems to emerge at the end of utterances, serving the function of checking the mutual ground and/or prolonging the attitude/intention of the speaker, the circular gestures in the second study predominantly occur at the beginning or in the middle of the utterances. This finding is mainly based on the analysis that, except for monitoring mutual ground and confirmation, the co-occurring speech patterns (e.g. modification, deixis, preposition phase discourse marker, vague language etc.) mostly emerge at the beginning and middle of the utterances. Although circular

gestures are multifunctional, those circular gestures positioned at the beginning of utterances primarily seem to play an important role of segmenting discourses (e.g. occurring with discourse markers and at the beginning of clauses, etc.) and maintaining the flow of talk (e.g. co-occurring with conjunctions). Those in the middle of the utterances are likely to act as emphasising noteworthy information (e.g. co-occurring with noun phrases, intensification, compare and contrast, etc.) and expressing referential meaning (e.g. describing on-going actions, deixis, etc.). Similar to the post-stroke hold phases in the first study, the circular gestures can also occur approaching the end of the utterances followed by a pause in an attempt to elicit responses from the other party; however, the number is only minimal (less than 2%). These findings reveal that, in accordance with certain meanings of the circular gestures, they tend to occur at the beginning or in the middle of utterances rather than at the end.

Overall, it seems that the positions of the gestures emerging in discourse are, to some extent, related to the pragmatics of that gesture. As demonstrating and interpreting attitudes appropriately during a certain point in communication is a crucial element for establishing and maintaining interpersonal relationships, further research on this issue would undeniably be invaluable. In addition, as demonstrated by the second study, corpus tools can be somewhat efficient in automatically generating a macro picture that captures the dominant positions when a recurrent gesture occur (refer back to Chapter 5, 5.2.4. for details).

6.3. Meaning coordination between speech and gesture and metaphor

The contents of this section relate to the second research question concerning meaning coordination between speech and gesture. Whereas the co-occurrences of the variations of speech and gesture have already been

presented in detail in Chapter 4 and 5, the reasons for these co-occurrences, i.e. why they co-occur, have yet been explored. In view of the metaphoric nature of the emerging gestures, the theories from the established line of research on speech, gesture and metaphor (Lakoff & Johnson, 2008; Müller, 2009) seem relevant. Hence, the sub-section 6.3.1 will illustrate how Conceptual Metaphor Theory (Lakoff & Johnson, 2008) and the notion of image schema (Johnson, 1987) in cognitive linguistics are particularly useful explanatory tools for interpreting the emerging gesture-speech mappings in both studies. In addition, 6.3.2 further explains the role each gesture parameter (i.e. the trajectory, space, speed, etc. of a gesture) played in the relationship between speech, gesture and metaphor, followed by a brief description of how gesture provides a further viewpoint on cognition.

6.3.1. Conceptual Metaphor Theory as an explanatory tool

The study of metaphor has been revolutionized methodologically and theoretically since being removed from literary texts and introduced into the field of cognitive linguistics (Lakoff & Johnson, 2008). Conceptual Metaphor Theory holds that metaphor is the cognitive process of understanding one thing/the source domain out of another/the target domain, and such a metaphoric mapping is deeply rooted in our repetitive bodily experiences in the physical world (Chui, 2011; Cienki, 2008). In other words, our bodies, including our hands, have experienced recurring patterns of physical movements, object handling and concept comprehension. These routinized motor-sensory patterns then become image schema, which are crucial for abstract and metaphoric conceptualisation and reasoning (Lakoff & Johnson, 2008). As our physical experiences can be somewhat culturally specific, research has shown some cultural and linguistic differences in metaphorical conceptualisation (Gu et al., 2017; Núñez & Sweetser, 2006).

Due to the particularly significant status of the hand in interacting with the physical world (Streeck, 2009b), it is only natural that Conceptual Metaphor Theory and image schema have been further pursued in gestures. Numerous studies have been carried out and have revealed the representation of metaphors in gesture (Cienki & Müller, 2008b), providing new evidence for metaphoric thinking and its root in recurrent bodily experiences. A good illustration would be the PAST-IS-LEFT, FUTURE-IS-RIGHT metaphoric link, which is revealed when people moving their hands from left to right when saying something in the past refer to the future. This analogic association may well be embodied in English speakers' daily experiences of writing from left to right. There is also a growing trend of exploring the multimodal nature of metaphor, i.e. metaphors expressed in different modes in both speech and gesture and beyond (Forceville & Urios-Aparisi, 2009; Müller & Cienki, 2009).

For the current thesis, owing to the metaphoric nature of the emerging gestures, metaphoric conceptualisation is a useful tool for justifying the association between variations of gesture forms and meanings across contexts (Calbris, 2011; Ladewig, 2011; McNeill, 2005; Müller & Cienki, 2009). The results of the first study show that when the TFSs have the meaning of checking understanding in discourse, they are dominantly accompanied by beat/beat-like gestures and holding. Beat gestures are one of the most frequently occurring hand movements in daily life and they are perceived as metaphoric in terms of projecting the hand's action of pressing or marking something important, the notion of emphasising (Calbris, 2011). In other words, cognitively speaking, the function of underlining something important and noteworthy in the beat gestures may be entrenched owing to the repetitious real-life experience of the hand as it presses or underlines something important with some objects such as a pen or a knife (Calbris, 2011). The notion of significance embodied in the beat gestures are well

coordinated with the meaning of checking understanding in "(do) you know/see what I mean". As the speakers use speech to directly ask whether mutual understanding has been reached between the two parties, they use gesture to express the importance of such an understanding.

The same approach can be applied to holding. Although holdings were originally coded as post-stroke hold phases, their high co-occurrence with TFSs checking understanding and further pilot analysis confirmed their association with the function of waiting for responses in interaction. Such a use seems to be derived from our bodies' habit of holding still while waiting for something. Therefore, there might be a high possibility that holding is unwittingly performing a vital interpersonal role in interaction, and this is worthy of further exploration.

There are three gestures co-occurring with the current TFSs which embody the same metaphor as IDEAS ARE OBJECTS, these are Palm Up Open Hand gesture, the bounded-space gesture and the fling/throwing away gesture. They are all recurrent gestures that can be frequently observed in real-life communication. Four Palm Up Open Hand gestures were found to be performed by one speaker (Supervisor 3), evenly distributed in the contexts of TFSs checking understanding and as discourse markers. Produced by the same speaker, the only one bounded-space metaphoric gesture emerges with a TFS as a discourse marker. These two gestures embody the metaphor of IDEAS ARE OBJECTS that can be presented on the surface of an open hand (PUOH gesture) or held between two hands (bounded-space gesture) in front of the speaker. Both forms were categorised as conduit metaphoric gestures by McNeill (1992) with the function of presenting abstract ideas, and such a function might be schematised from the recurring physical experiences of presenting concrete objects by the hand in daily life. This notion is somewhat compatible with the TFSs, which intends to draw the hearer's attention to the presented ideas, though with different levels of mutual understanding.

The fling gesture, or throwing away gesture (Müller, Bressem, et al., 2013), has been found in the current study and in previous ones as having the meaning of abandoning or terminating something. Here, the two TFSs co-occurring with the fling gestures are both used as a way of claiming the speaker's confidence of mutual understanding by giving up self-repair. The fling away movement then embodies the metaphor of IDEAS ARE OBJECTS that can be thrown away. Such a metaphoric use of the fling gesture is very likely cognitively motivated by the commonplace physical experiences of throwing away something of no value (Müller, Ladewig, & Bressem, 2013).

With regards to the second study on the circular gestures, consistent with the previous research, the circular gestures tend to occur in those contexts where the meaning of continuous activities or actions is implied (Ladewig, 2014b). This meaning is most likely projected from the image schema of CYCLE as the core formational element of this recurrent gesture, which tends to be conceptualised from recurring bodily experiences of cyclic events such as the rolling wheels, the cycle of the seasons, the actions of cranking, etc. (Johnson, 1987; Ladewig, 2011). All these circular experiences in the real world can be metaphorically mapped onto abstract concepts such as the cranking of the mind, the continuity of time and the on-going processes of an activity.

In addition to speech that already has the meaning of continuity, similar metaphoric meaning association can also be established between the circular gestures and many other emerging speech patterns. For instance, their co-occurrence with dysfluent speech seems to symbolise the cranking of the mind while the speaker is searching for utterances. Also, when the circular gestures are synchronised with discourse linking devices such as clause stems (e.g. *there's*), prepositions (e.g. *and*), discourse makers (e.g. *I mean*), conjunctions, monitoring mutual ground (e.g. *you know what I mean*), and

confirmation (e.g. *yes*), the continuous rotational movements seem to suggest the development of conversation as an on-going process.

However, the meaning of continuity in the circular gestures seems not as coordinated with the meanings delivered by some other speech patterns such as noun phases, modification (quantifiers and qualifiers), negation, intensification, compare and contrast, vague languages. Detailed observation of those instances seems to indicate another possible metaphoric meaning of the circular gestures similar to emphasising function of the beat/beat-like gestures. The speakers seem to be using the circular gestures to emphasise that information in line with them. Such a metaphoric meaning of the circular gestures may stem from the hand's experience of underlining noteworthy information with circular lines while reading, or drawing a single circle outside a target object (Calbris, 2011).

6.3.2. Gesture parameters and their metaphoric links

There have been two seemingly contrasting views on the meaning of gestures. One suggests that the constituents of the gestures do not contribute to the meaning of the gesture, and the entire gesture gestalt (i.e. the entire gesture including its components) determines the meaning of each element (i.e. handshape, trajectory, movement, velocity, space) (McNeill, 1992). By contrast, the others maintain that variations of gesture parameters (i.e. handshape, trajectory, movement, velocity, space) of a certain gesture can contribute to meaning variations of the gestures (Calbris, 2011; Ladewig, 2014a, 2014b; Müller, Bressem, et al., 2013). Each parameter makes up a metaphoric link between gesture and meaning, and the parameter(s) activated each time is/are largely dependent on context (Calbris, 2011).

The results of the current study largely support the second standpoint. For instance, when co-occurring with the TFSs, the movement of beating of the

beat/beat-like gestures is the most relevant parameter for the metaphoric meaning of emphasis rather than the handshapes and movement directions, etc. Holding can also have different handshapes and space, etc., but the element of holding still maintains the most stable and relevant analogic link to the metaphoric meaning of waiting for confirmation of understanding. Similarly, for the PUOH and bounded-space gestures, it is the upper space supported by the hand and the bounded space held between the hands which activates the metaphor of IDEAS ARE OBJECTS. For the fling gestures and circular gestures, it is the throwing away and rotational movements that symbolise the core meanings of abandoning and continuity/emphasis, respectively; other parameters such as the velocity, handshape and space etc. can vary to a various degree depending on the contexts of use (refer back to Chapter 6 for details).

In addition, as can be observed from the discussion, the representation of the aforementioned metaphors can only lie in the gestures rather than speech, which then supports the argument that metaphor is not lexical, but modal independent (Cienki & Müller, 2008b; Forceville & Urios-Aparisi, 2009). Principally, metaphor can be represented in any form of sign, including speech, image, gesture and even intonation (Müller & Cienki, 2009). For instance, as the PUOH gesture embodies the metaphor of IDEAS ARE OBJECTS, the TFSs co-occurring with it does not contain any metaphor. This also applies to the bounded-space gesture, the fling gesture, and the circular gesture.

Hence, research on speech, gesture and metaphor provides us with concrete evidence that metaphoric thinking, even cognition in general, may well be rooted in recurring physical experiences (Lakoff & Johnson, 2008). Monomodal textual research on conceptual metaphor and image schema can only provide us with partial knowledge of metaphor in language. Indeed, the visual representation of metaphoric thinking in gesture and image seems more straightforward than speech, as human bodies are more adept at metonymically/iconically (Mittelberg, 2014; Mittelberg & Waugh, 2009,

2014), spatially and visually representing the chosen elements of the corresponding image/the source domain in the physical world. This is something that could not be provided by speech (Cienki, 2008).

Overall, the thesis further confirms the value of understanding metaphoric links between speech and gesture while interpreting the meaning of the gestures in certain discourse contexts. This is especially true when a researcher is targeting abstract expressions such as the current TFSs and metaphoric gestures. The benefits of gesture and cognitive linguistics are mutual. Whereas gesture provides another window on the human cognition, theories in cognitive linguistics can also be of great value to gesture analysis. There is perhaps no other area in linguistics that embraces gesture as fully as cognitive linguistics (Cienki, 2016). It can be assumed that gestures will play an increasingly significant role in cognitive research within and beyond the field of linguistics.

6.4. Functions of gestures

This section will focus on demonstrating how the current thesis contributes to the exploration of gesture functions. As Chapter 4 and 5 have illustrated the discourse functions of the gestures, the discussion here relates to the existing debate revolving around two major functions of gestures: lexical retrieval and communication. Instead of one function overriding the other, the current thesis supports the increasingly accepted view that gestures perform both productive and communicative functions (Holler & Wilkin, 2011; Kita, 2000; McNeill, 2005). Based on the results of the thesis, 6.4.1. will demonstrate that, although gestures may be playing a role in speech production, it is by no means the pre-dominant function. This will be followed by a discussion in 6.4.2. regarding the ways in which gestures contain information that is absent in speech and serve a role in communication. As previous research on the

communicative value of gestures mainly drew on experimental data (this will be reviewed briefly), the natural communication in supervision meetings offers new evidence on how the common ground between the interlocutors can affect the performance of gestures. Sub-section 6.4.3. concentrates on some new findings of the thesis that expand our understanding of a particular type of gesture investigated by a group of researchers, namely, interactive gestures, i.e. gestures that mainly appear in interaction. The final part of this section will provide a summary of the new discourse-based functional framework proposed by the current thesis, which drew on more detailed discussions in both Chapter 4 and 5.

6.4.1. For speech production

Some researchers have claimed that the primary function of gestures was to assist internal lexical retrieval and speech production, mainly based on the high occurrence of iconic gestures around the hesitation pauses and during the pauses immediately before their associated speech (Beattie & Aboudan, 1994; Butterworth & Beattie, 1978; Hadar & Butterworth, 1997). They have argued that the communicative function of the gestures was overrated and even negligible. Without denying the possible role of gestures in speech production, the results of the current thesis indicate that assisting lexical retrieval may not be the primary function of gesture. Examining the stroke phases of all the gestures co-occurring with the TFSs, only a few occur next to dysfluent speech with the majority occurring in fluent speech. For the circular gestures, only 47 out of 570 instances co-occur with dysfluent speech and most occur in fluent speech. It seems plausible to assume the cognitive role played by the gestures in view of their occurrence in the neighbourhood of dysfluent speech, such as the circular gestures do here. However, it is premature to claim that

assisting speech production is the primary function and overrides the communicative value of gestures.

Moreover, the lexical retrieval hypothesis scarcely seems to hold water due to its inability to account for some phenomena generating from patients with speaking disorders. For instance, if the main function of gestures were to facilitate speaking, it should occur more frequently among people with speaking disorders such as stuttering (excessive repeating and prolonging of sounds causing speech disruption) (Mayberry & Jaques, 2000) and aphasia (McNeill, 1985, 2005). However, gestures were disrupted with their associated speech components during stuttering, indicating an extremely high level of speech-gesture synchrony and co-expressiveness (Mayberry & Jaques, 2000). Similarly, different types of aphasia were found to lead to the disappearance of corresponding gestures (McNeill, 1985, 2005). As Broca's aphasics can produce grammatically disconnected (e.g. lack of referring expressions) but semantically rich speech, they can produce complex iconic gestures but not beat gestures. In contrast, Wernicke-type aphasics' disruption of producing semantically connected contents but maintaining word sequences led to numerous beat gestures but not referential gestures. Research on the coordination of gesture and speech in stuttering and aphasia patients not only disproved lexical retrieval as the main purpose of gesturing, but also served as a strong evidence of a deep neurotic synchrony between speech and gesture.

In addition to the lexical retrieval hypothesis, there may be other reasons for the occurrence of gestures ahead of their relevant speech. One justification might lie in the type of gestures they examined and the special use of gestures for expression. That is, as most previous research (Beattie & Aboudan, 1994; Butterworth & Beattie, 1978; Hadar & Butterworth, 1997) used iconic gestures as a demonstration (i.e. concrete gestures depicting objects/motions), it might be easier for gestures to mimic physical appearance/movements than

speech. Speech needs to be semantically and syntactically analysed, sequenced and produced (Kendon, 1980), the process of which tends to take more time than gesture production. Gestures' imagery, motor-sensory thinking method makes them more efficient in encoding physical objects and actions than speech (Kita, 2000; McNeill, 2005). For this reason, gestures can be executed more quickly than speech.

An even more straightforward explanation is that, as the strokes of the gestures aim to be temporally synchronised with their lexical affiliates, sometimes the preparation phases or the stroke phases have to begin earlier in attempt to perform complicated movements (Kendon, 2004). This is especially true of those large iconic gestures. A further reason relates to the communicative function of gestures in turn-taking. In an experiment comparing the onset of gestures in social conditions, non-social conditions and monologue/social conditions, gestures were found to occur significantly more frequently in unfilled pauses in social interactions compared to the other contexts (Beattie & Aboudan, 1994). It was claimed that this indicated gestures' conversational function of suppressing the other party from taking the floor in interaction.

To sum up, the current thesis suggests that lexical retrieval might not be the predominant function of the gestures. Moreover, there can be other factors that cause the gestures to appear close to dysfluent speech and/or slightly ahead of their associated speech. In addition to the possible function of assisting speech production, the communicative function of gestures cannot be ignored.

6.4.2. For communication

It seems self-evident that gestures contain meaning that may or may not exist in the co-occurring speech and, hence, tend to play a vital role in

communication. For instance, in the first study, when numerous instances of TFSs checking understanding co-occur with beat/beat-like gestures suggesting the meaning of emphasis, this meaning of importance is only delivered in the gesture rather than in "(do) you see/know what I mean". Also, although those TFSs acting as discourse markers tend not to co-occur with independent gestures/strokes, there are two instances accompanied by beat/beat-like gestures from Supervisors 4 and 6, adding the meaning of importance to the utterance. Similarly, three instances from Supervisor 3 are synchronised with two PUOH gestures and one bounded-space gesture, metaphorically holding and presenting the ideas while seeking for mutual understanding. In the second study, circular gestures often contain information that is absent from the co-occurring speech patterns. For example, when the circular gestures co-occurring with verbs such as 'describe', 'follow', and 'discuss', etc., the meaning of continuity or on-going process is only encoded in the gesture.

The findings of the thesis reveal the high coordination and cooperation between speech and gesture in meaning making in accordance with the speakers' intention in certain discourse contexts. Such an integration undoubtedly benefits listeners by providing more information in both speech and gesture. Previous experimental research has already found that more information can be obtained from video materials containing both speech and gesture information rather than just audio (Beattie & Shovelton, 1999). Also, in another experiment (McNeill, 2005), it was found that viewers of narration can receive information from both speech and gestural channels and then exchange the information between them while retelling the same scene. For example, whereas the narrator put the motion information in speech by saying "and he [came out] the pipe," and the manner information in the up-down bouncing hand, the viewer rearranged the manner information in speech by

saying, “and the cat bounces out the pipe,” and the motion in gesture by moving the hand up and down without bouncing.

The instance above is a good illustration of how listeners not only obtain information from both speech and gesture, but also rearrange the distribution of meaning in the two modalities. McNeill (2005) termed this speech-gesture unit as a Growth Point, this being a minimal unit where gesture and speech interact and the best way to represent meaning is worked out. A similar stance has been taken by Kita (2000) in his Information Packaging Hypothesis, in which the gestural/spatio-motoric thinking and speech/analytical thinking interplay and feed information mutually, converging into one utterance as the most communicatively effective final product in that particular context. Such a hypothesis differs from the Lexical Access Hypothesis and the Image Activation Hypothesis, where the information only flows from gesture to speech as the image in gestures activates the concepts (Hadar & Butterworth, 1997).

However, although benefit for listeners from the gestural information seems apparent, there remains a question as to whether speakers produced gestures for communicative purposes regardless of being conscious of doing so (Holler & Wilkin, 2009). Concerning this issue, numerous controlled experiments have been previously conducted, pointing towards the communicative nature of gestures in interactions. For example, by manipulating the visibility between participants, early research (Alibali et al., 2001) found that speakers produced significantly more referential gestures, though not beat gestures, when the listeners were in sight than when the listeners were behind a screen. Going further, research (Bavelas et al., 2008) has shown that, compared to telephone dialogues and monologues, participants involved in face-to-face communication produced more life-size gestures that contained information absent from speech; speakers even verbally referred to their gestures and produced interactive gestures.

The communicative nature of gestures is also reflected in the influence of recipient feedback on the gestures. One study revealed that speakers not only gestured more often when narrating known information, but also when the listeners were more attentive (Jacobs & Garnham, 2007). These findings were interpreted as evidence of how the degree of attentiveness of the listeners could motivate or demotivate the speakers to gesture. A further study also showed that speakers designed their gestures in accordance with the location of the listeners (Özyürek, 2002).

In addition to research enquiries focusing on the impact of visibility and recipient feedback on gesture production, one factor that is particularly relevant to the current study is common ground, which could also constitute an influential and independent factor on the uses of gestures. For instance, in a communicative task, speakers predominantly encoded the size information of the entities in gesture, or in both gesture and speech, when talking to recipients who did not share the size knowledge. On the contrary, where listeners shared the size information, the speakers tended not to deliver this in gestures, but mainly in speech (Holler & Stevens, 2007). Researchers emphasised that the results showed how the production of gestures was affected by the communicative intention of the speakers. That is, for the known recipients, speakers were only reminding them of the size representation of the entities, but, for the unknown listeners, they aimed to build up brand-new image and using gestures seemed the most effective way of doing this.

Similar results have been generated in a study where a smaller number and less precise gestures were performed while retelling stories to previous listeners rather than to new ones (Galati & Brennan, 2014). Taking a slightly broader angle, another experimental study showed that, when considering both speech and gesture, speakers were found to contain less semantic information in both speech and gesture for known listeners; however, proportionately more iconic and beat gestures were produced compared to

speech. This then indicates that, though the information contained in speech is as low as expected when mutual ground has been established, gesture takes up the communicative weight in order to produce equally informative multimodal utterances (Holler & Wilkin, 2009).

What can be inferred from the aforementioned research on common ground is that, when the common ground has been built up between the interlocutors, the gestures tend to be smaller in number and less salient, or precise, in form. This is then confirmed by the first study of the current research. As the previous studies on the effect of common ground on gesture production all drew on experimental data, the current study using a much more natural corpus provides new evidence for the communicative nature of gestures. The results show that, when TFSs indicate the meaning of checking understanding and less common ground between participants, they tend to co-occur with more salient, independent strokes rather than other gesture phases (e.g. preparation phases, retraction phases). On the other hand, when co-occurring with TFSs acting as discourse markers with more common ground, the speakers tend not to gesture.

Therefore, these results explicitly indicate how common ground between the interlocutors can greatly affect the uses of gestures in interaction, which further suggests their communicative value. When the supervisors have greater doubt about mutual understanding between them and the students, they use both speech and gesture in TFSs; on the contrary, when there is less or no such doubt, the supervisors only use speech.

6.4.3. Interactive gestures

The metaphorical gestures emerging in the first study co-occurring with the TFSs checking understanding are markedly similar to a sub-group of gestures identified and researched by Bavelas and his colleagues as being especially

designed for interaction, these they term interactive gestures (Bavelas et al., 1995, 1992). Targeting conversations, they divided gestures into three broad types, namely, self- and object-oriented gestures/adaptors, hand signals/emblems and conversational gestures. Their main research interest lay in conversational gestures, which they categorised into interactive gestures and topical gestures, with the latter sub-category constituting the majority. Interactive gestures were defined as gestures in interaction that refer to the interlocutors rather than the topical contents. It was found that, compared to monologue and blocked dialogue (participants cannot see each other), interactive gestures were more prevalent in face-to-face conversation. However, the number of topical gestures (gestures referring to the contents) were not significantly affected.

They (Bavelas et al., 1995, 1992) also provided some typical examples of interactive gestures as illustrations of their functions, all of which involved the element of an open hand facing upwards in the gesture form. One example that is particularly similar to the instances co-occurring with "(do) you know/see what I mean" in the current study was a Palm Up Open Hand gesture synchronised with *you know* at the end of an utterance, which aimed to check the mutual ground between the interlocutors. However, seeking for agreement, understanding and help was just one function served by the interactive gestures. Others included: a). delivering gestures (delivering information from speaker to listener, meaning, 'As you know'), b). citing gestures (referring to listener's previous speech, meaning, 'As you said earlier') and c). turn gestures (referring to turn-taking issues, meaning, 'Your turn now').

Although the results of the current study have generated similar examples to the research on interactive gestures, more research is required to identify and categorise interactive gestures in terms of their functions and the corresponding gesture forms. The current thesis is able to generate some new insights. Previous research on the interactive gestures only provided four

representations of gesture forms, which all seem to belong to the open hand gesture family (Bavelas et al., 1995, 1992). However, in the first study, only a few PUOH gestures (produced by only one speaker) co-occurring with the TFSs as seeking for understanding were identified. By contrast, a larger number of beat/beat-like gestures and holding emerged in such uses. In the second study, many instances of circular gestures were also found associated with the function of monitoring mutual understanding. Therefore, based on the current dataset, beat/beat-like gestures, holding and circular gestures might all belong to the category of interactive gestures. What can be clearly inferred from the current study is that there may well exist a wider range of gesture forms in interactive gestures than the few gesture forms exemplified by the previous researchers. Hence, more research drawing on large corpora of face-to-face interactions collected from natural communicative contexts would surely broaden our knowledge of the functions and forms of interactive gestures.

Face-to-face communication is one of the most crucial and prevalent forms of communication. Whereas previous research has mainly focused on how speech, such as discourse markers (e.g. *you know*) and vague languages (e.g. *kind of*) (Cutting, 2007), can be adopted in order to involve the listeners and enhance interpersonal relationships (Carter, Goddard, Reah, Sanger, & Bowring, 2001; Stubbe & Holmes, 1995), the current thesis and the research on interactive gestures has revealed the significant role of gestures in this process.

6.4.4. Discourse functions

The current thesis holds the view that gestures are multifunctional and might have values in both speech production and communication. However, based on the fine-grained analysis of the speech co-occurring with the gestures, a

discourse-based functional framework for gestures was able to be constructed. The rationale behind the proposal for such a framework is that, as the previous distinction between cognition and communication is more cognitive-oriented and based on experimental datasets, a discourse framework built upon the exact speech co-occurring with the gestures drawing from natural dataset such as the NMMC is necessary. Only a brief discussion will be given in this sub-section as detailed descriptions have been provided in Chapters 4 and 5.

Based on the datasets of both studies, five discourse functions have been proposed, including 1) contributing to referential meaning (both concrete and abstract), 2) segmenting utterances, 3) underlining noteworthy information, 4) maintaining coherence, and 5) enhancing inter-subjectivity. The first three functions have already been widely acknowledged and discussed in the existing literature in gesture studies (Calbris, 2011; Kendon, 2004; McNeill, 1992, 2005) and the current research follows their definitions. Contributing to referential meanings refers to the semantic or metaphoric meaning delivered by the gestures, which might or might not exist in the speech channel. Segmenting utterances means their ability to segment the speech they co-occur with from the rest of the utterances. Typical examples include those co-occurring with formulaic expressions such as "(do) you know/see what I mean", "you know" and "sort of". The third function, underlining noteworthy information, usually refers to the way in which some gestures, such as beats and circular gestures, can mark-up certain component(s) of speech as a way of emphasis. Instances include those beat/beat-like gestures co-occurring with "(do) you know/see what I mean" and circular gestures with noun phrases such as "city scapes" and "hotel buildings". However, it could be argued that, any gesture could be serving such a purpose compared to those unaccompanied speech.

However, little focus has been paid to the latter two functions: maintaining coherence and enhancing inter-subjectivity. For instance, circular

gestures seem to play a role of maintaining the flow of communication when they co-occur with dysfluent speech. The coherence function of gestures is also implied in those circular gestures in line with linking expressions such as "and" or clause stems at the beginning of utterances, serving the function of oiling the wheels of the conversation. In addition, holding the hand with/without speech while waiting for the other party to take the floor also reflects this function of linking the turns.

The function of inter-subjectivity differs from the rest of the categories because of its interpersonal rather than transactional nature (Fung & Carter, 2007). Some examples of circular gesture co-occurring with vague language and TFSs (Adolphs et al., 2004; Cutting, 2007) are good illustrations of this function as they seem to help soften the utterances to make them more polite. Similarly, different types of metaphoric gestures co-occurring with "(do) you know/see what I mean" as a means of seeking mutual understanding and involving the listeners also have an interpersonal element in them.

As already indicated in the previous discussion, these functions are not exclusive to each other as gestures can play multiple functions simultaneously (Calbris, 2011). For instance, when a beat gesture co-occurs with a TFS, it adds a meaning of emphasis to the TFS, segments the TFS from the rest of the speech and underlines the TFS visually. Also, when a circular gesture is synchronised with a vague expression such as "sort of", it segments the speech and emphasises it.

This framework of gestural discourse functions involves a wide range of categories, which can be conveniently applied to the analysis of other gestures. Instead of focusing on the cognitive function, the categories largely underline the transactional and interpersonal role both speech and gesture perform in various discourse contexts based on the exact speech co-occurring with them. However, it is possible that other gestures might not be able to perform all the functions listed in this framework, or they might perform more. As the current

framework was built on a limited number and types of gestures, further research aiming for the same or different gestures with larger datasets will help refine the current discourse-based functional framework of gestures.

6.5. Summary

This chapter has demonstrated some valuable contributions made by the current thesis drawn from the NMMC. In terms of examining the temporal relationship between speech and gesture, the first study has offered a solution to the problem of identifying the lexical affiliates of non-iconic gestures, by fixating the analysis on the same group of functional formulaic sequences such as "(do) you know/see what I mean". In addition, the first study has generated evidence that, similar to iconic gestures, the stroke phases of the metaphoric gestures also mostly tend to align with their associated speech. The onset of those strokes tended to precede or synchronise with the TFSs. Building upon the findings of iconic gestures, results from non-iconic gestures further confirmed the close cognitive relationship between speech and gesture (Kendon, 2004; McNeill, 2005). Meanwhile, the high frequency of co-occurrence of formulaic sequences and gestures indicated the processing of formulaic sequences as a multimodal speech-gesture ensemble.

Both studies of the current thesis have found the Conceptual Metaphor Theory (Lakoff & Johnson, 2008) and the concept of image schema particularly useful for interpreting the emerging patterns co-occurring with recurrent speech (i.e. the TFSs) and recurrent gesture (i.e. circular gestures). Context is vital for analysing which analogical link(s) between speech and gesture has/have been activated among all the links provided by different gesture parameters (e.g. velocity, handshape, trajectory, space, etc.) (Calbris, 2011). During the process, the widely accepted view that metaphoric thinking is rooted in schematised physical experiences of the body has been confirmed,

and that metaphor is inherently modal-independent and multimodal (Forceville & Urios-Aparisi, 2009).

Regarding the functions of gestures, the findings of both studies supported the dual functions of gesture in both speech production and communication. The thesis also expanded our knowledge of the form of interactive gestures (Bavelas et al., 1992) and suggested how the research on interactive gestures could be further pursued. Moreover, based on both studies, a new discourse-oriented framework for categorising gesture functions was proposed.

Chapter 7. Conclusion

7.1. Thesis overview

The thesis started with the aim of exploring multimodal corpus-based approaches for investigating multimodal units of meaning in speech and gesture. For this purpose, two exploratory studies were rigorously designed and effectively executed. The first used a formulaic sequence, i.e. frequently occurring multiword expressions, as an analytical unit for demonstration and the second used a recurrent gesture. Each study aimed to discover the patterned uses of speech and gesture as multimodal ensembles, and build up a speech-gesture profile representing the multimodal units of meaning of the chosen formulaic sequence and the chosen recurrent gesture, respectively.

The analysis of each study revolved around three aspects in accordance with the three research questions: exploring the temporal relationship between speech and gesture, identifying emerging speech-gesture associations and examining the discourse functions of gesture. The thesis had a cross-disciplinary nature and utilised a variety of qualitative and quantitative methods and tools from different theoretical traditions. The main methods applied throughout the thesis stemmed from the areas of discourse analysis, corpus linguistics and gesture studies, together with monomodal and multimodal analytical toolkits such as AntConc, CLAN and ELAN.

To address the aforementioned questions, Chapter 2 reviewed relevant concepts, theories, frameworks and research in multimodality, corpus linguistics, formulaic sequences, gesture studies, and the multimodal units of meaning. Chapter 3 introduced and justified the use of the Nottingham Multimodal Corpus as the data source, and detailed the construction of a five-phase gesture coding scheme in accordance with the data. It also presented the formats designed for speech-gesture transcription.

Based on the existing knowledge reviewed in Chapter 2, and the data and framework for analysing and representing speech-gesture alignment in Chapter 3, the fourth chapter provided the specific approaches and procedures for exploring the speech-gesture profile of the target formulaic sequence (TFS), "(do) you know/see what I mean", during which the three research questions were answered. Chapter 5 focused on the second study, reporting the methods and processes of establishing a multimodal unit of meaning for the circular gestures. The answers to the three research questions were also determined.

The approaches, findings and implications presented in Chapters 4 and 5 have furthered our knowledge regarding the interplay between speech and gesture as two integral components in the multimodal units of meaning, and the role of gestures in interactions. The drawbacks of traditional linguistic research on the units of meaning drawing on textual corpora alone was indicated, especially with respect to the partial description of meaning. The limitations of the qualitative approach to the meaning of recurrent gestures was also revealed, providing an incomplete picture of their meaning variations in real-life uses. Moreover, throughout the thesis, concepts, methods and tools in monomodal corpus linguistics have proved particularly valuable for examining, foregrounding and describing the temporal and functional relationship between the verbal and nonverbal data streams.

Chapter 6 demonstrated the new insights that the thesis provided for some important issues and debates in the field of gesture studies. They included 1) new methods to examine the temporal relationship between non-iconic gestures and speech, and implications for investigating the positions of recurrent gestures in discourse; 2) the conceptual metaphoric theory as an explanatory tool for speech-gesture meaning associations, and evidence for the links between gesture, cognition and the physical world; 3) implications for the dual functions of gesture for both speech production and

communication, meanwhile emphasising the interactive nature of some gestures.

7.2. Limitations and avenues for future research

As emphasised throughout the thesis, the primary goal was not in making broad generalisations for the multimodal units of meaning of the TFSs and the circular gestures, but exploring multimodal corpus-based approaches for investigating the multimodal units of meaning. The narrow representation of datasets has been admitted, being limited to the small number of instances drawn from a limited range of speakers in one genre related to the pedagogical context, i.e. supervision meetings between supervisors and postgraduate students in the NMMC (refer back to 3.2.1. for a discussion of the limitations of the corpus). The first study mainly drew on one formulaic sequence with less than 100 instances from one female supervisor in one supervision meeting in one subject area (literature). Based on a much larger number of instances (570 from 8 videos involving more than 10 speakers), the second study might be more robust than the first one, but it was still restricted due to the uneven number of instances from each speaker and the narrow representation of the entire corpus. The limitations of the corpus and the two studies call for caution when trying to generalise the findings. However, such limitations were necessary due to limited time and finance available during the PhD.

Another limitation concerns the adaption of Sinclair's (2004) unit of meaning model to the description of speech-gesture relationship in the second study, which explored the relationship between the circular gestures and their co-occurring speech. Such an adaption may seem unusual when considering a few factors. To begin with, Sinclair's model is constructed based on written language and is, therefore, usually applied to the analysis of

textual corpora, whereas the current study drew on a multimodal corpus and used the model to describe speech and gesture relationship. Hence, the nature of the datasets is entirely different. Secondly, as has been clarified in detail in 5.2.6 (Chapter 5), the concepts, i.e. colligation, semantic preference and semantic prosody, were not used in the exact same way here as in the original model. The three concepts originally refer to the emerging grammatical, semantic and functional patterns in the concordance list surrounding the searched lexical item, but are applied here as general concepts to categorise all the speech patterns co-occurring with the circular gestures. Hence, all the grammatical patterns were assigned to colligation, the semantic patterns to semantic preference, and the functional ones to semantic prosody. In addition, one of the patterns, collocation, that refers to the exact word(s) frequently co-occurring with the searched item was not used here as the circular gestures tended to co-occur with different lexical items. Despite the above aspects, the three concepts were effective in terms of categorising the emerging speech patterns and describing the speech-gesture relationship. Borrowing linguistic concepts derived from textual-based research to multimodal data is necessary and common at this initial stage of exploration as it adds coherence to the field as we gain more knowledge and develop further. It also helps researchers understand the multimodal research better as they are already familiar with those concepts and theories.

Several directions for extending the current thesis are outlined as follows:

1) The most straightforward next move is to verify and refine the findings of the two studies drawing on a much larger corpus with a greater number of instances. Such a corpus could represent a similar or different communicative context, allowing for contrastive analysis, or a wide range of contexts.

2) A further way of expanding the thesis, drawing on larger corpora, would be to add one or more streams of data to the analysis of the multimodal units of meaning, such as head movements, prosody, facial expressions.
3) This approach can also be used to investigate the multimodal units of meaning in other multiword strings (e.g. *you know*, *I mean*, etc.) or recurrent gestures (e.g. beat gestures, holding, etc.).
4) Taking a different perspective, as the manual coding of gestures is extremely time-consuming, future research could address some technological issues based on the current data. Such endeavours could involve developing technology for (semi-)automatic segmentation of gesture phases, and/or (semi-)automatic search and recognition of the recurrent gesture, and/or (semi-)automatic construction of gesture concordances.
5) The thesis can also be extended by collecting perceptual data, such as interviews, which can provide a further perspective for examining the role and nature of gesture in communication. For instance, speakers could be asked specifically what they meant by saying "(do) you/know see what I mean" and/or by performing perform certain gestures. The listeners could also be asked whether they have noticed the gestures and how they interpret their meanings. This will require collecting a new corpus from a fresh group of participants, from whom queries can be close to the time of recording. Such data could also have certain implications for the level of consciousness of gesture perception and production.

7.3. Implications for applied linguistics

In the previous sections of the thesis, many implications of the current thesis were discussed for the areas of research on formulaic sequences and the

multimodal unit of meaning (refer to Chapter 4, 4.5. and Chapter 5, 5.4.), and the speech-gesture relationship (refer to Chapter 6). This section will demonstrate its implications for other areas of applied linguistics, with a primary focus on pragmatics, particularly corpus pragmatics, and a brief discussion of others.

The past decade has witnessed an increasing number of scholarly works in corpus pragmatics, a field that integrating corpus linguistics approaches into the study of language function in context, i.e. pragmatics (Thomas, 1995). Earlier volumes that propose the applications of corpus approach to pragmatics include those from Adolphs (2008), Jucker, Schreier and Hundt (2009), Romero-Trillo (2008), and O'Keeffe, Clancy and Adolphs (2011). More recent ones include the yearbooks of corpus linguistics and pragmatics edited by Romero-Trillo (2013, 2014, 2015, 2016), signposting each year's new trends, and a handbook of corpus pragmatics edited by Aijmer and Rühlemann (2015). Among the extensive publications in corpus pragmatics, a dominant number of them focus on identifying the form and function of pragmatic markers in different contexts (including discourse markers as a sub-category of pragmatic markers) (also see a recent review on pragmatic markers and other strands of corpus pragmatic research in Rühlemann & Clancy, 2018). With the assistance of corpus tools, common analyses include the frequency distribution, collocation and position (i.e., turn initial, medium or final) of pragmatic markers in the target text types. The availability of socio-cultural information in the corpora that most researchers draw on affords them opportunities to explore pragmatic markers with regards to situational and social variables such as region, gender, age, language levels in learners, formality, interpersonal relationship, social status, etc. For instance, by comparing a Hong Kong English learner corpora (secondary pupils) with a British spoken pedagogic sub-corpus of CANCODE, Fung and Carter (2007) investigated the scope of discourse markers used in the two groups, and found

out a limited range of functional varieties in English learners compared to native speakers: learners in their study used a variety of discourse markers serving referential functions (*and*, *but*, *because*, *OK*, *so*, etc.), but were restricted in using discourse markers having interpersonal, structural and cognitive functions (*yeah*, *really*, *say*, *sort of*, *I see*, *you see*, *well*, *right*, *actually*, *cos*, *you know*, etc.). Aijmer (2013) presented her research on *well*, *in fact* and *actually* (adversative pragmatic markers), and *and things* and *and stuff like that* in her well-known monograph. She also studied *I think* based on the British component of the International Corpus of English (ICE-GB), demonstrating its functional variations in terms of frequency, key collocations, text types (broadcast debate vs. conversation) (Aijmer, 2015). Fraser's (2015) research stands out as it explored the co-occurrences of contrastive discourse markers (e.g. *but*, *on the other hand*) and imperative discourse markers (e.g. *so*, *as a result*). Expanding the research target, there was also a study on pragmatic markers (*Uh*, *Um*) in adolescents with autism spectrum disorder (Irvine, Eigsti, & Fein, 2016).

In addition to pragmatic markers, pragmatic research on speech acts is another strand that can be compatible with the bottom-up search method in corpus linguistics, though studies of this type are less pervasive compared to those on pragmatic markers. Although an accurate automatic search for speech act functions (e.g. apology, thanking, suggesting) is still technically challenging, based on existing knowledge, researchers can predetermine conventionalised speech act expressions that can fulfill certain speech act functions and then run automatic corpus searches of them. The searched speech act expressions can then be investigated in terms of frequency, collocation, co-text and function. Based on naturally occurring language, researchers can re-examine the previous taxonomies in speech act theories and refining existing frameworks by incorporating new categories emerged from the corpora. Some well-known scholarly works of this nature include Aijmer's

(1996) research on thanking, apologies, requests and apologies drawing on the London-Lond Corpus of Spoken English, and Adolphs' (2008) research on suggestions (e.g. *how about*, *why don't you*) based on CANCODE. Among many others, corpus investigations were carried out on expressive speech acts in the pragmatically annotated SPICE Ireland corpus (Ronan, 2015) and in educational e-chats in both English and Spanish (Maíz-Arévalo, 2017).

As fruitful and inspiring as the previous research on the function of pragmatic markers and speech act expressions can be, most of them focus on recorded spoken discourse that is rendered into a textual record (or transcript). Other modes of communication (e.g. gesture, prosody, facial expressions, etc.) are usually not taken into consideration, yet they are crucial when it comes to performing speech acts and other pragmatic functions. Treating a transcript of spoken interaction in the same way as we would treat a corpus of texts that originate from a written record is therefore problematic in a number of respects. One apparent drawback is that a pragmatic function that has been assigned on the basis of a transcript might be incorrect when the information of the accompanying prosody or gesture is not available. In the current thesis, this becomes evident during the coding of the formulaic sequences "(do) you know/see what I mean" as the involvement and analysis of the nonverbal responses from the students (i.e. head-nodding) turned out to be crucial in order to correctly distinguish the three functions (i.e. checking understanding, discourse marker and a combination of both) (refer to Chapter 4, 4.2.2. for detail). Moreover, in view of the speech-gesture profile built up for "(do) you know/see what I mean" (refer to Chapter 4, 4.5.3. for detail), it suggests that gestures can alter the function of the whole unit of meaning. For example, a TFS having the meaning of checking understanding accompanied by a beat/beat-like gesture can deliver a stronger sense than one that is not accompanied by gesture. Similar results have also been found in the study on head nods and response tokens, i.e. the intensity of head movements can

influence the meanings of response tokens as multimodal units (Knight, 2011) (refer to Chapter 2, 2.8. for a discussion of the research). Hence, what can be inferred from the current research on "(do) you know/see what I mean" and other multimodal corpus research is that, in order to allow a more accurate and nuanced description of the form and function of multiword expressions in spoken language, other modes such as gesture will need to be explored together with speech drawing on multimodal corpora.

From a different perspective, whereas previous research in pragmatics solely centered on speech, earlier gesture studies have demonstrated the role of gesture in performing certain speech act functions that cannot be inferred from speech. These include Kendon's (1995, 2004) research on the illocutionary forces of certain gestures such as the open hand palm up (offering, presenting, etc.) and down gestures (negating, rejecting, etc.), and Ladewig's research on the cyclic gesture (encouraging) (refer to Chapter 2, 2.5. for a detailed discussion of the pragmatic functions of gesture families and recurrent gestures). In the current thesis, the pragmatic functions of gesture are also foregrounded in many instances. This is shown in the discussion of the function of checking understanding performed by some gestures (e.g. beat/beat-like gestures, holding and some cyclic gestures), as well as the function of emphasising noteworthy information by the beat/beat-like gestures and some cyclic gestures. These findings then suggest the necessity for future corpus pragmatic research to investigate those circumstances where gestures are used as the main channel for realising speech act functions. Such research could be based on daily communication where gestures are subtly used by speakers to perform speech act functions, manage turn-taking, and seek for responses, etc.; it could also draw on datasets that are collected from special contexts where gestures are simply more effective in transferring information (noisy environment such as construction site, restaurant, etc.), or in some socio-cultural contexts where

certain speech is forbidden. Such research will deepen our knowledge of the pragmatic functions of gesture in various socio-cultural contexts, as well as its relationship with other semiotic modes (e.g. speech, objects in the environment, etc.).

In addition to the apparent implications of the current thesis for pragmatics, it also has implications for many other fields in applied linguistics, some of which deserve at least a brief mention here. One of these areas is the idiosyncratic features of multimodal formulaic language in individual speakers. The idiosyncrasy of human behaviour is a widely acknowledged, but rarely investigated phenomenon. This is suggested by the particular high frequency of uses of "(do) you see/know what I mean" in one particular supervisor in the case study (Video 6), and the performance of the Palm Up Open Hand gestures synchronised with the TFSs by the supervisor in Video 3. These findings indicate that people may have preferences for certain formulaic sequences and gestures, and such characteristics in speech and gesture use forms vital components of individual speaking styles.

A further insight relates to the association between the sociolinguistic concept of power and the uses of multimodal formulaic sequences. This is indicated by the fact that all the TFSs, "(do) you know/see what I mean", are used by the supervisors. As this formulaic sequence mostly has the connotation of checking mutual understanding between speakers, the use of it by the supervisor alone, to a large extent, suggests their superior position in the pedagogical context in relation to their students. It is likely that the supervisors tend to use this formulaic sequence to ensure the common ground and to involve the students in interaction.

An additional implication is relevant to the topic of culture and recurrent gesture. As existing research on gesture tends to focus on the uses of particular recurrent gestures in a specific culture, the similar meaning identified in the circular movements from the English speakers in this thesis and German

speakers in Ladewig's research (2011) indicates the similar uses of this gesture in the two cultures. Hence, as some gestures can be universally observed, such as the beat gestures, open hand palm up/down gestures, etc., instead of focusing on cultural-specific gestures, knowledge of the similarities and differences of the same gesture in various cultural groups is also vital for understanding intercultural communication.

7.4. Applications of the approaches

The approaches explored by the current thesis examine the multimodal units of meaning composed by speech and gesture, and the approaches can potentially be applied to a variety of applications. The following are two highlighted areas in second language (L2) teaching and human-computer interaction (HCI), which are of particular relevance to this thesis.

One application is the teaching of multimodal formulaic sequences in second language (L2) teaching. It has long been recognised that, even for highly proficient post-childhood learners, acquiring formulaic language can be very challenging (Paquot & Granger, 2012). The fundamental reason that prevents post-childhood language learners from achieving a native-like instinct of formulaicity may be that they tend to analyse and break down the formulaic sequences into their components rather than processing them as wholes as child L1 learners do (Wray, 2004). Such a tendency of word-by-word processing in learners may also be further strengthened by the lack of input of formulaic language from textbooks and teacher talks, and the biased philosophy of teaching and testing word-sized vocabulary in L2 teaching and learning. The significance of teaching and learning formulaic sequences in L2 learners has led researchers to explore different pedagogical methods for addressing the issue. For instance, some researchers have proposed using and designing corpora for teaching formulaic language,

especially in the field of English for Academic Purposes (EAP) (Biber, 2009; Simpson-Vlach & Ellis, 2010), and much attention has been given to the methods used to extract lists of multiword expressions that can be applied in teaching. The rationale underlying this type of research is that each register is characteristic of distinctive lexical strings, and identifying and teaching such strings in, for example, academic context can help learners to improve their academic writing or speaking (Hyland, 2012; Swales, 1990).

Other researchers have directly focused on the effectiveness of various teaching interventions in teaching formulaic sequences. One common intervention method is awareness raising and noticing in language learners through explicitly foregrounding formulaic language during the teaching process, such as underlying the target sequences, looking them up, and engaging with them in different teacher-led and student-centred activities (refer to Boers & Lindstromberg, 2012 for a comprehensive review of different experiments and interventions on teaching formulaic sequences). The method of noticing is mostly proposed based on the Noticing Hypothesis (Schmidt, 1990) in language acquisition, which suggests that perceiving unconsciously (i.e. subliminal learning) does not lead to acquisition, and the consciousness level of noticing, i.e. being aware of something, is necessary and sufficient to turn input into intake. An investigation of the acquisition of collocations in Japanese English learners indicates that, for any significant incidental learning to occur, a learner would have to read and listen to a graded reader for at least 10 to 15 times (Webb, Newton, & Chang, 2013). In fact, they (Webb, Newton, & Chang, 2013) reported that even after encountering the sequences for 15 times in reading and listening, the success rate of recall was only around 50%; the percentage for the 10-times group was only 30%. Since L2 learners usually do not encounter many formulaic sequences as frequently as such in their environment, helping them to notice and raise consciousness in class is essential for actual learning. For instance, Bishop's (2004)

experimental research showed that increasing the typographical salience by underlying and colouring the target formulaic sequences significantly enhanced their chance of being looked up by students, which in turn led to better performance in the following comprehension task in the experimental group. In another study, Boers et al. (2006) found out that the students in the experimental group who had been taught and encouraged to notice and learn formulaic sequences used more of them in the follow-up oral test than those in the control group did, who were taught in a traditional grammar-lexis dichotomy method.

In a recent study (Lindstromberg, Eyckmans, & Connabeer, 2016), a text reconstruction task, dictogloss, was applied to enable learners to not only notice the target formulaic sequences but also produce texts with them. In the dictogloss task, a sheet with target formulaic sequences was provided to the students in the experimental group while they were required to reconstruct the whole text after listening to it and taking notes for a few times. A blank sheet without the formulaic sequences was provided to the control group. Results showed that the experimental group significantly outperformed the comparative group in terms of the recall and production of those formulaic sequences in a delayed reconstruction test a week after.

Among many others, the studies above reveal the benefits of consciousness raising and noticing in teaching formulaic language for language learners, nonetheless, the short-term effects after noticing them for only a few times do not necessarily guarantee long-term memorisation and actual acquisition (Petchko, 2010). In fact, none of the aforementioned research has reported any impact of the intervention on the retention of formulaic language. The contributions to learning and retention of formulaic sequences made by other interventions have also been investigated. These include memorising long oral texts that bypass grammatical analysis (Wray, 2002, 2008), focused instruction of formulaic language in a fluency workshop

to enhance speaking fluency (Thomson, 2017; Wood, 2009), dictionary lookups (Dziemianko, 2010; Laufer, 2011), and corpus search (Wu, Witten, & Franken, 2010), to name just a few.

Whereas the proposed methods to teach formulaic sequences discussed to this point maintain at the textual level, there has been little research on the ways in which other modes such as prosody and gesture can impact the learning outcomes of formulaic sequences. A few empirical studies (Boers & Lindstromberg, 2005; Boers, Lindstromberg, & Eyckmans, 2012; Lindstromberg & Boers, 2008a, 2008b) have been conducted exploring the effect of sound patterning on the recall of formulaic language, suggesting that multiword expressions with sound repetitions such as alliteration (i.e. the repetition of the initial consonants such as *cut corners*) and assonance (i.e. two or more words starting with different initial consonants and the same vowel such as *turn a blind eye to*) are more likely to be noticed and memorised. Such studies indicate that formulaic sequences with moderately high frequency and sound repetition patterning are worthwhile teaching (Boers & Lindstromberg, 2005).

In terms of gesture, there has been ample evidence indicating that observing gestures can facilitate learning and recall of knowledge such as mathematical equivalence (Congdon et al., 2017; Cook, Duffy, & Fenn, 2013; Koumoutsakis, Church, Alibali, Singer, & Ayman-Nolley, 2016; Singer & Goldin-Meadow, 2005), symmetry (Valenzeno, Alibali, & Klatzky, 2003), and vocabulary (Demir & Goldin-Meadow, 2016; Kushch, Igualada, & Prieto, 2018; Porter, 2016). For instance, Cook, Duff and Fenn (2013) compared the learning gains of solving mathematical equivalence problems (3+5+8 = ___ + 8) in two groups of children (n=184, ages = 7-10), one was taught with a videotaped lesson including both speech and gesture, and the other including only speech. The results showed that children watching the lesson with both speech and gesture not only performed better than those in

the other group, but also improved their performance after 24 hours' delay and transferred the knowledge to different problems. They claimed that such findings suggest the benefits of gesture for learners in comprehending abstract mathematical concepts and consolidating knowledge by providing visual and concrete motor representations (Cook et al., 2013). In a similar vein, another study investigated the potential influence of medium (live lesson *versus* video lesson) on instructing mathematical equivalence to individual students (Koumoutsakis et al., 2016). The results showed that there was no major effect on learning and transfer from the medium of teaching, but speech instruction with gesture led to significantly more gains in such aspects than speech-only instruction.

Gesture has also long been found conducive to L1 vocabulary learning in children. Children rely on parents' gestures to comprehend word meanings before speaking, and the amount and meanings of parents' gestures are predictive of the size of comprehension vocabularies at the following stages of language development (refer to Demir & Goldin-Meadow, 2016 for a detailed review of the role of gesture in child language development). Gesture can also facilitate L2 learning. Focusing on beat gestures, research has shown that, when they were performed naturally and accompanied by prosodic prominence in speech (i.e. the focal pitch accent in speech), they helped Catalan native speakers learn Russian words better than those who were exposed to no gesture or to beat gestures without the presence of the focal pitch accent in speech (Kushch et al., 2018). This confirms the results of other studies on beat gesture's role in, for instance, facilitating recall of spatial directions (Austin & Sweller, 2014) and isolated words (though only in adult participants, but not children) (So, Sim Chen-Hui, & Low Wei-Shan, 2012). In the meantime, these findings of beat gestures (Austin & Sweller, 2014; Kushch et al., 2018; So et al., 2012) run counter to the previous argument that only iconic gestures, i.e. gestures have meanings, can facilitate

memory recall, but not non-representational gestures, such as beat gestures (Feyereisen, 2006). In particular relevant to the current thesis, Porter's (2016) research is possibly the only research to date that investigates the effect of observing gesture in teaching formulaic chunks in French lessons, which reveals the benefit of showing gestures in short-term memory retention, but not long-term. Hence, this suggests that, although gesture can facilitate learning, long-term retention still takes time and revisions.

Compared to *observing* gestures discussed above, the positive impact of *performing* gestures is also evident. For instance, research on child L1 language development has found out that children's gesture is a strong indicator of their later lexical and syntactical development (Iverson & Goldin-Meadow, 2005), i.e. the meanings children gesture tended to become their verbal language, and the initial combinations of a gesture and a word tended to develop into two words. In addition, in a series of experiments conducted by Cook, Yip and Goldin-Meadow (2010), they found out that those participants that performed gestures in the immediate test (describing animated actions) outperformed those without performing gestures in the delayed cued and free recall tasks. This, therefore, suggests that performing gestures while describing those actions makes the information more memorable than without performing them.

To summarise, previous research on the effectiveness of different pedagogical interventions in teaching formulaic sequences indicate that it is necessary for instructors to raise L2 learners' awareness and explicitly teach formulaic sequences in class. Moreover, in view of the potential benefits of involving gesture in learning and memory retention revealed by existing research, it is plausible to propose a multimodal approach for teaching formulaic sequences in L2 learning. As straightforward as the approach appears, one crucial question that has not been raised or answered is what gesture to teach along with formulaic language. Whereas the gestures

performed in the previous studies mentioned above are mostly based on the researcher or instructor's intuition, the multimodal corpus-based approach proposed by the current thesis points towards a more reliable and valid method to decide what gestures to teach when teaching formulaic sequences. Such a method would be particularly useful for teaching those functional discourse markers such as "(do) you know/see what I mean", whose co-occurring gestures are much difficult to decide based on intuition than other sequences (such as concrete actions such as *jump up/down*, *come out of*).

If the findings of the first study on the TFS "(do) you know/see what I mean" are replicated by future research, students can be informed that they or their interlocutors may use certain gestures, such as a beat/beat-like gesture with(out) holding and holding, to foreground the meaning of checking understanding. On the other hand, they may not gesture if the TFS is used as a discourse marker. During the process of explicit teaching, corpus evidence can be used for illustration, and activities such as role-play can be designed to help the students practice with the TFS. It is likely that further explanations and performance of gesture in the teaching interventions can not only facilitate the students processing the TFS as a whole, but also, according to the research discussed above, leave more memory traces with the students and lead to more efficient learning. The apparent downside of applying such a method is the requirement of extensive research on the patterned uses of gestures in certain frequently occurring functional formulaic language such as "you know", "I mean", "sort of", etc. before we can actually teach them. However, such work will be worthwhile considering the usefulness of such sequences in communication.

In addition to teaching gestures together with formulaic sequences, it is also advisable to teach the meanings of recurrent gestures based on empirical evidence, and to provide information regarding the multiple functions of gestures in general in L2 education. Hence, the approach used in the second

study can be adopted to examine the meaning variations of recurrent gestures, and such findings can then be integrated into the English language teaching curriculum. Using the circular gesture as an example, such teaching can be implemented at two stages. The first would be the explicit teaching of some frequently occurring (e.g. the top 5) speech patterns co-occurring with the circular gestures, or the most frequent ones in each aspect of the unit of meaning model, i.e. colligation, semantic preference, semantic prosody. The second stage would be the instruction of their discourse functions (e.g.: demarcating speech, emphasising noteworthy information, etc.) as demonstrated in this thesis. Again, corpus datasets can be adapted as teaching materials for various activities such as group discussion and role play. Teaching gestures is very likely to raise the learners' awareness of their functions and importance, therefore, encouraging more attention to be paid to them in face-to-face communication. This is, without doubt, beneficial in terms of enhancing students' communication skills.

Human-computer interaction (HCI) is also an important area of application for the thesis. The generation and animation of nonverbal behaviours including gestures in agents rely on the input of pre-set gestures in the behaviour generation systems. Such systems include the behaviour expression animation toolkit (BEAT）(Cassell, Vilhjálmsson, & Bickmore, 2004), and the behaviour mark-up language (BML) (Vilhjálmsson et al., 2007), etc. Such systems have a library of gestures that are retrieved during the gesture animation process, the contents of which mostly stem from previous findings on the semantic and prosodic relationship between speech and gesture, i.e. certain lexis in certain contexts are pre-assigned to be associated with a gesture in the architecture. However, at times such contents have to rely on animators' intuition as solid empirical evidence of the functional relationship between speech and gesture is lacking.

Having said that, in order to enhance the behaviour generation

architectures and HCI technology, it is vital to have sufficient knowledge of the form-meaning associations between speech and gesture, which can be pre-stored in the system. Hence, to achieve that, findings of the temporal and functional relationships between speech and gesture drawn from research such as this thesis can be very useful. For instance, if the findings of the first study are validated in the future, the rules such as accompanying "(do) you know/see what I mean" with beat gestures or holding as checking understanding and no gesture for discourse markers can be pre-stored in the animation systems. Similarly, the associations between the circular gestures and their co-occurring lexical and grammatical patterns can also be applied.

Many agents have been built, showing the potential positive effects of agents on users in different contexts, such as a pedagogical agent (Meij, Meij, & Harmsen, 2015) and military training agents (Goldberg & Cannon-Bowers, 2015), etc. The investigations of agents have indicated that, in order to build up trust between human users and agents, human-human interactions in the target context of HCI will first need to be thoroughly and systematically investigated, both verbally and nonverbally (Ring, Shi, Totzke, & Bickmore, 2015). Hence, for this purpose, applying the approaches developed in this thesis for exploring the multimodal units of meaning in different communicative contexts are particularly relevant and valuable.

7.5. Summary

This chapter has provided an overview of the thesis, including the topic, the aims, the research questions, the general approach as well as the main findings, implications and chapter structure. It has also demonstrated implications of the findings for other areas in applied linguistics, including corpus pragmatics, the idiosyncrasy of formulaic sequences, the power relationship in the pedagogical context and the uses of gestures in different

cultures. The chapter has also proposed two areas of applications that are particularly relevant to this thesis: one is the teaching of multimodal formulaic sequences in the L2 learning context, and the other is the application of semantic and temporal associations between speech and gesture in the behaviour generation systems in HCI.

In a word, the thesis is mostly valuable and relevant in terms of continuing the recent methodological and theoretical shift of moving from the monomodal to multimodal in (applied) linguistics. It is believed that this thesis can provide numerous areas, such as corpus linguistics, discourse analysis, sociolinguistics and pragmatics, etc., with a more fine-grained approach to explore theories, frameworks and issues.

Appendix I Speech transcription conventions for the Nottingham Multimodal Corpora

(Adolphs, 2008, pp. 137–138)

Transcription convention	**Symbol**	**Explanation**
Speaker codes	<S01>, <S02>, <S0F>,<S0M>	Each speaker is numbered with separate codes. Where it has been impossible to identify a particular speaker, which is sometimes the case in conversations with many different speakers, the transcriber has indicated whether the speaker was male <S0M> or female <S0F>.
Extralinguistic information	<$E>. . .</$E>	This includes laughter, coughs and transcribers' comments about speech on the tape, including notes on anonymised elements, e.g. <$E> company name </$E>
Unintelligible speech	<G?>, <G1>	Unintelligible speech is marked with these brackets. Where the number of unintelligible syllables can be determined, this is indicated within the tag, e.g. <G1>, <G2>, etc. Where this number cannot be determined the transcriber has used the symbol <G?>
Guess	<$H>. . . </$H>	Where the accuracy of the transcription is uncertain, the sequence of words in question is placed between these two angle brackets.
Interrupted sentence	+	When an utterance is interrupted by another speaker this is indicated by using a + sign at the end of the interrupted utterance and at the point where the speaker resumes his or her utterance: <S 01> You could + <S 02> Yeah <S 01> + come round tonight
Overlap	<$O1>. . .</$O1>	Some parts of the corpus have been coded for overlapping speech. For this purpose, the overlap is indicated by numbered angle brackets: <$O1>. . .</$O1>, <$O2>. . .</$O2> <S02> Oh. So I'll have to re-calibrate those <$O1> then <\$O1>? <S01> <$O1> Yes. <\$O1> I'm afraid <$O2> so. You will have <\$O2> <S02> <$O2> Oh. Will it <\$O2> not take the previous calibration?

Unfinished sentence, repeat, or false start	<$=> . . . </$>	Unfinished sentences of any type are placed between these angle brackets. This type of coding aids subsequent parsing procedures performed on the data, as well as analyses of patterns of spoken grammar in the corpus.
Punctuation	. ?	A full stop or question mark is used to mark the end of a sentence (depending on intonation). 'Sentences' are anything that is felt to be a complete utterance, such as: <S02> Erm is MX there? <S01> He's got somebody with him at the moment MX. <S02> Right. 'Right' is considered as a sentence here. Unfinished sentences are not followed by a full stop.

Appendix II A sample of participant consent form in NMMC

Research Participant Consent Form

The School of English Studies at the University of Nottingham is a leading centre in Applied Linguistics and as such, we are always seeking to improve our understanding of human communication and spoken interaction. Researchers in the School are currently part of a 3-year ESRC funded project. The aim of this Project is to investigate the relationship between aspects of language and gesture. In order to do this we are building a corpus of 250,000 words, made up of videos of lectures (50%) and supervisions of postgraduate students (50%). We are working with colleagues in Computer Science who have devised an image tracking technique that will allow us to track the movement of the hands and align this with an orthographic transcription of the same event.

Any information that is used will be used anonymously wherever possible and only with your written consent. Your data will be held in confidence and stored securely.

Please read the following options carefully and chose the one you feel most comfortable with. Please be assured that you will be able to change the option you have chosen or withdraw your consent at any time. Please tick all the statements that apply to the level of consent you are prepared to give.

We request your consent with regard to the following video recording
Name of participant:
Type of recording:
Date of recording:
Location of recording:

I consent to having my supervision/lecture transcribed and the video and transcription will be added to a corpus. I understand that I will not be identifiable to anyone outside the Project and that the data will be used for research purposes only. [__]

I also consent to still images from my supervision/lecture being used in any publications that are based on the work done as part of the Project. [__]

I also consent to moving images and the sound of my voice being used by researchers in the University of Nottingham for both internal and external presentations. I understand that the data will not be anonymised. [__]
Signed: …… Date: ……

Thank you for your time and cooperation in helping us with our research.

Appendix III Consent form for researchers (signed)

Permissions – For researchers

I agree that I will use data from the ESRC-funded Nottingham Multi Modal Corpus project for research purposes only. If I wish to publish examples of any data taken from the Corpus, I will seek permission to do this from a member of the Research Team. I agree not to show the data to anyone or display the information in a public place.

Signed: Chen Yaoyao Date: 15.April.2015

Thank you for your cooperation.

Guide to CD-ROM data disk

File 4.1 Speech episodes for all instances of “(do) you know/see what I mean”

File 4.2 Transcript of gestures co-occurring with “(do) you know/see what I mean” for Supervisor 6 in Video 6

File 4.3 Transcript of gestures co-occurring with “(do) you know/see what I mean” for other supervisors

File 4.4 Color-coded transcript for speech-gesture alignment in “(do) you know/see what I mean” in all videos

File 5.1 Speech patterns co-occurring with circular gestures in all speakers

File 5.2 Transcript for circular gestures (noun phrase)

File 5.3 Transcript for circular gestures (modification)

File 5.4 Transcript for circular gestures (deixis)

File 5.5 Transcript for circular gestures (dysfluent speech)

File 5.6 Transcript for circular gestures (clauses)

File 5.7 Transcript for circular gestures (preposition phrase)

File 5.8 Transcript for circular gestures (discourse marker)

File 5.9 Transcript for circular gestures (conjunction)

File 5.10 Transcript for circular gestures (on-going action)

File 5.11 Transcript for circular gestures (intensification)

File 5.12 Transcript for circular gestures (compare and contrast)

File 5.13 Transcript for circular gestures (negation)

File 5.14 Transcript for circular gestures (monitoring mutual ground)

File 5.15 Transcript for circular gestures (vagueness)

File 5.16 Transcript for circular gestures (confirmation)

References

Adolphs, S. (2006). *Introducing Electronic Text Analysis: A Practical Guide for Language and Literary Studies*. London: Routledge.

Adolphs, S. (2008). *Corpus and Context: Investigating Pragmatic Functions in Spoken Discourse*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Adolphs, S., & Carter, R. (2007). Beyond the word: new challenges in analysing corpora of spoken English. *European Journal of English Studies*, *11*(2), 133–146.

Adolphs, S., & Carter, R. (2013). *Spoken Corpus Linguistics: From Monomodal to Multimodal*. London: Routledge.

Adolphs, S., Brown, B., Carter, R., Crawford, P., & Sahota, O. (2004). Applying corpus linguistics in a health care context. *Journal of Applied Linguistics*, *1*(1), 9–28.

Adolphs, S., Knight, D., & Carter, R. (2011). Capturing context for heterogeneous corpus analysis: Some first steps. *International Journal of Corpus Linguistics*, *16*(3), 305–324.

Aijmer, K. (1996). *Conversational Routines in English: Convention and Creativity*. London: Longman.

Aijmer, K. (2013). *Understanding Pragmatic Markers: A Variational Pragmatic Approach*. Edinburgh: Edinburgh University Press.

Aijmer, K. (2015). Pragmatic markers. In K. Aijmer & C. Rühlemann (Eds.), *Corpus Pragmatics: A Handbook* (pp. 195–218). Cambridge: Cambridge University Press.

Aijmer, K., & Rühlemann, C. (Eds.). (2015). *Corpus Pragmatics: A Handbook*. Cambridge: Cambridge University Press.

Aist, G., Allen, J., Campana, E., Galescu, L., Gómez Gallo, C., Stoness, S., Swift, M., & Tanenhaus, M. (2006). Software architectures for incremental

understanding of human speech. In *Proceedings of the International Conference of Spoken Language Processing (ICSLP), Pittsburg, PA, USA, September 2006* (pp. 1922–1925).

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: some gestures are meant to be seen. *Journal of Memory and Language*, *44*(2), 169–188.

Allwood, J., Björnberg, M., Grönqvist, L., Ahlsen, E., & Ottesjö, C. (2000). The spoken language corpus at the Department of Linguistics, Göteborg University. *Forum: Qualitative Social Research [Online]*, *1*(3). Retrieved from http://www.qualitative-research.net/index.php/fqs/article/view/1026/2215

Ashby, S., Bourban, S., Carletta, J., Flynn, M., Guillemot, M., Hain, T., & Wellner, P. (2005). The AMI meeting corpus. In *Proceedings of Measuring Behavior 2005, Wageningen* (pp. 4–8).

Atkins, S., & Harvey, K. (2010). How to use corpus linguistics in the study of health communication. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 605–619). London: Routledge.

Atkinson, J. M., & Heritage, J. (1984). Transcription notation. In J. M. Atkinson & J. Heritate (Eds.), *Structures of Social Action* (pp. ix–xvi). Cambridge: Cambridge University Press.

Austin, E. E., & Sweller, N. (2014). Presentation and production: The role of gesture in spatial communication. *Journal of Experimental Child Psychology*, *122*(1), 92–103.

Baker, P. (2008). "Eligible" bachelors and "frustrated" spinsters: Corpus linguistics, gender and language. In J. Sunderland, K. Harrington, & H. Stantson (Eds.), *Gender and Language Research Methodologies* (pp. 73–84). London: Palgrave.

Baker, P. (2010). *Sociolinguistics and Corpus Linguistics*. Edinburgh:

Edinburgh University Press.

Ball, M. J., & Rahilly, J. (1999). *Phonetics: The Science of Speech*. London: Routledge.

Bannard, C., & Lieven, E. (2012). Formulaic Language in L1 Acquisition. *Annual Review of Applied Linguistics*, *32*, 3–16.

Bardovi-Harlig, K. (2012). Formulas, Routines, and Conventional Expressions in Pragmatics Research. *Annual Review of Applied Linguistics*, *32*, 206–227.

Barlow, M. (2004). Collocate. *Houston: Athelstan Publications*.

Bavelas, J. B., Chovil, N., Coates, L., & Lori, R. (1995). Gestures Specialized for Dialogue. *Personality and Social Psychology Bulletin*, *21*(4), 394–405.

Bavelas, J. B., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language*, *58*(2), 495–520.

Beattie, G. (2003). *Visible Thought: The New Psychology of Body Language*. London/New York: Routledge.

Beattie, G., & Aboudan, R. (1994). Gestures, pauses and speech: An experimental investigation of the effects of changing social context on their precise temporal relationships. *Semiotica*, *99*(3–4), 239–272.

Beattie, G., & Shovelton, H. (1999). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica*, *123*(1), 1–30.

Beattie, G., & Shovelton, H. (2002). What properties of talk are associated with the generation of spontaneous iconic hand gestures? *British Journal of Social Psychology*, *41*(3), 403–417.

Bednarek, M., & Martin, J. R. (Eds.). (2010). *New Discourse on Language: Functional Perspectives on Multimodality, Identity, and Affiliation*. London/New York: Continuum.

Bednarek, M., & Martin, J. R. (Eds.). (2010). *New Discourse on Language:*

Functional Perspectives on Multimodality, Identity, and Affiliation. London/New York: Continuum.

Bertrand, R., Blache, P., Espesser, R., Ferré, G., Meunier, C., Priego-Valverde, B., & Rauzy, S. (2006). Le CID: Corpus of interactional data-protocoles, conventions, annotations. *Travaux Interdisciplinaires Du Laboratoire Parole et Langage d'Aix En Provence (TIPA), 25*, 25–55.

Bezemer, J., & Kress, G. (2016). *Multimodality, Learning and Communication: A Social Semiotic Frame*. Abingdon/New York: Routledge.

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, *14*(3), 275–311.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, *25*(3), 371–405.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Bickmore, T., & Cassell, J. (2005). Social dialongue with embodied conversational agents. In J. C. J. van Kuppevelt, L. Dybkjær, & N. O. Bernsen (Eds.), *Advances in Natural Multimodal Dialogue Systems* (pp. 23–54). Dordrecht: Springer Netherlands.

Birdwhistell, R. L. (1970). *Kinesics and Context: Essays on Body-motion Communication*. Philadelphia: University of Pennsylvania Press.

Bishop, H. (2004). The effect of typographic salience on the look up and comprehension of unknown formulaic sequences. In N. Schmidt (Ed.), *Formulaic Sequences: Acquisition, Processing, and Use* (pp. 227–244). Philadelphia: John Benjamins Publishing Company.

Blache, P., Bertrand, R., & Ferré, G. (2008). Creating and exploiting multimodal annotated corpora. In *Proceedings of Sixth International Conference on Language Resources and Evaluation (LREC), [online], 2008* (pp. 110–115). Retrieved from

www.lrec-conf.org/proceedings/lrec2008/pdf/449_paper.pdf

Boers, F., & Lindstromberg, S. (2005). Finding ways to make phrase-learning feasible: The mnemonic effect of alliteration. *System*, *33*(2), 225–238.

Boers, F., & Lindstromberg, S. (2012). Experimental and Intervention Studies on Formulaic Sequences in a Second Language. *Annual Review of Applied Linguistics*, *32*, 83–110.

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: putting a Lexical Approach to the test. *Language Teaching Research*, *10*(3), 245–261.

Boers, F., Lindstromberg, S., & Eyckmans, J. (2012). Are alliterative word combinations comparatively easy to remember for adult learners? *RELC Journal*, *43*(1), 127–135.

Bressem, J., & Ladewig, S. H. (2011). Rethinking gesture phases: Articulatory features of gestural movement? *Semiotica*, *184*, 53–91.

Bressem, J., Ladewig, S. H., & Müller, C. (2013). Linguistic annotation system for gestures. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & S. Teßendorf (Eds.), *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction (Volume 1)* (pp. 1098–1124). Berlin: De Gruyter Mouton.

Brookes, H. (2004). A repertoire of South African quotable gestures. *Journal of Linguistic Anthropology*, *14*(2), 186–224.

Brookes, H. (2005). What gestures do: Some communicative functions of quotable gestures in conversations among Black urban South Africans. *Journal of Pragmatics*, *37*(12), 2044–2085.

Brookes, H. (2011). Amangama amathathu 'The three letters': The emergence of a quotable gesture (emblem). *Gesture*, *11*(2), 194–218.

Burdelski, M., & Cook, H. M. (2012). Formulaic language in language socialization. *Annual Review of Applied Linguistics*, *32*, 173–188.

Butterworth, B., & Beattie, G. (1978). Gesture and silence as indicators of

planning in speech. In R. N. Campbell & P. T. Smith (Eds.), *Recent Advances in the Psychology of Language: Formal and Experimental Approaches* (pp. 347–360). New York/London: Springer.

Cafaro, A., Wagner, J., Baur, T., Dermouche, S., Torres, T. M., Pelachaud, C., & Valstar, M. (2017). The NoXi database: Multimodal recordings of mediated novice-expert interactions. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction - ICMI'2017* (pp. 350–359). AMC, New York, USA.

Calbris, G. (1990). *The Semiotics of French Gestures*. Bloomington: Indiana University Press.

Calbris, G. (2003). From cutting an object to a clear cut analysis: Gesture as the representation of a preconceptual schema linking concrete actions to abstract notions. *Gesture, 3*(1), 19–46.

Calbris, G. (2008). From left to right…: coverbal gestures and their symbolic use of space. In A. Cienki & C. Müller (Eds.), *Metaphor and Gesture* (pp. 27–53). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Calbris, G. (2011). *Elements of Meaning in Gesture*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Campbell, N. (2009). Tools and resources for visualising conversational-speech interactionle. In M. Kipp, J. C. Martin, P. Paggio, & D. Heylen (Eds.), *Multimodal Corpora: From Models of Natural Interaction to Systems and Applications* (pp. 176–188). Springer.

Carter, R., & Adolphs, S. (2008). Linking the verbal and visual: New directions for corpus linguistics. *Language and Computers, 64*(1), 275–291.

Carter, R., & McCarthy, M. (2006). *Cambridge Grammar of English: A Comprehensive Guide*. Cambridge: Cambridge University Press.

Carter, R., Goddard, A., Reah, D., Sanger, K., & Bowring, M. (2001). *Working with Texts: A Core Introduction to Language Analysis* (2nd ed.). London: Routledge.

Cassell, J., Vilhjálmsson, H. H., & Bickmore, T. (2004). Beat: the behavior expression animation toolkit. In *Proceedings of SIGGRAPH '01* (pp. 163–185). Springer.

Chen, L., Travis-Rose, R., Parrill, F., Han, X., Tu, J., Huang, Z., & Huang, T. (2005). VACE multimodal meeting corpus. In *Proceedings of the Workshop on Machine Learning for Multimodal Interaction (MLMI) [online]*. Retrieved from http://www.umiacs.umd.edu/~mharper/papers/M-B-1.pdf

Chen, Y., & Adolphs, S. (2017). Multimodal corpus-based investigations of the gesture-speech relationship. University of Birmingham: The 9th International Corpus Linguistics Conference.

Chui, K. (2005a). Temporal patterning of speech and iconic gestures in conversational discourse. *Journal of Pragmatics*, *37*(6), 871–887.

Chui, K. (2005b). Topicality and gesture in Chinese conversational discourse. *Language and Linguistics*, *6*(4), 635–654.

Chui, K. (2009). Linguistic and imagistic representations of motion events. *Journal of Pragmatics*, *41*(9), 1767–1777.

Chui, K. (2011). Conceptual metaphors in gesture. *Cognitive Linguistics*, *22*(3), 437–458.

Chui, K. (2012). Gestural manifestation of knowledge in conceptual frames. *Discourse Processes*, *49*(8), 599–621.

Church, R. B., Kelly, S., & Holcombe, D. (2013). Temporal synchrony between speech, action and gesture during language production. *Language and Cognitive Processes*, *00*(3), 345–354.

Cienki, A. (2008). Why study metaphor and gesture. In A. Cienki & C. Müller (Eds.), *Metaphor and Gesture* (pp. 5–25). Amsterdam/Philadelphia: John Benjamins Punlishing Company.

Cienki, A. (2013). Image schemas and mimetic schemas in cognitive linguistics and gesture studies. *Review of Cognitive Linguistics*, *11*(2), 417–432.

Cienki, A. (2016). Cognitive linguistics, gesture studies, and multimodal communication. *Cognitive Linguistics*, *27*(4), 603–618.

Cienki, A., & Müller, C. (2008a). Metaphor, gesture, and thought. In R. W. Gibbs (Ed.), *The Cambridge Handbook of Metaphor and Thought* (pp. 483–501). Cambridge: Cambridge University Press.

Cienki, A., & Müller, C. (Eds.). (2008b). *Metaphor and Gesture*. Amsterdam/Philadelphia: John Benjamins Punlishing Company.

Congdon, E. L., Novack, M. A., Brooks, N., Hemani-Lopez, N., O'Keefe, L., & Goldin-Meadow, S. (2017). Better together: Simultaneous presentation of speech and gesture in math instruction supports generalization and retention. *Learning and Instruction*, *50*, 65–74.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, *29*(1), 72–89.

Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, *32*, 45–61.

Cook, S. W., Duffy, R. G., & Fenn, K. M. (2013). Consolidation and transfer of learning after observing hand gesture. *Child Development*, *84*(6),

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2010). Gesturing makes memories that last. *Journal of Memory and Language*, *63*(4), 465–475.

Cutting, J. (Ed.). (2007). *Vague Language Explored*. Basingstoke: Palgrave Macmillan.

Dahlmann, I., & Adolphs, S. (2007). Pauses as an indicator of psycholinguistically valid multi-word expressions (MWEs)? *Proceedings of the Workshop on a Broader Perspective on Multiword Expressions*. Prague, Czech Republic: Association for Computational Linguistics.

Dahlmann, I., & Adolphs, S. (2009). Spoken corpus analysis: multimodal approaches to language description. In P. Baker (Ed.), *Contemporary Corpus Linguistics* (pp. 136–150). London/New York: Continuum

International Publishing.

Demir, Ö. E., & Goldin-Meadow, S. (2016). Gesture's Role in Learning and Processing Language. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 275–283). San Diego: Academic Press.

Desmond, M., Collet, P., Marsh, P., O'Shaughnessy, M., & Morris, D. (1979). *Gestures: Their Origins and Distribution*. *England: Jonathan Cape. Ltd*. New York: Stein and Day.

Duncan, S. D. (2006). Co-expressivity of speech and gesture: Manner of motion in Spanish, English, and Chinese. In *27th Berkeley Linguistic Society Annual Meeting* (pp. 353–370). Berkeley, CA: Berkeley University Press.

Duncan, S. D., Cassell, J., & Levy, E. (Eds.). (2007). *Gesture and the Dynamic Dimension of Language: Essays in Honor of David McNeill*. Amsterdam/Philadelphia: John Benjamins Publishing.

Dziemianko, A. (2010). Paper or electronic? The role of dictionary form in language reception, production and the retention of meaning and collocations. *International Journal of Lexicography*, *23*(3), 257–273.

Efron, D. (1941). *Gesture, Race and Culture*. The Hague: Mouton and Co.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, *1*(1), 49–98.

Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal Teddy Bear. *Annual Review of Applied Linguistics*, *32*, 17–44.

Ellis, N. C., & Simpson-Vlach, R. (2009). Formulaic language in native speakers: Triangulating psycholinguistics, corpus linguistics, and education. *Corpus Linguistics and Linguistic Theory*, *5*(1), 61–78.

Ellis, N. C., Simpson-vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *Tesol Quarterly*, *42*(3), 375–396.

Enfield, N. J. (2009). *The Anatomy of Meaning: Speech, Gesture, and*

Composite Utterances (Vol. 8). Cambridge: Cambridge University Press.

Enfield, N. J. (2013a). A ‘composite utterances’ approach to meaning. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & S. Teßendorf (Eds.), *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction (Volume 1)* (pp. 689–707). Berlin: Mouton de Gruyter.

Enfield, N. J. (2013b). Doing fieldwork on the body, language, and communication. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Body - Language - Communication. Volume 1 : An International Handbook on Multimodality in Human Interaction* (pp. 974–981). Berlin/Boston: De Gruyter.

Erman, B. (2001). Pragmatic markers revisited with a focus on you know in adult and adolescent talk. *Journal of Pragmatics*, *33*(9), 1337–1359.

Ferré, G. (2010). Timing relationships between speech and co-verbal gestures in spontaneous French. In *Language Resources and Evaluation, Workshop on Multimodal Corpora*. Malta.

Feyereisen, P. (2006). How could gesture facilitate lexical access? *Advances in Speech-Language Pathology*, *8*(2), 128–133.

Fligelstone, S., Pacey, M., & Rayson, P. (1997). How to generalize the task of annotation. In R. Garside, G. Leech, & A. McEnery (Eds.), *Corpus Annotation: Linguistic Information from Computer Text Corpora* (pp. 122–136). London.

Forceville, C. J., & Urios-Aparisi, E. (Eds.). (2009). *Multimodal Metaphor*. Berlin/New York: Mouton de Gruyter.

Fraser, B. (2015). The combining of Discourse Markers - A beginning. *Journal of Pragmatics*, *86*, 48–53.

Fricke, E. (2013). Towards a unified grammar of gesture and speech: A multimodal approach. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & S. Teßendorf (Eds.), *Body-Language-Communication: An*

International Handbook on Multimodality in Human Interaction (Volume 1) (pp. 733–754). Berlin: De Gruyter Mouton.

Fung, L., & Carter, R. (2007). Discourse markers and spoken english: Native and learner use in pedagogic settings. *Applied Linguistics*, *28*(3), 410–439.

Galati, A., & Brennan, S. E. (2014). Speakers adapt gestures to addressees' knowledge: Implications for models of co-speech gesture. *Language, Cognition and Neuroscience*, *29*(4), 435–451.

Garofolo, J., Laprun, C., Michel, M., Stanford, V., & Tabassi, E. (2004). The NIST meeting room pilot corpus. In *Proceedings of the 4th Language Resources and Evaluation Conference, (LREC) 2004 [online]*. Retrieved from https://catalog.ldc.upenn.edu/docs/LDC2004S09/LREC04-NIST_MR_Paper.pdf

Garside, R., & Rayson, P. (1997). Higher-level annotation tools. In R. Garside, G. Leech, & A. McEnery (Eds.), *Corpus Annotation: Linguistic Information from Computer Text Corpora* (pp. 179–193). London: Longman.

Garside, R., & Smith, N. (1997). A hybrid grammatical tagger: CLAWS4. In R. Garside, G. Leech, & A. McEnery (Eds.), *Corpus Annotation: Linguistic Information from Computer Text Corpora* (pp. 102–121). London: Longman. Retrieved from http://ucrel.lancs.ac.uk/papers/HybridTaggerGS97.pdf

Goldberg, B., & Cannon-Bowers, J. (2015). Feedback source modality effects on training outcomes in a serious game: Pedagogical agents make a difference. *Computers in Human Behavior*, *52*, 1–11.

Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, *32*(10), 1489–1522.

Goodwin, C. (2007). Environmentally coupled gestures. In S. D. Duncan, J. Cassell, & E. Levy (Eds.), *Gesture and the Dynamic Dimension of*

Language: Essays in Honor of David McNeill (pp. 195–212). Amsterdam: John Benjamins Publishing Company.

Gu, Y., Mol, L., Hoetjes, M., & Swerts, M. (2017). Conceptual and Lexical Effects on Gestures : the Case of Vertical Spatial Metaphors for Time in Chinese. *Language, Cognition and Neuroscience*, *32*(8), 1048–1063.

Gullberg, M. (1998). *Gesture as a Communication Strategy in Second Language Discourse: A Study of Learners of French and Swedish*. Lund: Lund University Press.

Gullberg, M. (2006). Some reasons for studying gesture and second language acquisition (Hommage à Adam Kendon). *IRAL-International Review of Applied Linguistics in Language Teaching*, *44*(2), 103–124.

Gullberg, M. (2010). Methodological reflections on gesture analysis in second language acquisition and bilingualism research. *Second Language Research*, *26*(1), 75–102.

Gullberg, M. (2011). Language-specific encoding of placement events in gestures. In J. Bohnemeyer & E. Pederson (Eds.), *Event Representation in Language and Cognition* (pp. 166–188). Cambridge: Cambridge University Press.

Gullberg, M., & Kita, S. (2009). Attention to speech-accompanying gestures: Eye movements and information uptake. *Journal of Nonverbal Behavior*, *33*(4), 251–277.

Gullberg, M., & McCafferty, S. G. (2008). Introduction to gesture and SLA: Toward an integrated approach. *Studies in Second Language Acquisition*, *30*(02), 133–146.

Hadar, U., & Butterworth, B. (1997). Iconic gestures, imagery, and word retrieval in speech. *Semiotica*, *115*(1–2), 147–172.

Halliday, M. A. K. (1978). *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Arnold.

Hanna, B. E. (1996). Defining the emblem. *Semiotica*, *112*(3–4), 289–358.

Harrison, S. (2009). The expression of negation through grammar and gesture. In J. Zlatev, M. Andrén, M. J. Falck, & C. Lundmark (Eds.), *Studies in Language and Cognition* (pp. 405–419). Newcastle: Cambridge Scholars Publishing.

Harrison, S. (2010). Evidence for node and scope of negation in coverbal gesture. *Gesture*, *10*(1), 29–51.

Harrison, S. (2013). The temporal coordination of negation gestures in relation to speech. *TiGeR 2013 - Tiberg Gesture Research Meeting*, 1–4. Retrieved from http://tiger.uvt.nl/pdf/papers/harrison.pdf

Harvey, K., Churchill, D., Crawford, P., Brown, B., Mullany, L., Macfarlane, A., & McPherson, A. (2008). Health communication and adolescents: what do their emails tell us? *Family Practice*, *25*(4), 304–311.

Have, P. ten. (1999). *Doing Conversation Analysis*. London: Sage.

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

Hazel, S., Mortensen, K., & Rasmussen, G. (2014). Introduction: A body of resources – CA studies of social conduct. *Journal of Pragmatics*, *65*, 1–9.

Herrera, D., Novick, D., Jan, D., & Traum, D. (2010). The UTEP-ICT cross-cultural multiparty multimodal dialog corpus. In *Proceedings of the LREC Workshop on Multimodal Corpora, Mediterranean Conference Centre, Malta, 18 May 2010* (pp. 49–54). Retrieved from http://people.ict.usc.edu/~traum/Papers/LREC2010-UTEP-ICT.pdf

Holler, J., & Beattie, G. (2002). A micro-analytic investigation of how iconic gestures and speech represent core semantic features in talk. *Semiotica*, *142*(1/4), 31–70.

Holler, J., & Beattie, G. (2003). How iconic gestures and speech interact in the representation of meaning: Are both aspects really integral to the process? *Semiotica*, *146*, 81–116.

Holler, J., & Beattie, G. (2004). The interaction of iconic gesture and speech in talk. In A. Camurri & G. Volpe (Eds.), *Gesture-Based Communication in*

Human-Computer Interaction (pp. 63–69). Berlin: Springer.

Holler, J., & Stevens, R. (2007). The effect of common ground on how speakers use gesture and speech to represent size information. *Journal of Language and Social Psychology*, *26*(1), 4–27.

Holler, J., & Wilkin, K. (2009). Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task. *Language and Cognitive Processes*, *24*(2), 267–289.

Holler, J., & Wilkin, K. (2011). Co-speech gesture mimicry in the process of collaborative referring during face-to-face dialogue. *Journal of Nonverbal Behavior*, *35*(2), 133–153.

Holler, J., Shovelton, H., & Beattie, G. (2009). Do iconic hand gestures really contribute to the communication of semantic information in a face-to-face context? *Journal of Nonverbal Behavior*, *33*(2), 73–88.

Hough, J., Tian, Y., Ruiter, L. De, Betz, S., Kousidis, S., Schlangen, D., & Ginzburg, J. (2016). DUEL: A multi-lingual multimodal dialogue corpus for disfluency,exclamations and laughter. In *10th edition of the Language Resources and Evaluation Conference*. Slovenia. Retrieved from https://core.ac.uk/download/pdf/47081542.pdf

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, *27*(1), 4–21.

Hyland, K. (2012). Bundles in Academic Discourse. *Annual Review of Applied Linguistics*, *32*, 150–169.

Irvine, C. A., Eigsti, I. M., & Fein, D. A. (2016). Uh, Um, and autism: Filler disfluencies as pragmatic markers in adolescents with optimal outcomes from autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *46*(3), 1061–1070.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for

language development. *Psychological Science*, *16*(5), 367–371.

Jacobs, N., & Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language*, *56*(2), 291–303.

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation Analysis: Studies from the First Generation.* (pp. 13–31). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Jewitt, C. (2013). Multimodal methods for researching digital technologies. In S. Price, C. Jewitt, & B. Brown (Eds.), *The Sage Handbook of Digital Technology Research* (pp. 250–265). London: Sage.

Johnson, M. (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago/London: The University of Chicago Press.

Jucker, A. H., Schreier, D., & Hundt, M. (Eds.). (2009). *Corpora: Pragmatics and Discourse*. Amsterdam: Rodopi.

Kapoor, A., & Picard, R. W. (2001). A real-time head nod and shake detector. In *Proceedings from the Workshop on Perceptive User Interfaces*. ACM. Retrieved from https://dam-prod.media.mit.edu/x/files/pub/tech-reports/TR-544.pdf

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. *The Relationship of Verbal and Nonverbal Communication*, *26*(26), 207–227.

Kendon, A. (1992). Some recent work from Italy on "Quotable Gestures (Emblems)." *Journal of Linguistic Anthropology*, *2*(1), 92–108.

Kendon, A. (1995). Gestures as illocutionary and discourse structure markers in Southern Italian conversation. *Journal of Pragmatics*, *23*(3), 247–279.

Kendon, A. (1996). An agenda for gesture studies. *Semiotic Review of Books*, *7*(3), 1–22.

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge:

Cambridge University Press.

Kendon, A. (2015). Gesture and sign. In K. Allan (Ed.), *The Routledge Handbook of Linguistics* (p. 33). London/New York: Routledge.

Kendon, A., Sebeok, T. A., & Umiker-Sebeok, J. (Eds.). (1981). *Nonverbal Communication, Interaction, and Gesture: Selections from SEMIOTICA*. Hague: De Gruyter Mouton.

Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London/New York: Longman.

Kipp, M. (2001). Anvil-a generic annotation tool for multimodal dialogue. In *7th European Conference on Speech Communication and Technology, Aalborg, Denmark, September 3-7, 2001* (pp. 1367–1370).

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and Gesture* (pp. 162–185). Cambridge: Cambridge University Press.

Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: A review. *Language and Cognitive Processes*, *24*(2), 145–167.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*(1), 16–32.

Kita, S., Özyürek, A., Allen, S., Brown, A., Furman, R., & Ishizuka, T. (2007). Relations between syntactic encoding and co-speech gestures: Implications for a model of speech and gesture production. *Language and Cognitive Processes*, *22*(8), 1212–1236.

Kita, S., van Gijn, I., & van der Hulst, H. (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. In I. Wachsmuth & M. Fröhlich (Eds.), *Gesture and Sign Language in Human-Computer Interaction: International Gesture Workshop Bielefeld, Germany, September 17–19, 1997 Proceedings* (pp. 23–35). Springer.

Knight, D. (2011). *Multimodality and Active Listenership: A Corpus Approach*. London: Bloomsbury.

Knight, D., & Tennent, P. (2008). Introducing DRS (The Digital Replay System): A tool for the future of Corpus Linguistic research and analysis. In N. Calzolari (Ed.), *Proceedings of the 6th Language Resources and Evaluation Conference (LREC), Palais des Congrés, Marrakech, Morocco* (pp. 26–31). European Language Resources Association.

Knight, D., Evans, D., Carter, R., & Adolphs, S. (2009). HeadTalk, HandTalk and the corpus: Towards a framework for multi-modal, multi-media corpus development. *Corpora*, *4*(1), 1–32.

Knight, D., Evans, D., Carter, R., & Adolphs, S. (2009b). Redrafting corpus development methodologies: Blueprints for 3rd generation "multimodal, multimedia''corpora. *Corpora*, *4*(1), 1–32.

Kok, K. (2017). Functional and temporal relations between spoken and gestured components of language. *International Journal of Corpus Linguistics*, *22*(1), 1–26.

Koumoutsakis, T., Church, R. B., Alibali, M. W., Singer, M., & Ayman-Nolley, S. (2016). Gesture in instruction: Evidence from live and video lessons. *Journal of Nonverbal Behavior*, *40*(4), 301–315.

Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and Gesture* (pp. 261–283). Cambridge: Cambridge University Press.

Kress, G., & van Leeuwen, T. (1996). *Reading Images*. London: Routledge.

Kuiper, K. (2004). Formulaic performance in conventionalised varieties of speech. In N. Schmitt (Ed.), *Formulaic Sequences: Acquisition, Processing and Use* (pp. 37–54). Amsterdam: John Benjamins Publishing Company.

Kushch, O., Igualada, A., & Prieto, P. (2018). Prominence in speech and gesture favour second language novel word learning. *Language, Cognition and Neuroscience*, *33*(8), 992–1004.

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Ladewig, S. H. (2011). Putting the cyclic gesture on a cognitive basis. *CogniTextes. Revue de l'Association Française de Linguistique Cognitive*, (Volume 6).

Ladewig, S. H. (2014a). Recurrent gestures. In C. Müller, A. Cienki, F. Ellen, S. H. Ladewig, D. McNeill, & J. Bressem (Eds.), *Body - Language - Communication: An International Handbook on Multimodality in Human Interaction (Volume 2)* (pp. 1558–1574). Berlin/Bosten: De Gruyter Mouton.

Ladewig, S. H. (2014b). The cyclic gesture. In C. Müller, A. Cienki, F. Ellen, S. H. Ladewig, D. McNeill, & J. Bressem (Eds.), *Body - Language - Communication: An International Handbook on Multimodality in Human Interaction (Volume 2)* (pp. 1605–1618). Berlin/Bosten: De Gruyter Mouton.

Ladewig, S. H., & Bressem, J. (2013). New insights into the medium hand: Discovering recurrent structures in gestures. *Semiotica*, *2013*(197), 203–231.

Lakoff, G., & Johnson, M. (2008). *Metaphors We Live by*. Chicago: University of Chicago press.

Laufer, B. (2011). The contribution of dictionary use to the production and retention of collocations in a second language. *International Journal of Lexicography*, *24*(1), 29–49.

Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, *26*(10), 1457–1471.

Lin, P. (2012). Sound evidence: The missing piece of the jigsaw in formulaic language research. *Applied Linguistics*, *33*(3), 342–347.

Lindstromberg, S., & Boers, F. (2008a). Phonemic repetition and the learning of

lexical chunks: The power of assonance. *System, 36*(3), 423–436. h

Lindstromberg, S., & Boers, F. (2008b). The mnemonic effect of noticing alliteration in lexical chunks. *Applied Linguistics*, *29*(2), 200–222.

Lindstromberg, S., Eyckmans, J., & Connabeer, R. (2016). A modified dictogloss for helping learners remember L2 academic English formulaic sequences for use in later writing. *English for Specific Purposes, 41*, 12–21.

Louw, B. (1993). Irony in the text or insincerity in the writer?: The diagnostic potential of semantic prosodies. In M. Baker, G. Grancis, & E. Tognini-Bonelli (Eds.), *Text and Technology: In honour of John Sinclair* (pp. 157–176). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Lücking, A., Bergman, K., Hahn, F., Kopp, S., & Rieser, H. (2010). The Bielefeld speech and gesture alignment corpus (SaGA). In *Proceedings of the LREC Workshop on Multimodal Corpora, Mediterranean Conference Centre, Malta, 18 May 2010* (pp. 92–98). Retrieved from https://pub.uni-bielefeld.de/download/2001935/2534579

Machin, D. (2016). *Introduction to Multimodal Analysis*. London: Bloomsbury.

Maíz-Arévalo, C. (2017). Expressive speech acts in educational e-chats. *Soprag, 5*(2), 151–178.

Mana, N., Lepri, B., Chippendale, P., Cappelletti, A., Pianesi, F., Svaizer, P., & Zancanaro, M. (2007). Multimodal corpus of multi-party meetings for automatic social behavior analysis and personality traits detection. In *Proceeding of Workshop on Tagging, Mining and Retrieval of Human-Related Activity Information at ICMI'07, Nagoya, Japan* (pp. 9–14). Retrieved from https://i3.fbk.eu/sites/i3.fbk.eu/files/multimodal_corpus_ms2.pdf

Mayberry, R. I., & Nicoladis, E. (2000). Gesture reflects language development evidence from bilingual children. *Current Directions in Psychological Science, 9*(6), 192–196.

McCarthy, M., & Carter, R. (1995). Spoken grammar: what is it and how can we teach it? *ELT Journal*, *49*(3), 207–218.

McCowan, S., Bengio, D., Gatica-Perez, G., Lathoud, F., Monay, D., Moore, P., Bourlard. (2003). Modelling human interaction in meetings. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hong Kong, April 2003* (pp. 748–751).

McEnery, T., & Wilson, A. (1996). *Corpus Linguistics*. Edinburgh: Edinburg University Press.

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, *92*(3), 350.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago/London: The University of Chicago Press.

McNeill, D. (2005). *Gesture and Thought*. Chicago/London: The University of Chicago Press.

McNeill, D. (2008). Unexpected metaphors. In A. Cienki & C. Müller (Eds.), *Metaphor and Gesture* (pp. 155–170). Amsterdam/Philadelphia.

Meij, H. van der, Meij, J. van der, & Harmsen, R. (2015). Animated pedagogical agents effects on enhancing student motivation and learning in a science inquiry learning environment. *Educational Technology Research and Development*, *63*(3), 381–403.

Meyer, C. F. (2002). *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press.

Mittelberg, I. (2008). Peircean semiotics meets conceptual metaphor: Iconic modes in gestural representations of grammar. In A. Cienki & C. Müller (Eds.), *Metaphor and Gesture* (pp. 115–154). Amsterdam/Philadelphia: John Benjamins Punlishing Company.

Mittelberg, I. (2014). Gestures and iconicity. In C. Müller, A. Cienki, F. Ellen, S. H. Ladewig, D. McNeill, & J. Bressem (Eds.), *Body - Language - Communication: An International Handbook on Multimodality in Human*

Interaction (Volume 2) (pp. 1712–1732). Berlin/Bosten: De Gruyter Mouton.

Mittelberg, I., & Waugh, L. R. (2009). Metonymy First, Metaphor Second: A Cognitivesemiotic Approach to Multimodal Figures of Thought in Co-speech Gesture. In C. J. Forceville & E. Urios-Aparisi (Eds.), *Multimodal Metaphor* (pp. 329–356). Berlin/New York: Mouton de Gruyter.

Mittelberg, I., & Waugh, L. R. (2014). Gestures and metonymy. In C. Müller, A. Cienki, F. Ellen, S. H. Ladewig, D. McNeill, & J. Bressem (Eds.), *Body - Language - Communication: An International Handbook on Multimodality in Human Interaction (Volume 2)* (pp. 1747–1766). Berlin/Bosten: De Gruyter Mouton.

Mondada, L. (2007). Multimodal resources for turn-taking pointing and the emergence of possible next speakers. *Discourse Studies*, *9*(2), 194–225.

Mondada, L. (2014). The local constitution of multimodal resources for social interaction. *Journal of Pragmatics*, *65*, 137–156.

Müller, C. (2009). Gesture and language. In K. Malmkjaer (Ed.), *Routledge's Linguistics Encyclopedia* (pp. 214–217). Abington/New York: Routledge.

Müller, C., & Cienki, A. (2009). Words, Gestures, and Beyond: Forms of Multimodal Metaphor in the Use of Spoken Language. In C. J. Forceville & E. Urios-Aparisi (Eds.), *Multimodal Metaphor* (pp. 297–398). Berlin/New York: Mouton de Gruyter.

Müller, C., Bressem, J., & Ladewig, S. H. (2013). Towards a grammar of gestures: A form-based view. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & S. Teßendorf (Eds.), *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction (Volume 1)* (pp. 707–733). Berlin: De Gruyter Mouton.

Müller, C., Ladewig, S. H., & Bressem, J. (2013). Gestures and speech from a

linguistic perspective: A new field and its history. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & S. Teßendorf (Eds.), *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction (Volume 1)* (pp. 55–81). Berlin: De Gruyter Mouton.

Nicoladis, E., Pika, S., Yin, H. U. I., & Marentette, P. (2007). Gesture use in story recall by Chinese–English bilinguals. *Applied Psycholinguistics, 28*(04), 721–735.

Norris, S., & Maier, C. D. (Eds.). (2014). *Interactions, Images and Texts: A Reader in Multimodality*. Berlin: Walter de Gruyter.

Núñez, R. E., & Sweetser, E. (2006). With the Future Behind Them : Convergent Evidence From Aymara Language and Gesture in the Crosslinguistic Comparison of Spatial Construals of Time, *30*, 401–450.

O'Keeffe, A., & Adolphs, S. (2008). Response tokens in British and Irish discourse: corpus, context and variational pragmatics. In K. P. Schneider & A. Barron (Eds.), *Variational Pragmatics: A Focus on Regional Varieties in Pluricentric Languages* (pp. 69–98). Amsterdam: John Benjamins Publishing Company.

O'Keeffe, A., Clancy, B., & Adolphs, S. (2011). *Introducing Pragmatics in Use*. London: Routledge.

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From Corpus to Cassroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press.

Özyürek, A. (2002). Do speakers design their cospeech gestures for their addressees?: The effects of addressee location on representational gestures. *Journal of Memory and Language*, *46*(4), 688–704.

Paggio, P., & Navarretta, C. (2017). The Danish NOMCO corpus: Multimodal interaction in first acquaintance conversations. *Language Resources and Evaluation*, *51*(2), 463–494.

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, *32*, 130–149.

Parrill, F. (2008). Form, meaning, and convention: A comparison of a metaphoric gesture with an emblem. *Metaphor and Gesture*, 195–217.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and Communication*, *191*, 225.

Payrató, L. (1993). A pragmatic view on autonomous gestures: A first repertoire of Catalan emblems. *Journal of Pragmatics*, *20*(3), 193–216.

Petchko, K. (2010). Input enhancement, noticing, and incidental vocabulary acquisition. *Asian EFL Journal*, (13), 228–225.

Pfeiffer, T. (2013a). Documentation of gestures with data gloves. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & S. Teßendorf (Eds.), *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction (Volume 1)* (pp. 868–879). Berlin: Mouton de Gruyter.

Pfeiffer, T. (2013b). Documentation of gestures with motion capture. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & S. Teßendorf (Eds.), *Body - Language - Communication (Volume 1) : An International Handbook on Multimodality in Human Interaction* (pp. 857–868). Berlin/Boston: De Gruyter.

Porter, A. (2016). A helping hand with language learning: Teaching French vocabulary with gesture. *Language Learning Journal*, *44*(2), 236–256.

Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, *13*(4), 519–549.

Read, J., & Nation, P. (2004). Measurement of formulaic sequences. In N. Schmidt (Ed.), *Formulaic Sequences: Acquisition, Processing and Use* (pp. 23–35). Amsterdam: John Benjamins Publishing Company.

Rehm, M., Nakano, Y., Huang, H. H., Lipi, A. A., Yamaoka, Y., & Grüneberg, F. (2008). Creating a standardized corpus of multimodal interactions for

enculturating conversational interfaces. In *Proceedings of the IUI-Workshop on Enculturating Interfaces (ECI), Gran Canaria, 2008*. ACM Press.

Reppen, R. (2010). Building a corpus: what are the key considerations? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (1st ed., pp. 31–37). London; New York, N.Y.: Routledge.

Reppen, R., Fitzmaurice, S. M., & Biber, D. (2002). *Using Corpora to Explore Linguistic Variation*. Amsterdam: John Benjamins Publishing Company.

Richards, K., & Seedhouse, P. (Eds.). (2007). *Applying Conversation Analysis*. *Applying Conversation Analysis*. Basingstoke: Palgrave Macmillan.

Ring, L., Shi, L., Totzke, K., & Bickmore, T. (2015). Social support agents for older adults: longitudinal affective computing in the home. *Journal on Multimodal User Interfaces*, *9*(1), 79–88.

Romero-Trillo, J. (2008). *Pragmatics and Corpus Linguistics: A Mutualistic Entente*. Berlin: Walter de Gruyter.

Romero-Trillo, J. (Ed.). (2013). *Yearbook of Corpus Linguistics and Pragmatics: New Domains and Methodologies*. Dordrecht: Springer.

Romero-Trillo, J. (Ed.). (2014). *Yearbook of Corpus Linguistics and Pragmatics: New Empirical and Theoretical Paradigms*. Cham: Springer.

Romero-Trillo, J. (Ed.). (2015). *Yearbook of Corpus Linguistics and Pragmatics: Current Approaches to Discourse and Translation Studies*. Cham: Springer.

Romero-Trillo, J. (Ed.). (2016). *Yearbook of Corpus Linguistics and Pragmatics: Global Implications for Society and Education in the Networked Age*. Cham: Springer.

Ronan, P. (2015). Categorizing expressive speech acts in the pragmatically annotated SPICE Ireland corpus. *ICAME Journal*, *39*(1), 25–45.

Rossini, N. (2012). *Reinterpreting Gesture as Languuage: Language "in Action."* Amsterdam: IOS Press.

Rühlemann, C., & Clancy, B. (2018). Corpus linguistics and pragmatics. In C. Ilie & N. R. Norrick (Eds.), *Pragmatics and its Interfaces* (pp. 241–266). John Benjamins Publishing Company.

Salvato, G. (2008). The representation of gestures in Italian textbooks and workbooks. *Italica*, *85*(1), 1–26.

Sanchez-Cortes, D., Aran, O., Jayagopi, D. B., Mast, M. S., & Gatica-Perez, D. (2013). Emergent leaders through looking and speaking: From audio-visual data to multimodal recognition. *Journal on Multimodal User Interfaces*, *7*, 39–53.

Schegloff, E. A. (1984). On Some Gestures' Relation to Talk. In J. M. Atkinson & J. Heritage (Eds.), *Structures of Social Action* (pp. 266–296). Cambridge: Cambridge University Press.

Schiel, F., & Mögele, H. (2008). Talking and looking: the smart web multimodal interaction corpus. In *Proceedings of Sixth International Conference on Language Resources and Evaluation (LREC) 2008 [online]*. Retrieved from http://www.lrec-conf.org/proceedings/lrec2008/pdf/510_paper.pdf

Schiel, F., Steininger, S., & Türk, U. (2002). The smart Kom multimodal corpus at BAS. In *Proceedings of the 3rd Language Resources and Evaluation Conference (LREC), 2002 [online]*. Retrieved from http://www.lrec-conf.org/proceedings/lrec2002/pdf/49.pdf

Schmidt, R. W. (1990). The role of consciousness in Second Language Acquisition. *Applied Linguistics*, *11*(2), 129–158.

Schmitt, N. (Ed.). (2004). *Formulaic Sequences: Acquisition, Processing, and Use*. Amsterdam: John Benjamins Publishing Company.

Schmitt, N., & Underwood, G. (2004). Exploring the processing of formulaic sequences through a self-paced reading task. In N. Schmitt (Ed.), *Formulaic Sequences: Acquisition, Processing and Use* (pp. 173–189). Amsterdam: John Benjamins Publishing Company.

Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived recurrent

clusters psycholinguistically valid? In N. Schmidt (Ed.), *Formulaic Sequences: Acquisition, Processing and Use* (pp. 127–151). Amsterdam: John Benjamins Publishing Company.

Schourup, L. (1999). Discourse markers. *Lingua, 107*, 227–265.

Seedhouse, P., & Knight, D. (2016). Applying digital sensor technology: A problem-solving approach. *Applied Linguistics*, *37*(1), 7–32.

Seyfeddinipur, M. (2006). *Disfluency: Interrupting Speech and Gesture*. Nijmegen: MPI-Series in Psycholinguistics.

Sherzer, J. (1991). The Brazilian Thumbs-Up Gesture. *Journal of Linguistic Anthropology, 1*(2), 189–197.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, *31*(4), 487–512.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press .

Sinclair, J. (1996). The search for units of meaning. *Textus*, *9*(1), 1000–1032.

Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. Routledge.

Singer, M. A., & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science*, *16*(2), 85–89.

Sloetjes, H., & Wittenburg, P. (2008). Annotation by category-ELAN and ISO DCR. In *The 6th International Conference on Language Resources and Evaluation (LREC 2008)* (pp. 816–820).

Smith, N., Hoffmann, S., & Rayson, P. (2008). Corpus tools and methods, today and tomorrow: Incorporating linguists' manual annotations. *Literary and Linguistic Computing, 23*(2), 163–180.

So, W. C., Sim Chen-Hui, C., & Low Wei-Shan, J. (2012). Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall? *Language and Cognitive Processes*, *27*(5), 665–681.

Sperberg-McQueen, C., & Burnard, L. (1995). The design of the TEI encoding

scheme. *Computers and the Humanities*, *29*(1), 17–39.

Stam, G. (2006). Thinking for speaking about motion: L1 and L2 speech and gesture. *IRAL-International Review of Applied Linguistics in Language Teaching*, *44*(2), 145–171.

Steen, F., & Turner, M. B. (2013). Multimodal construction grammar. In M. Borkent, B. Dancygier, & J. Hinnell (Eds.), *Language and the Creative Mind*. Stanford: C S L I Publications/Center for the Study of Language & Information. Retrieved from http://cogweb.ucla.edu/crp/Papers/SteenTurner_2013_MultimodalConstructions.pdf

Streeck, J. (2009a). Forward-gesturing. *Discourse Processes*, *46*(2–3), 161–179.

Streeck, J. (2009b). *Gesturecraft: The Manu-facture of Meaning*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Streeck, J. (2010). Ecologies of gesture. In J. Streeck (Ed.), *New Adventures in Language and Interaction* (pp. 223–242). Amsterdam/Philadelphia: John Benjamins Publishing.

Stubbe, M., & Holmes, J. (1995). You know, eh and other "exasperating expressions": An analysis of social and stylistic variation in the use of pragmatic devices in a sample of New Zealand English. *Language and Communication*, *15*(1), 63–88.

Swales, J. M. (1990). *Genre Analysis*. Cambridge: Cambridge University Press.

Talmy, L. (1975). Semantics and syntax of motion. In J. P. Kimball (Ed.), *Syntax and Semantics* (pp. 181–238). New York: NY: Academic Press.

Thomas, J. A. (1995). *Meaning in Interaction: An Introduction to Pragmatics*. New York: Routledge.

Thompson, P. (2010). Building a specialised audio-visual corpus. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 93–103). London/New York: Routledge.

Thomson, H. (2017). Building speaking fluency with multiword expressions.

TESL Canada Journal, *34*(3), 26–53.

Tognini-Bonelli, E. (1996). *Corpus Linguistics at Work*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Tree, J. E. F., & Schrock, J. C. (2002). Basic meanings of you know and I mean. *Journal of Pragmatics*, *34*(6), 727–747.

Trojanová, J., Hrúz, M., Campr, P., & Železný, M. (2008). Design and recording of Czech audio-visual database with impaired conditions for continuous speech recognition. In *Proceedings of Sixth International Conference on Language Resources and Evaluation (LREC)*, Heidelberg.

Tsuchiya, K. (2013). *Listenership Behaviours in Intercultural Encounters: A Time-aligned Multimodal Corpus Analysis*. Amsterdam: John Benjamins Publishing Company.

Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic Sequences: Acquisition, Processing, and Use* (pp. 153–172). Amsterdam: John Benjamins Publishing Company.

Valenzeno, L., Alibali, M. W., & Klatzky, R. (2003). Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology*, *28*(2), 187–204.

van Son, R. J. J. H., Wesseling, W., Sanders, E., & van der Heuvel, H. (2008). The IFADV corpus: A free dialog video corpus. In *Proceedings of Sixth International Conference on Language Resources and Evaluation (LREC), 2008 [online]*. Retrieved from http://www.lrec-conf.org/proceedings/lrec2008/slides/132.pdf

Vilhjálmsson, H., Cantelmo, N., Cassell, J., Chafai, N. E., Kipp, M., Kopp, S., & Pelachaud, C. (2007). The behavior markup language: Recent developments and challenges. In *Proceedings of the 7th international conference on Intelligent Virtual Agents, Paris, France, September 17-19, 2007* (pp. 99–111).

Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning*, *63*(1), 91–120.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). Elan: A professional framework for multimodality research. *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation.*

Wolf, J. C., & Bugmann, G. (2006). Linking speech and gesture in multimodal instruction systems. In *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication, (RO-MAN06), 2006 [online]* (pp. 141–144). Retrieved from https://www.swrtec.de/swrtec/publications/Ro-Man-06_V10_A4.pdf

Wood, D. (2009). Effects of focused instruction of formulaic sequences on fluent expression in second language narratives: A case study. *Canadian Journal of Applied Linguistics*, *12*(1), 39–57.

Woods, H. B. (1991). Social differentiation in Ottawa English. In J. Cheshire (Ed.), *English around the World: Sociolinguistic Perspectives* (pp. 134–150). Cambridge: Cambridge University Press.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, *21*(4), 463–489.

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Wray, A. (2008). *Formulaic Language: Pushing the Boundaries*. Oxford: Oxford University Press.

Wray, A. (2012). What do we (think we) know about formulaic language?: An evaluation of the current state of play. *Annual Review of Applied Linguistics*, *32*, 231–254.

Wu, S., Witten, I. H., & Franken, M. (2010). Utilizing lexical data from a Web-derived corpus to expand productive collocation knowledge. *ReCALL*, *22*(1), 83–102.

Zadeh, A., Zellers, R., Pincus, E., & Morency, L. P. (2016). MOSI: Multimodal corpus of sentiment intensity and subjectivity analysis in online opinion videos. *IEEE Intelligent Systems*, *31*(6), 82–88.

Železný, M., Krňoul, Z., Císař, P., & Matoušek., J. (2006). Design, implementation and evaluation of the Czech realistic audio-visual speech synthesis. *Signal Processing*, *83*(12), 3657–3673.

Zima, E. (2014). *English multimodal motion constructions. A construction grammar perspective.* Retrieved from http://uahost.uantwerpen.be/linguist/SBKL/sbkl2013/Zim2013.pdf